

Analyzing an Archive of 20 Years of T&S Research

Looking back to move forward: How 20 years of empirical T&S research unveils a better path forward

Michael Bochkur Dratver and Matthew Katsaros



Introduction

The Trust & Safety Foundation Archive

As part of their “History of Trust & Safety” project, the Trust & Safety Foundation (TSF) constructed an archive of peer-reviewed articles, books, and book chapters written about various online T&S topics. The aim of this archive was to begin to organize all of the academic literature written on the T&S. The archive, constructed during 2023, contains articles published between 1974 and 2023 (the archive only has a partial record of 2023, given it was created as articles were still being published that year). The TSF described the creation of the archive as such:

We prioritized our search in the publications and conferences for the following disciplines: Communications, Human-Computer Interaction (HCI), Law, Media Studies, and Science & Technology Studies (STS). We began our search with the following keywords/phrases: content moderation, online moderation, online harassment, online safety, online governance, cybersafety, and platform governance. We used these search terms to find relevant research across several research databases including Google Scholar, the Association for Computing Machinery Digital Library, LexisNexis, Scopus, and Web of Science. We read the resulting research and added relevant items to the Zotero archive. Additionally, leveraging the sources we read, we identified other relevant scholarship based on citations and author keywords. –TSF Founders

This archive of 1,288 works made up the corpus that we analyzed for this project.

Process

A Focus on Empirical Research

As the T&S field matures, policymakers, regulators, and industry practitioners are increasingly hungry for empirical evidence to guide their work. When analyzing the T&S archive, we took a particular focus on trying to better understand the breadth of empirical work that has been conducted in the T&S literature.

Labeling Process

As part of the Justice Collaboratory’s Social Media Governance Initiative, we downloaded a the TSF Zotero archive as a spreadsheet including the title, abstract, a link to the article, and article metadata (authors, publisher, article type, publishing date, etc). Our postbaccalaureate fellow, Michael Bochkur Dratver, then began a systematic coding of all 1,288 articles in the archive using a codebook that was developed collaboratively among those in our lab and feedback from others interested in the project.

First, each article was assessed as to whether it was empirical or not. No additional labels were added to articles which were determined not to be empirical. Those that were empirical were then labeled according to the following categories:

- Which platform(s), if any, was being studied?
- Was the study design observational, quasi-experimental, an experimental lab study, or an experimental field study?
- What were the methodologies used to collect the data used in the study?
- What types of data were used in the study?
- What were the different topics of the study?
- What level of collaboration existed between the researchers and the platform or groups of platform users (like community moderators)?

