COMMUNICATION METHODS AND MEASURES
https://doi.org/10.1080/19312458.2025.2605946

Routledge
Taylor & Francis Group

The platform matters: cross-platform differences in data donation willingness, behavior, and bias

Lion Wedel[a], Jakob Ohme[a], Anna-Theresa Mayer[a], Felix Gaisbauer[a], and Yangliu Fan[a,b]

[a]Weizenbaum Institute, Berlin, Germany; [b]Institute for Media and Communication Studies, Freie Universität Berlin, Berlin, Germany

ABSTRACT

Data donations are a method to access user-level digital trace data, as they provide fine-grained measures of content exposure on social media. The interest in data donation as a data collection method is accompanied by a broad uncertainty about the reasons that drive the donation of data by users. The current literature lacks comparative analysis across various platforms. This study investigates platform-specific predictors for data donation behavior of a non-probability quota sample of German social media users (N = 2,296) for YouTube, Facebook, Instagram, and TikTok and the resulting non-response biases. Based on the analysis of 340 data donation packages, we find that participants are less likely to donate TikTok data compared to the other platforms. Gender is the main driver during the willingness step for drop offs, while political leaning is a key predictor for all platforms except Facebook during the donation stage. Data donors tend to self-report less active social media usage with news and political content than those who do not donate data. Our findings highlight the importance of considering platform-specific differences in expected donation rates, biases, and the potential for discrepancies between indicated willingness and actual donation behavior when designing and interpreting data donation studies.

Introduction

With the opportunity for social media users to request the data large online platforms have harnessed and stored about them at any time, a novel data collection method for user trace data is emerging in recent years: asking participants to donate their data download packages (DDPs) that they can personally request from digital platforms. This fine-grained and nonreactive approach to accessing user trace data receives increasing interest in computational communication science (e.g., Breuer et al., 2022; Hase et al., 2024; Ohme & Araujo, 2022; Ortega et al., 2023; Otto et al., 2024). We understand data donations specifically as data acquired by participants that platform providers have collected on them, which are subsequently donated for research purposes. Other types of self-collected data (e.g., screenshot upload) and automated tracking approaches (e.g., tracking apps Reeves et al., 2021; Toth, 2023) must be distinguished from the data donation approach discussed in this paper, as they differ in key characteristics of data collection methods, such as reactivity, the ability to obtain informed consent, and the temporal scope and detail covered by the collected data (Christner et al., 2021; Otto et al., 2024; Wu-Ouyang & Chan, 2023).

CONTACT Lion Wedel lion.wedel@weizenbaum-institut.de Weizenbaum Institute, V. Hardenbergstraße 32, Berlin 10623, Germany

Supplemental data for this article can be accessed online at https://doi.org/10.1080/19312458.2025.2605946

DDPs can provide novel detail and insight into user behavior, offering a passive and scalable approach to gathering user information (Ohme et al., 2024). These fine-grained data give researchers a promising basis for applying computational methods, such as automated content analysis (Semenova et al., 2024; van Hoof et al., 2024; Wedel et al., 2024), linkage analysis (De Vreese et al., 2017; Pouwels et al., 2023), network analysis (Kahlert, 2024), and sequence analysis (Fan et al., 2025), in various research contexts, such as search behavior on political referendums (Blassnig et al., 2023), the content of direct messages on Instagram (Verbeij et al., 2024), and the context in which such traces are generated (Ortega et al., 2022).

Importantly, this data collection method is not without caveats. The process of requesting, downloading, and finally donating DDPs requires an increased effort from participants, which makes the donation rate (the fraction of eligible and consenting participants that eventually donate) less predictable than the response rates in traditional survey data collection (e.g., Hase & Haim, 2024; Pfiffner, Reiss, et al., 2024; Welbers et al., 2024). Therefore, data collected in such a way are prone to various sample and non-response biases, including undercoverage, self-selection, and compliance biases (Boeschoten et al., 2022; Bosch et al., 2024; Breuer et al., 2020).

The promise of data donations as a data base and the difficulties in its collection have received increasing attention in the research community. Research looked at predictors of participants' willingness to donate their data and of the actual donation behavior separately and how biased DDP samples are relative to the target population (e.g., Hase & Haim, 2024; Keusch et al., 2019; Strycharz et al., 2024). Previous research found that factors such as age, gender, pro-social attitude, monetary incentive, and participants' political orientation can affect users' willingness and actual donation behavior (e.g. Kmetty et al., 2024; Pfiffner & Friemel, 2023; Welbers et al., 2024).