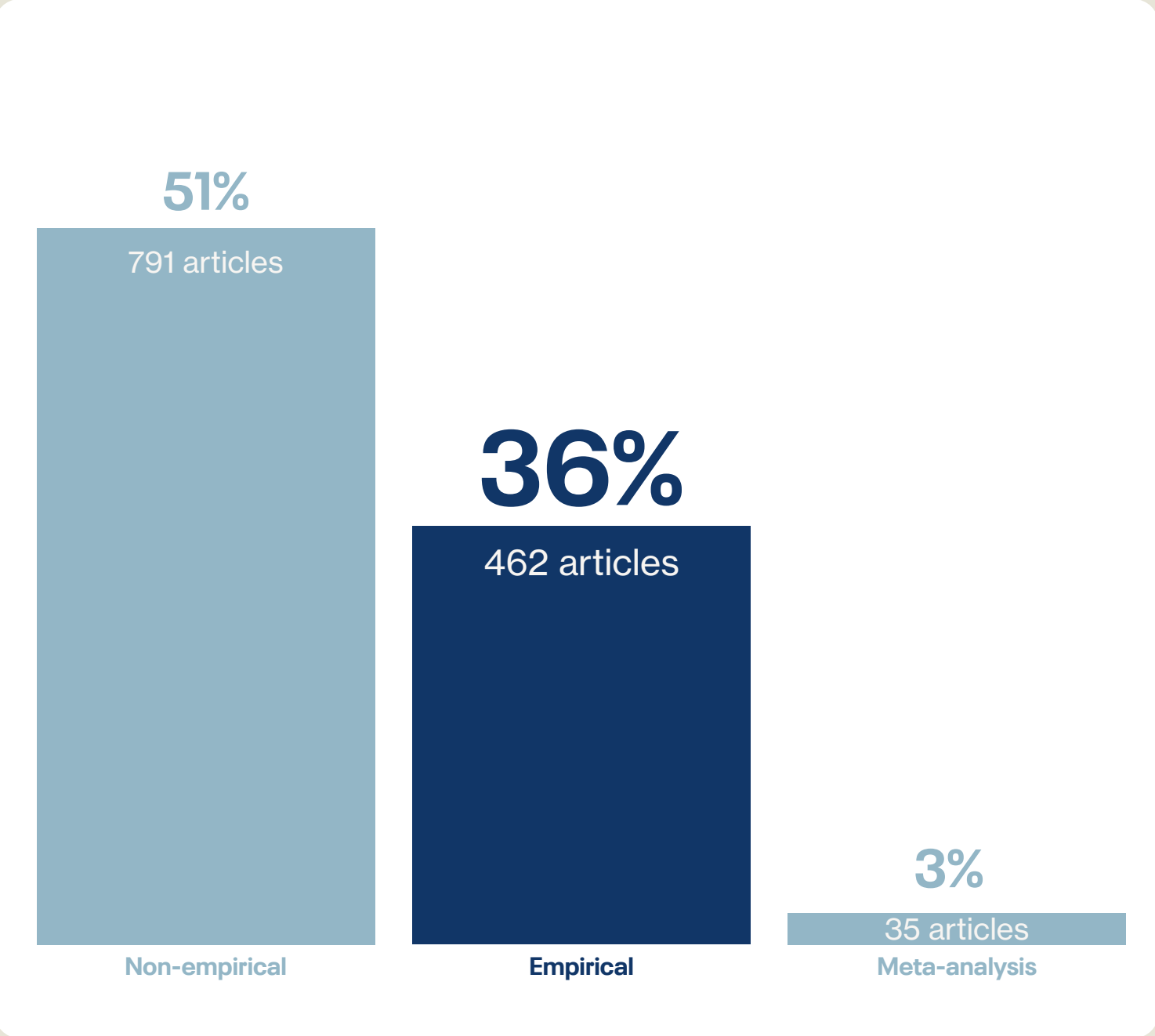
Articles were read in as much detail as was necessary to assign the appropriate labels. In many cases, the abstract was used to determine whether an article was empirical or not. For articles that were empirical, it was usually necessary to read significant portions of the article, with a specific focus on the methods section. Each article was labeled only by the single annotator. A future iteration of this project may include employing multiple annotators to ensure there is agreement on applied labels. The fully annotated dataset is available on Airtable for anyone to view, filter, download, and interact with:



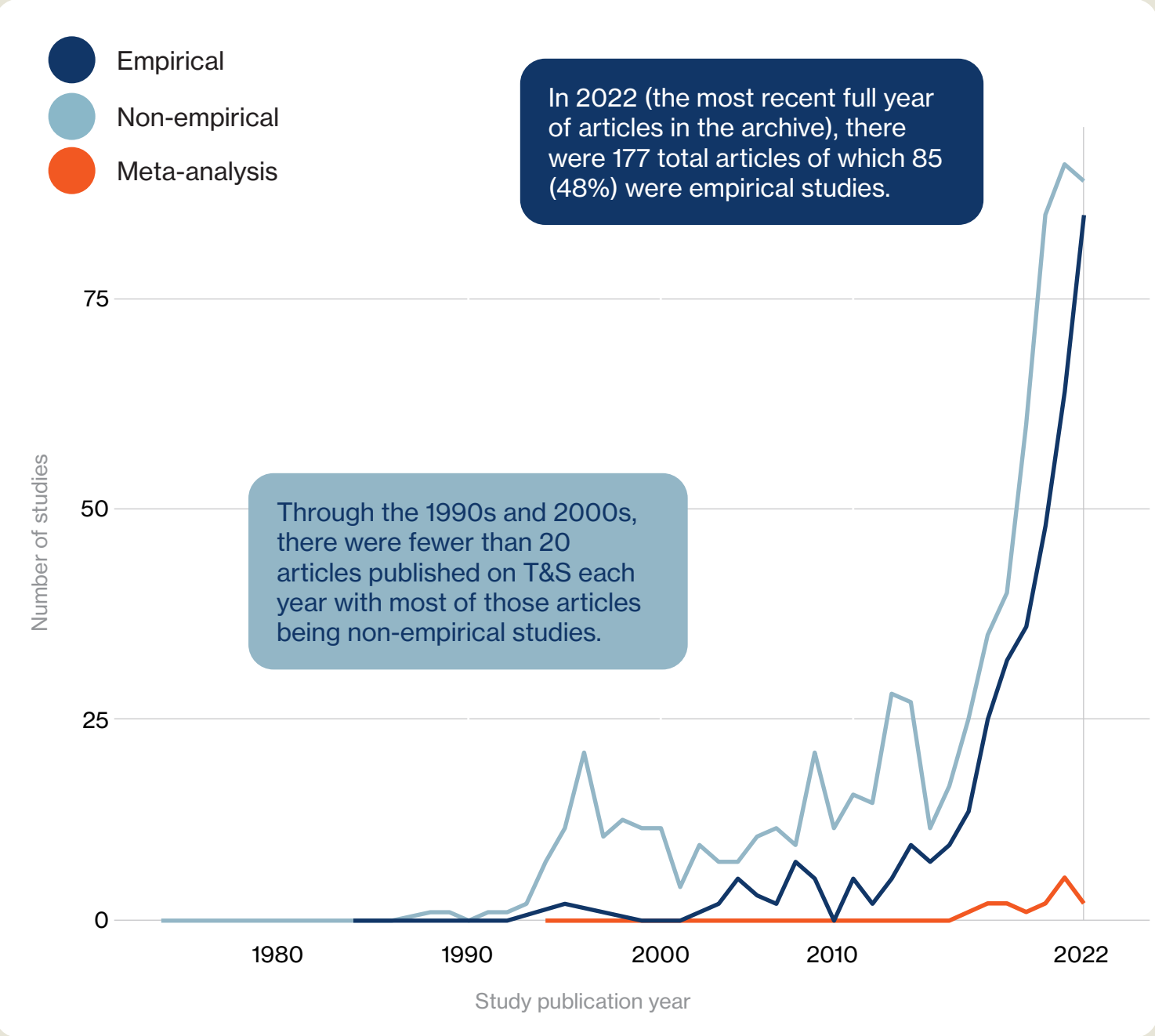
bit.ly/ts_archive_labeling

Findings

Across the archive, we found 462 articles (36% of the archive) that were **empirical studies**.