However, two important research gaps remain. First, it is an ongoing question how the sub-sample that is willing to donate and the sub-sample that actually donates differ. Recent work on this comparative perspective is limited to findings on Dutch Facebook users showing that older age plays a role in drop-out rates at both stages of the process, while being female only impacts willingness, not actual donation behavior (Strycharz et al., 2024) and German social media users where no major differences in significant predictors between willingness and actual donation behavior were found (Hase & Haim, 2024). Another study of German Facebook users found that being female reduced willingness, but had no impact on actual donation behavior (Keusch et al., 2024). Second, to date, little is known about whether established predictors are applicable across different platforms. Specifically, for the willingness to donate, evidence shows that participants are more willing to donate YouTube data compared to Google, Instagram, and Facebook (Pfiffner & Friemel, 2023). In addition, comparative findings on differences between data sources (health app, sensor, Facebook, etc.) or data sharing forms (sharing username, web tracking, etc.) beyond DDPs point to significant differences in biases and user behavior depending on the requested platforms (Beuthner et al., 2024; Silber et al., 2022). However, a systematic analysis of predictors of donation behavior for different social media platforms is still missing. Recent work on this focuses on one platform (Keusch et al., 2019; Strycharz et al., 2024) or does not differentiate between different platforms comparatively (e.g., Hase & Haim, 2024; Welbers et al., 2024). The lack of research on cross-platform donation behavior is problematic because understanding users' comprehensive social media behavior often requires considering their activity across multiple platforms, similar to how a single-device approach may not fully capture a user's digital footprint (Bosch et al., 2024).

This paper addresses both research gaps. Based on a German quota sample, we compare the predictors of willingness and actual donation behavior for YouTube, Facebook, Instagram, and TikTok users and the resulting non-response biases, giving future studies a more comprehensive basis to understand the potential cross-platform differences between drivers for data donation willingness and actual donation behavior. It also unravels the opportunity to improve data donation procedures if we understand who the users are that are generally willing to donate but do not donate in the end. Our findings show that it is a fallacy to assume similar biases for donor samples, regardless of

the platform for which they are asked to donate. We find differences between platforms and donation stages with respect to age, gender, use of social media for news and political information, and political leaning.

In the following, we begin by establishing our data collection pipeline and the relevant terminology, followed by a review of previous research to establish our research questions. We then present our methods and results before concluding with a discussion of our findings.

The DDP collection pipeline

The DDP collection pipeline commonly follows a similar set of steps and was depicted for different purposes in previous work (Boeschoten et al., 2022; Strycharz et al., 2024). This study focuses specifically on the collection of data donations facilitated by a non-probability online panel provider and provides in this section a description of such data collection pipeline for contextualization and transparency. Although there are other recruitment strategies, access panels remain the most robust approach (Pfiffner, Reiss, et al., 2024).

We distinguish the DDP collection pipeline in this paper broadly into four phases (see Figure 1). Beginning with phase one, for which the panel provider invites their A) *panel members* who upon an initial informed consent procedure enter the survey as B) *consenting participants*. Only participants that are still needed to fulfill the set quotas (e.g., age, gender, education) proceed with phase two, the survey. For this study, only those respondents who complete the survey and who use at least one of the platforms for which we collect DDPs are C) *eligible respondents*. For participants who complete phase one and phase two, we have survey data to explore the differences of the collected sample in terms of what predicts participants to move on or drop out. In this pipeline, the survey is placed before the data donation. We designed it this way because a dataset without survey responses would be of limited use to analyze predictors for data donation willingness and data donation behavior, as well as resulting