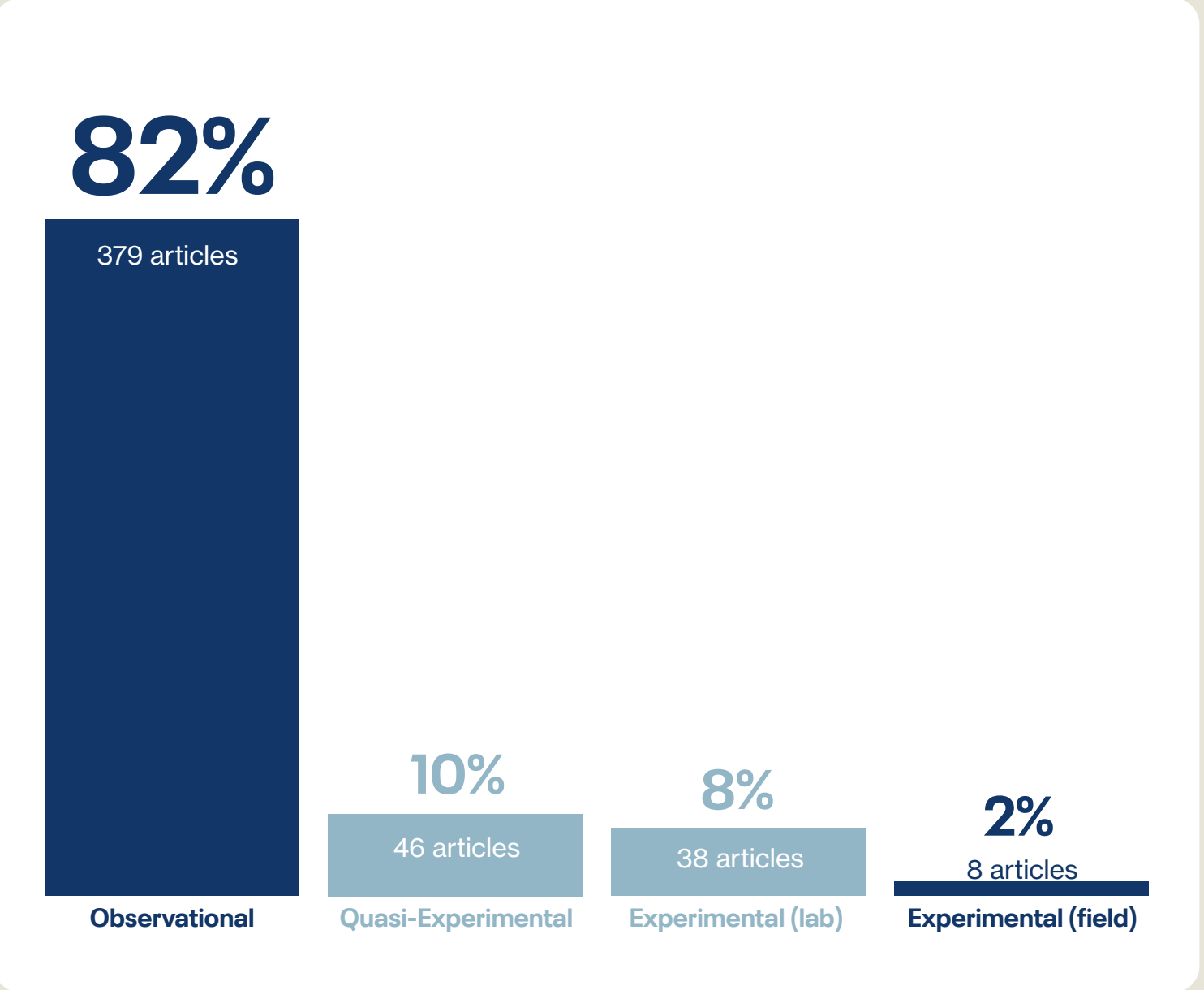


The study of T&S as a field has **increased significantly** over time.

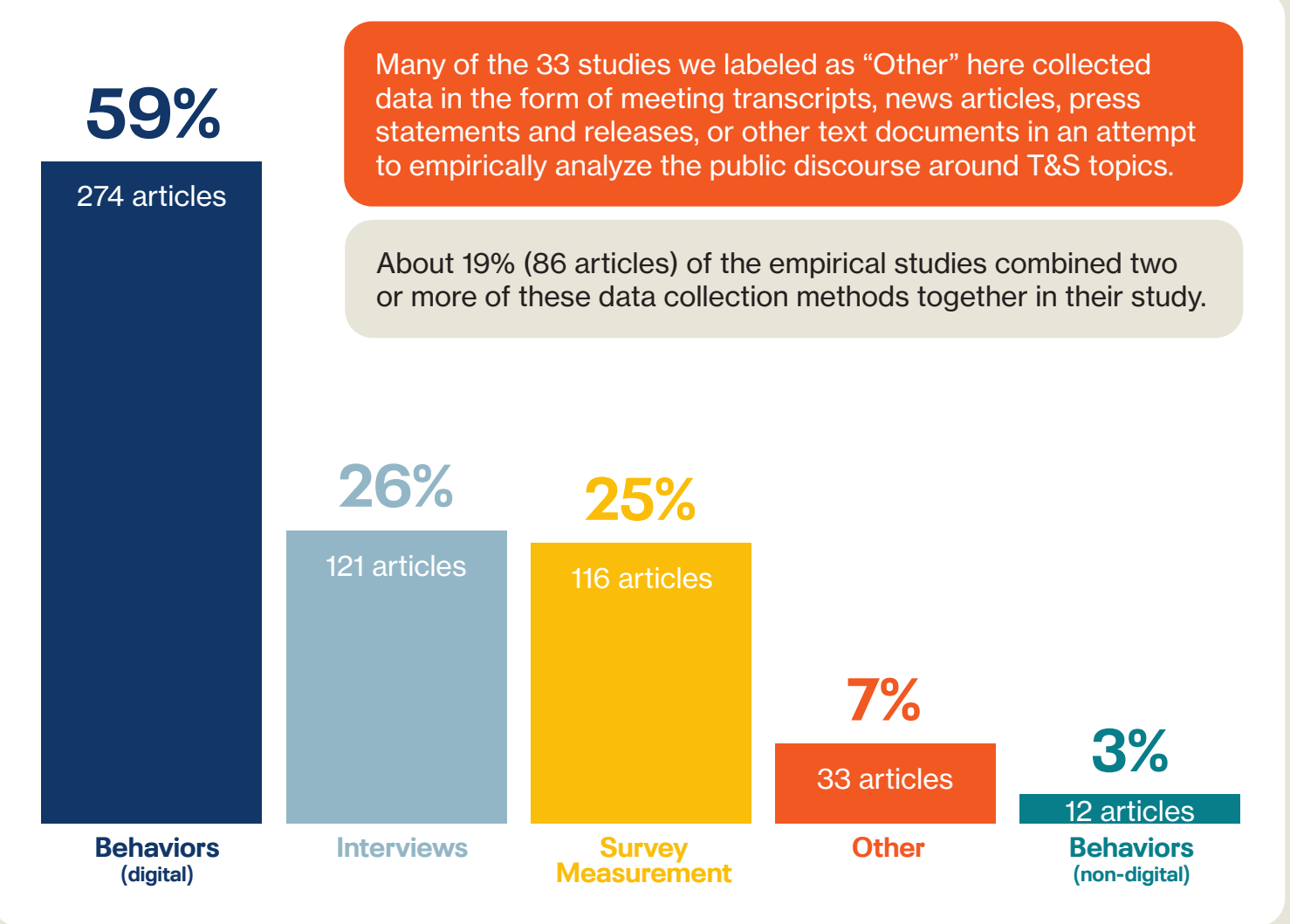


The remainder of the analysis focuses on the 36% of the archive containing the 462 empirical studies.

The vast majority of empirical studies in the archive are **non-experimental observational studies**. Of the 20% of experimental studies, exceedingly few were experimental field studies.



Over half of the empirical studies in the archive are collecting and analyzing **digital data and behaviors** – which includes user-generated content, interactions with user-generated content (likes, shares, etc), user account information, actions taken against content or accounts.