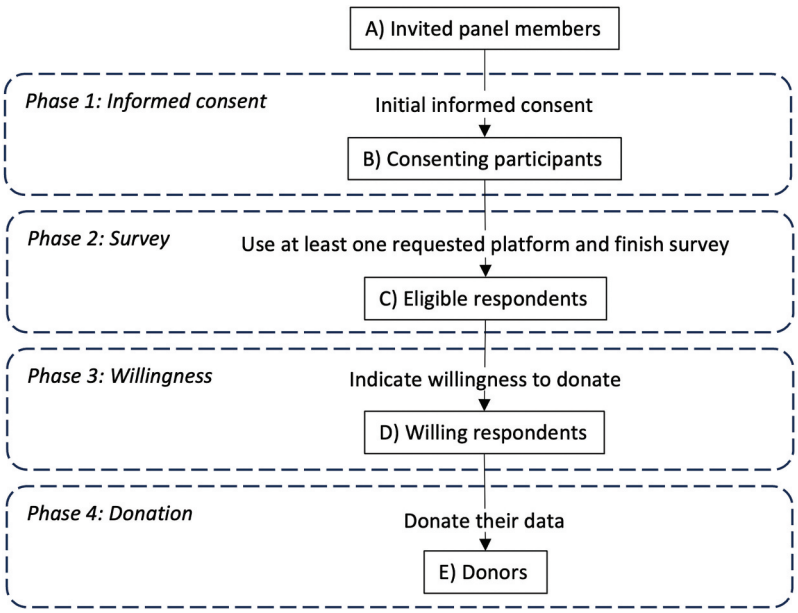

Figure 1. Flowchart of the data donation collection pipeline based on a non-probabiltiy sample and a survey prior to the data donation. Own depiction adapting previous work to our study specifics (Boeschoten et al., 2022; Strycharz et al., 2024). Depicted are four crucial phases from informed consent until the DDP donation. We depict the sample at each point in the process in rectangles and the condition that respondents need to fulfill to advance in the flow in between.

non-response biases. Front-loading the survey also helped us logistically in a cross-platform data donation collection by allowing us to pre-filter users for the platforms they use to ensure that we only presented them with relevant options during the data donation steps, avoiding a confusing and overwhelming experience (Carrière et al., 2025).

After finishing the survey, participants decide in the third phase (willingness) if they want to continue with the data donation. Participants who continue in this phase are considered D) *willing respondents* (see also Keusch et al., 2024). Other descriptions of this subset of respondents have been *respondents after finishing the survey* (Strycharz et al., 2024; Welbers et al., 2024) or *respondents with the intention of donating* (Hase et al., 2024). We follow the terminology of willingness here to provide an alignment in interpretation with the so-called *willingness* or *vignette* studies (Gilbert et al., 2021; Kmetty et al., 2024; Pfiffner, Reiss, et al., 2024). Willingness in these studies refers to the participants' intention to continue with a hypothetical data donation after an initial informed consent and survey. Since these studies are survey-only, they offer a cost-effective way to assess expected data donation behavior when exploring a new method, prompting participants with various scenarios (vignettes) varying for platforms, types of traces, and question phrasing which would not be suitable in the context of an actual data donation study. Finally, we reach the donation phase. Here, participants are typically referred to a respective platform such as *Port* (Boeschoten et al., 2023) or the *Data Donation Module* (DDM) (Pfiffner, Witlox, et al., 2024). At this stage, participants must request and subsequently donate their data, which results in the final sample of E) *donors*.

This paper focuses on the characteristics of participants who drop out during the willingness phase and the donation phase. We also look at the non-response bias that results from the data collection process – in other words, we consider systematic differences in demographics and behavior characteristics of the final sample of E) donors vs. C) the sample of *eligible respondents*.

Background and research questions

Data donation willingness

The *willingness rate*, as the fraction of participants who agree to continue with the data donation after finishing an initial survey (reported in percentages), ranged in previous studies on German samples from 52% (Hase & Haim, 2024; aggregated across Facebook, YouTube, Instagram and Twitter, participation in the lottery for a 50 euro gift card) to 79% (Keusch et al., 2024; Facebook with a 2 euro incentive). For a study on Facebook DDP collection among Dutch panel members, 52% willingness rate was reported (Strycharz et al., 2024; no incentive reported). Given these previous results, we expect willingness rates around 50%. However, specifically the results from (Keusch et al., 2024) point to possible variations in willingness rates between platforms.

Importantly, the existing results do not provide a between-platform perspective with consistent national context and sampling strategy. This paper closes this gap by providing findings on the following question:

> **RQ1:** What are the willingness rates to donate data download packages of German social media users for Facebook, Instagram, YouTube, and TikTok?

In addition to bare willingness rates, researchers are interested in predictors that drive this intention. It is crucial to understand the reasons for participants' decision making throughout the data collection process to improve study designs, or at the very least to be cautious about potential biases introduced during the willingness phase. Previous vignettes and donation studies provide a basis for understanding willingness behaviors based on sociodemografics (Hase & Haim, 2024; Kmetty et al., 2024; Strycharz et al., 2024). However, fragmented national contexts, platforms, and sampling approaches make it challenging for researchers to draw conclusions for a single country on different platforms that can inform their research designs. In the German

context, for example, a previous study found that female participants were less likely to indicate willingness to donate (Keusch et al., 2024). However, the same predictor was not significant when reported on social media users without differentiating by platform (Hase & Haim, 2024). In the Dutch context, the role of gender is also unclear. For Google users (Chrome and YouTube), gender (f) showed a negative association (Welbers et al., 2024), but for Facebook users, none at all (Strycharz et al., 2024). Education was reported to not affect the willingness to donate data (Hase & Haim, 2024; Keusch et al., 2024; Strycharz et al., 2024) with the exception of the willingness to donate Chrome and YouTube DDPs by a Dutch online panel (Welbers et al., 2024). Age was reported to be not predictive of the willingness to donate data for German Facebook users and Dutch YouTube users (Keusch et al., 2024; Welbers et al., 2024), nor for the platform-aggregated donation willingness (Hase & Haim, 2024). However, for Dutch Facebook users, lower age meant higher willingness (Strycharz et al., 2024).

Beyond sociodemographic variables, previous results are even more scattered, since existing studies either do not specifically report other predictors due to their research design (Strycharz et al., 2024; Welbers et al., 2024) or include different sets of additional variables due to their study focus: Political leaning (right) was shown to be negatively connected to willingness in cross-platform German samples, while privacy concerns were only predictive of a German non-quota convenience sample, but not for other online panels (Hase & Haim, 2024; Keusch et al., 2024; Kmetty et al., 2024).

Given the different sampling strategies and predictor sets, the results are hardly comparable across platforms. However, comparable numbers would be invaluable for future research projects to inform sampling (e.g., quotas) decisions depending on the platform under scrutiny. The present study aims to provide such numbers for Germany's four most used social media platforms (Behre et al., 2023) and to answer the corresponding research question:

> **RQ2:** What predicts the willingness to donate data download packages of German social media users and how do findings vary between Facebook, YouTube, Instagram, and TikTok?

Data donation behavior

The data donation step results in notoriously high dropout rates despite prior informed consent and indication of willingness (e.g., Hase et al., 2024; Strycharz et al., 2024; Welbers et al., 2024). On the one hand, this can be explained due to the increased effort participants need to invest in the donation itself (see Groot Kormelink et al., 2025 for detailed descriptions of the challenges). On the other hand, people might have second thoughts about giving their data to research when they are confronted with what they are about to donate – certain beliefs or fears might be activated at this stage in the process, due to being uncomfortable with specific traces (Boeschoten et al., 2022; Groot Kormelink et al., 2025). The *donation rates* (the fraction of willing participants who donate their data) reported as percentages in past research range from 11.2% (Google Takeout, Welbers et al., 2024) and 24% (Facebook, Instagram, Twitter, and YouTube accumulative; Hase & Haim, 2024) to 28% (Facebook, Strycharz et al., 2024). In contrast, a study of German Facebook users reported a donation rate of 48% (Keusch et al., 2024). Studies with small sample sizes, that could afford to personally assist each participant's donation process even reached 69% for Instagram (van Driel et al., 2022) or 100% for 20 participants in a cross-platform donation setting (Groot Kormelink et al., 2025). Given the sample size of our study that does not allow such a setup, the expected donation rate for our study setup is therefore between 11 and 24%.

The varying sample strategies and national contexts make a cross-platform comparison, similar to willingness rates, difficult or even unfeasible. Consequently, this paper closes this gap for the following research question:

RQ3: What are the donation rates for data download packages of German social media users for Facebook, Instagram, YouTube, and TikTok?

In addition, we expand the evidence for multi-platform donation behavior. Previous insights based on student and online non-quota convenience samples showed that: "78% of the donors provided information for a single platform, while 16% donated data for two, 4% for three, and 2% for four" (Hase & Haim, 2024, p. 16). To provide comparable evidence from an online non-probability quota panel, we ask:

RQ4: What are the multi-platform donation rates for data download packages of German social media users for Facebook, Instagram, YouTube, and TikTok?