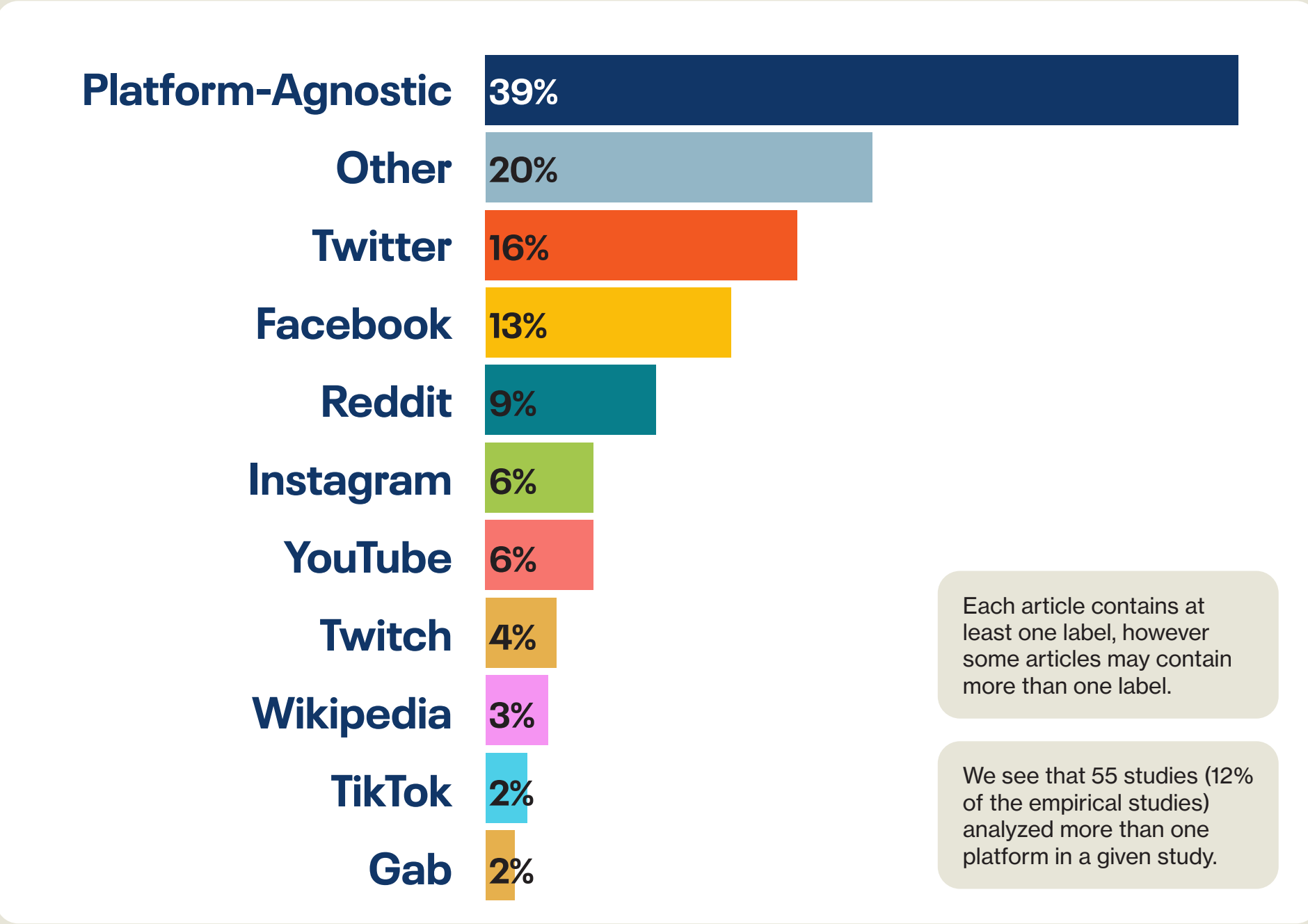


Many of the 33 studies we labeled as “Other” here collected data in the form of meeting transcripts, news articles, press statements and releases, or other text documents in an attempt to empirically analyze the public discourse around T&S topics.

About 19% (86 articles) of the empirical studies combined two or more of these data collection methods together in their study.

39% of the empirical studies in the archive are **platform-agnostic** or study **fictional platforms**. In some of these studies, researchers develop their own platform to study while in other cases researchers are using surveys to ask participants about general social media use, perceptions, behaviors, or media consumption.

Shown here are the 10 most studied platforms. **Twitter, Facebook, and Reddit** were the most frequently studied platforms. There was a long tail of less frequently studied platforms which, in total, made up 20% of the empirical studies in the archive. As many in the T&S community are aware, behaviors and phenomena that T&S research focuses on moves fluidly between and across platforms. Many have discussed **the importance of studying these issues across platforms**.