Furthermore, current research still lacks clarity on the platform differences in predictors of actual data donation behavior, and altogether for newer platforms such as TikTok. For example, being older led to a higher chance of dropping out of a data donation study for Facebook and Google users in the Netherlands (Strycharz et al., 2024; Welbers et al., 2024). However, it did not have an impact on the donation behavior of a German sample across platforms (Hase & Haim, 2024). Gender showed a significant correlation for Facebook users in the Netherlands, but not for Google users in the same country, nor in the case of social media users in Germany (Hase & Haim, 2024; Strycharz et al., 2024; Welbers et al., 2024). Higher education was associated with data donation for Google and Facebook users in the Netherlands, but not in Germany (Hase & Haim, 2024; Keusch et al., 2024; Strycharz et al., 2024; Welbers et al., 2024). Political orientation (right) reduced the donation behavior in Dutch and German samples alike (Hase & Haim, 2024; Welbers et al., 2024). Similarly to willingness predictors, beyond the above, evidence on other predictors (trust, privacy concerns, and platform use) is scattered, depending on research designs and study foci.

Current research only allows us to derive the impact of users' behavior for specific combinations of sampling, national context, and platform – or on accumulative analysis of different platforms (e.g., Hase & Haim, 2024). A within-country comparison of the predictors of data donation behavior across platforms is still lacking. Consequently, this paper seeks to provide much needed knowledge on the data donation behavior of German social media users for under-researched platforms in the German context (TikTok, YouTube, Instagram), and between platforms.

RQ5: What predicts the data donation of data download packages of German social media users and how do findings vary between Facebook, YouTube, Instagram and TikTok?

Data download package collection and non-response bias

The assumption that specific user characteristics influence the willingness and donation behavior consequently leads to the expectation that data donation samples are biased (Gil-López et al., 2023; Hase & Haim, 2024; Strycharz et al., 2024). Such biases can be introduced at various steps of the DDP data collection process (Boeschoten et al., 2022).

Past research has differentiated between different biases along the way of data donation collection, such as *self-selection biases* that can arise, among others, from certain panel members being systematically unwilling to participate in a study (Hase & Haim, 2024). Furthermore, when consenting and willing participants are asked to follow up on their indicated willingness, compliance errors – which may occur due to technical difficulties, a change of mind, or other factors – can lead to a *compliance bias* for the final sample from which the researchers retrieve the DDPs (Boeschoten et al., 2022). When systematic dropouts of participants who decide not to donate when confronted with their DDP and the data therein occur, we talk about *consent bias* (Hase & Haim, 2024). However, in the present study, our technical setup does not allow us to distinguish between the specific reasons for dropping out, nor

for the exploration of self-selection biases (we only have data on participants that finished the survey). Therefore, for this work, we summarize the various biases that emerge throughout the data donation collection process as *non-response bias* in alignment with previous research on the matter (Hase & Haim, 2024). This bias refers to systematic differences between the sample of respondents who successfully made a data donation and those who completed the initial informed consent and survey stage and reported using the respective platform (eligible participants).

Past work that distinguished between self-selection and compliance biases has shown that participants become successively younger in each data collection stage (non-quota online sample of Facebook users, Strycharz et al., 2024). Furthermore, a self-selection bias was found for female participants, while location (urban vs. rural) and level of education were not biased. For an online non-quota convenience sample, non-response biases were found: donor samples were politically left, showed increased awareness of algorithmic filtering, had less privacy concerns, and had higher technical skills (Hase & Haim, 2024). Understanding the biases present in data donation samples is essential for exploring methods to counter such biases (Hase & Haim, 2024; Pak et al., 2022). We explore the non-response bias of a German population quota sample across and between platforms:

> **RQ6:** What non-response biases are introduced throughout the collection of data download packages of German social media users and how do these vary between Facebook, YouTube, Instagram and TikTok?

Method

To answer our research questions, we collected survey data of social media users' interaction with news and political information online and DDPs across four different social media platforms. Data were collected from August 2024 to September 2024 in collaboration with a non-probability online panel provider (Bilendi), as previous work has shown that panel recruitment yields the best willingness and donations rates to carry out DDP data collection efforts (Pfiffner, Reiss, et al., 2024). Bilendi is one of the largest non-probability online panel providers in Germany with up to 316,000 panelists that can be invited based on gender, education, and age quotas.

Before starting data collection, ethical approval was obtained from the Research Ethics Committee of the Weizenbaum Institute. All participants provided their informed consent prior to participating in the study. The data underlying this study are publicly accessible for secondary use (Wedel et al., 2026).

Sample

To ensure a sufficient prevalence of platform users for each of the four platforms of interest, our sampling strategy targeted a quota sample of 2500 German Internet users. This comprised a primary sample of 2000 to meet national quotas for age, gender, and education, supplemented by a deliberate oversampling with 500 younger adults (18–27 years) to provide a robust basis for analyzing this a priori underrepresented demographic. 6697 respondents started the initial survey, 5051 passed the informed consent, 2883 of those passed the screening questions, and 2564 participants passed the two attention checks (e.g., asking for a specific answer in a question battery) built into the survey. 2457 participants completed the survey.

Due to missing data for age and location variables, non-binary responses for the gender variable, and the lack of using any of the requested platforms, a subset of 2296 users (93.49% of participants) were eligible for subsequent analysis. The sample satisfies the target quotas, including an oversampling of young adults of 16.37% (see Table 1).

Table 1. Comparison of population quotas, target quotas, and achieved sample quotas for the survey. Age and gender quotas were provided by the panel provider (Bilendi), and the education quota is based on data from the German Federal statistical office (Statistisches Bundesamt, 2022). Target quotas for age deviate from population quotas due to a conscious oversampling of the 18–27 age group. Columns may not sum to 100 due to rounding.

Variable	Values	Population (%)	Target (%)	Sample (%; n = 2296)
Age	18–27	16.63	33.00	33.80
	28–37	19.93	15.94	16.16
	38–47	18.75	15.00	15.59
	48–57	22.89	18.31	18.21
	58 and older	21.80	17.44	16.25
Education (ISCED)	high	30.50		30.00
	mid	54.80		55.55
	low	14.70		14.45
Gender	male	50.57		50.09
	female	49.42		49.91

Study design

Participants were recruited through the panel provider and informed about the study characteristics. Participants were incentivized to complete the survey according to the standard rate of the panel provider, regardless of their decision to donate data. However, to encourage data donation, participants were informed that they would receive an additional 5€ for their first data donation and 3€ for each subsequent donation.

Following informed consent, participants completed a 20-minute online survey that included questions on demographics, social media use, and other topics relevant to the larger research project. At the end of the survey, participants were asked to indicate their willingness to donate data from their social media accounts (YouTube, Facebook, Instagram and TikTok). These platforms were selected due to their high prevalence of use in Germany (excluding messaging apps; Behre et al., 2023). Participants were only given the option to donate data from platforms they had indicated using at least once in the past month, displayed in a random order of items in the questionnaire (see Appendix A).

Participants who expressed willingness to donate data for at least one platform were redirected to the PORT data donation platform (Boeschoten et al., 2023). PORT was chosen for its ability to 1) automatically filter raw data download packages (DDPs) upon upload to ensure compliance with data minimization principles and 2) provide participants with a transparent view of their data and the option to remove specific elements before donation, thus facilitating an additional form of informed consent and data transparency (Boeschoten et al., 2023). Detailed instructions on how to request, download, and donate their data via PORT were provided to the participants. Importantly, participants could view and selectively remove any data from their DDP before submitting the donation. Only at this point the data to be donated and in case of multiple donations the differences across platforms became truely visible to participants. The requested data included basic consumption data (e.g., watch history, likes, shares), as well as more sensitive data such as follower/following lists, which users are typically less willing to share (Pfiffner & Friemel, 2023). Appendix B provides a complete list of the data extracted for each platform.

Predictors and measurements

For each participant, we collected socio-demographic information (age, gender, education, location), the self-reported frequency of different social media activities in relation to news and political information, self-reported perceived freedom of expression online, political interest and political leaning. Given that all participants in the final sample passed two attention checks during the survey, we did not include the information on rapid survey completions used in previous studies (Welbers

et al., 2024). The choice for those predictors resulted from the combination of research interests and research economic considerations to include understudied predictors in a larger data collection effort.