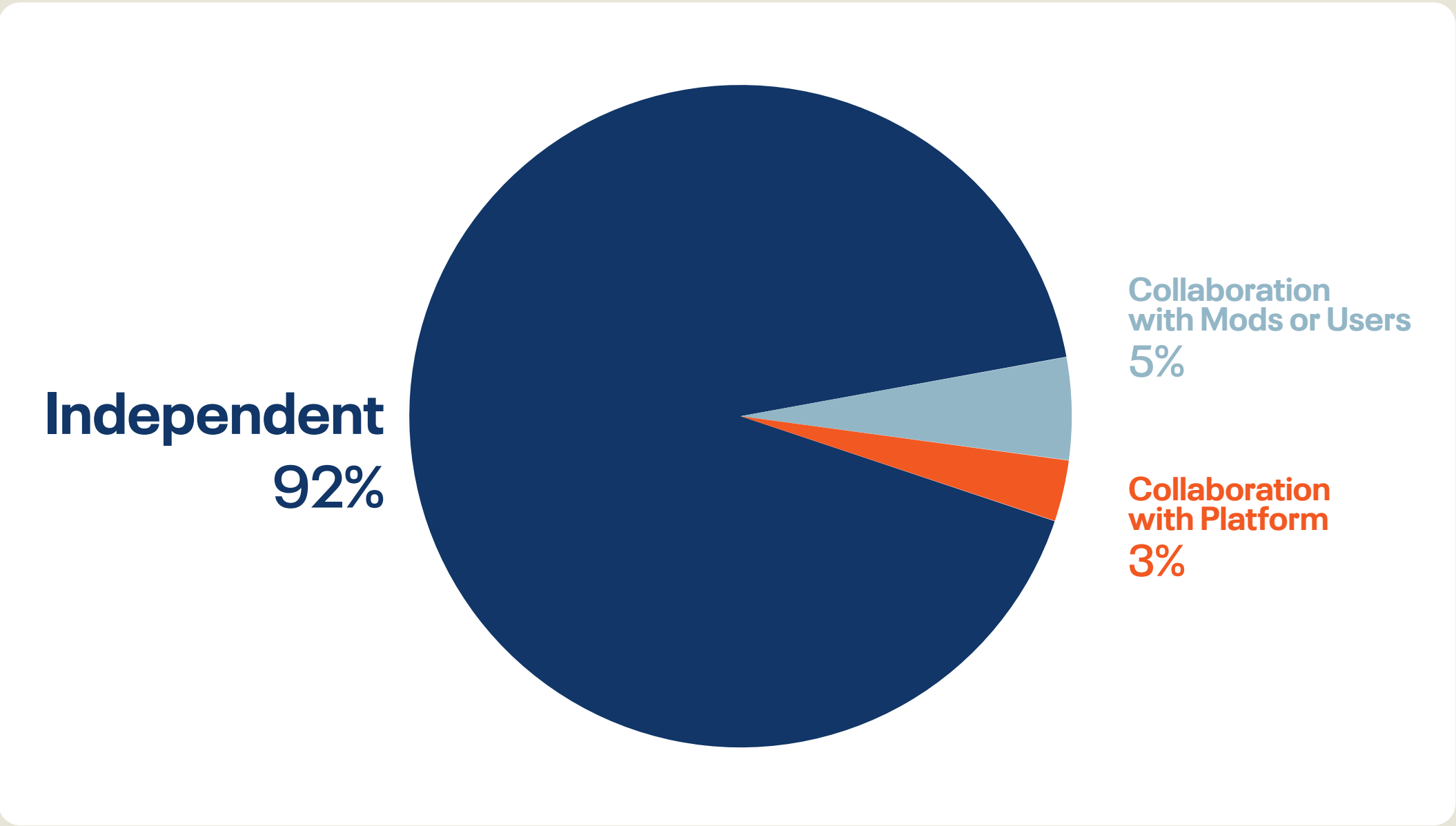
Each article contains at least one label, however some articles may contain more than one label.

We see that 55 studies (12% of the empirical studies) analyzed more than one platform in a given study.

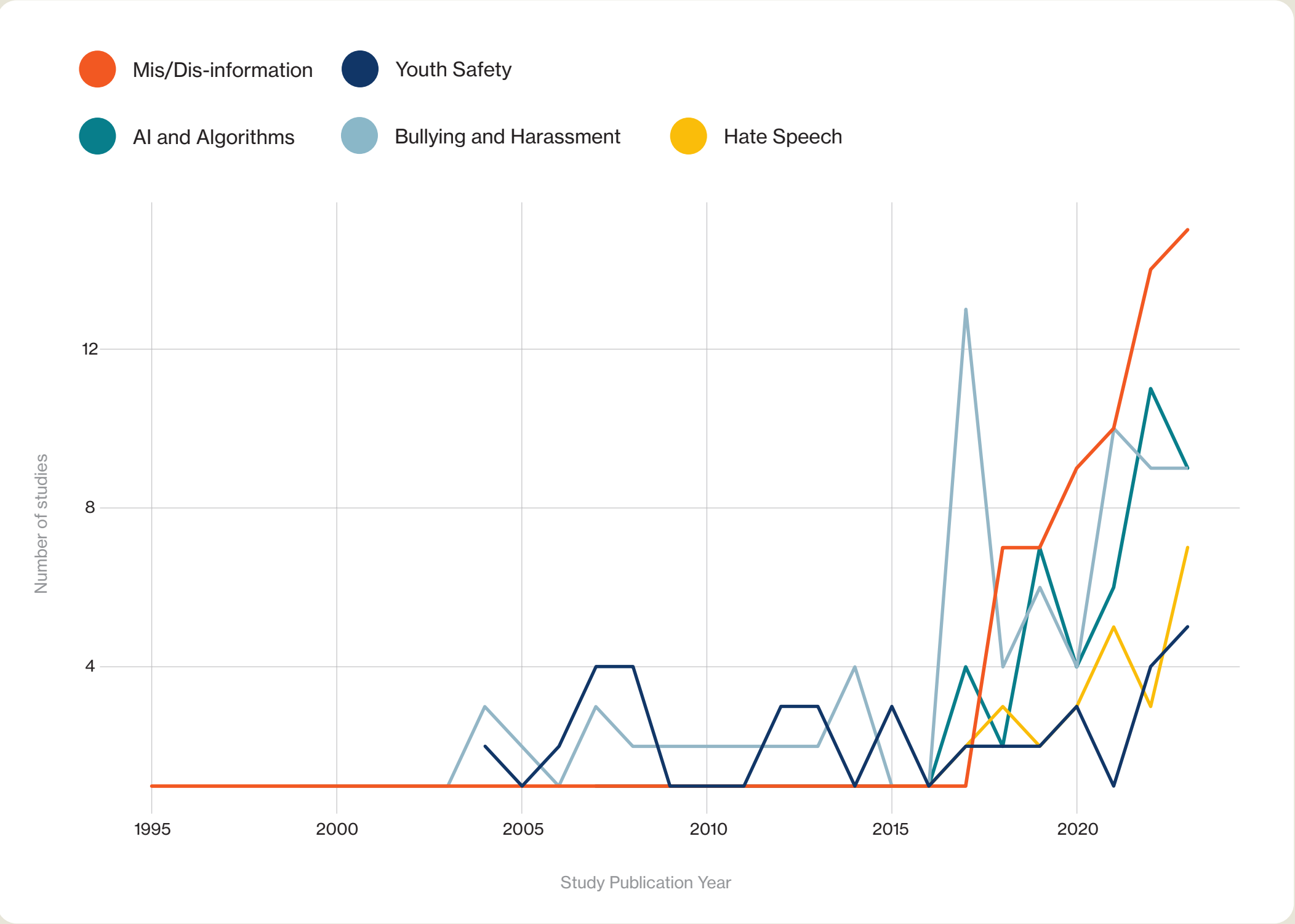
Nearly all of the empirical articles in this archive were done by **independent research teams** while **far fewer were done in collaboration or cooperation with the platform** being studied, or in **collaboration with specific groups of users**, often community moderators (aka mods).

While there are many who call for more research to be conducted independently from platforms (most notably the recent formation of the Coalition for Independent Technology Research), **we see an opportunity for more collaboration** in future T&S research.

- There is a major opportunity to **increase the frequency of involving user-stakeholders** like mods in future research.
- Additionally, we see a lot of opportunity for **more collaborations with platforms to publish research** which can increase transparency in an effort to help (re)build public trust.



About 10% of the empirical studies in the archive focus on **Youth Safety**. This topic is relatively consistent over time with just a few new studies being published each year on the topic. Conversely, studies focused on mis/disinformation are non-existent in the archive until 2018 but account for 20% of empirical articles in the archive published in 2023.



Discussion



Maintain and Expand the Archive

This archive is a terrific resource to both scholars and practitioners within the T&S community. Scholars can make use of this archive in teaching, curriculum development, and their own literature reviews.

As done with this project and recently in an article by the archive’s creators (Knittel & Menking, 2024) this archive itself can serve as a dataset to study and analyze. For practitioners, this archive can again serve to assist in training team members, but can also serve as a reliable resource to inform product and policy development as T&S teams expand into areas that may be new to them, but not to the T&S field at large.

We hope that the TSF continues to maintain and expand this resource as they have done so far (new articles have been added to the archive since our labeling project began!). We hope projects like this expand the community’s awareness of this resource and encourage more to engage with and contribute to the archive. You can access the archive and request to add to the archive on their website.



Field Studies

Through this project, we saw that there are exceedingly few experimental field studies contained in this archive – only 8 studies in an archive of nearly 1,300.

While observational and experimental lab studies are instrumental in building and testing theory as we develop our scientific understanding of a topic, it is equally important to see these how these experiments and interventions perform when deployed on platform.

The technology industry is well-known for its ability to run vast amounts of sophisticated A/B tests across their platform intended to help inform and guide product development. Specifically within T&S, many platforms often discuss “testing” new safety tools when making public announcements about these features. Given how often platforms refer to their development and testing of new safety features, it is discouraging to see how rarely that testing translates into published peer-reviewed research.



Meta Studies

As we can see, over recent years attention and scholars have been drawn to the T&S field. As this trend continues, there is increasing need for meta-studies which can tie together and summarize work across the field on a given topic.

We already see this trend starting within the existing archive and we imagine that this will continue as research in the field progresses.