Before discussing the included predictors and their operationalization in detail, we want to highlight an intuitive confounder that we did not measure: *privacy concerns*. Previous studies on non-probability online panels did not find a statistical connection between privacy concerns and data donation behavior (Keusch et al., 2024; Kmetty et al., 2024). An exception is a study of two convenience samples without quotas (Hase & Haim, 2024). Given the resource constraints and the empirical evidence from similar panels, we chose to exclude the measure of privacy concerns. However, we acknowledge that this decision is a limitation of our study. It is plausible that individuals who join non-probability online panels have, on average, lower privacy concerns than the general population. This potential self-selection bias could mask a true relationship between privacy concerns and data donation, and as such, our findings may not be generalizable to individuals who do not participate in such panels.

Socio-demographic variables

Four sociodemographic variables were included since they have been shown to play a role in data donation behavior in multiple studies with participants from different countries, including Germany, and across platforms (e.g., Keusch et al., 2024; Kmetty et al., 2024; Pfiffner & Friemel, 2023; Strycharz et al., 2024).

Age was measured on a numerical scale ($M = 39.81$; $SD = 14.89$). Only participants older than 18 years of age could participate. For *gender*, 50.09% ($n = 1{,}150$) indicated male, 49.91% ($n = 1{,}146$) indicated female. For *education*, the initial fine-grained survey question (What is your highest level of school or university education?) was converted to three levels based on the International Standard Classification of Education (ISCED) for the German education system (Statistisches Bundesamt, 2022): 30% reported low education (primary education; $n = 337$), 55.55% medium (secondary education; $n = 1.272$), and 14.45% high (tertiary education; $n = 687$) (see Table 1). For *location*, we assigned the measure *RegioStaR7* to each participant, for which we derived their commune from the coordinates provided by the survey platform (Qualtrics). *RegioStaR7* is a measurement developed by the German Federal Ministry of Digital and Transport. This measure considers the role of an area at the regional, local, and small-scale levels. Various spatial-structural statistics are considered, leading to an ordinal typology from rural to metropolitan (Sigismund, 2021). It was implemented as a seven-point scale starting at zero, with 0 describing the least urbanized areas in Germany ($M = 4.11$; $SD = 2.10$).

Social media engagement with news and politics

In addition, in this study, we focus on how users interact with news and political information on social media. Past research on German social media users has shown that the frequency of general social media use predicts whether people will donate their data across four platforms (Hase & Haim, 2024), but this was not the case for Facebook specifically (Keusch et al., 2024). However, if the frequent use of social media itself is a confounding variable in who participates in data donation, then the engagement with political content and news can be another important predictor. News and political content are highly revealing types of content which are closely related to another confounder (political orientation, e.g., Hase & Haim, 2024; Welbers et al., 2024) in data donation behavior. Therefore, it is essential to understand this relationship to properly plan and execute future data donation studies on the use of news and political information. Such studies could extend knowledge on political information and the use of news already reported by browser trace data (Loecherbach et al., 2024; Wojcieszak et al., 2024). Therefore, in the present study, we introduce a set of social media engagement measures.

A battery of items with a 7-point frequency scale (Multiple times daily – never) was presented to the participants: *If you inform yourself or others on platforms about news and political topics, how often are you . . .* , with seven different activities: 1.) *watching/seeing*; 2.) *posting*; 3.) *reacting* (likes, emojis); 4.) *commenting*; 5.) *discussing publicly*; 6.) *sharing publicly*; 7.) *sharing privately*. The scale for all activities was inverted for the analysis to reflect increased frequency with higher values. A subsequent

correlation and factor analysis resulted in the definition of two factors and two separate items in partial alignment with the constructs of high-effort and low-effort engagement (Heiss et al., 2019; Xenos et al., 2017, see Appendix C). The factor scores for high-effort and low-effort engagement were extracted and standardized with a mean of 0 and a standard deviation of 1 based on the results of all the survey participants ($n = 2{,}447$), not just the eligible, since for the factor score calculation the full power of the survey was exploited.

- *high-effort engagement* ($M = 0.02$; $SD = 0.97$, $min = -1.11$, $max = 2.24$): This factor captures the frequent engagement with news and political content through activities that require high effort from users on social media platforms. It is measured by the respective factor scores and includes the activities posting, commenting, and discussing publicly.
- *low-effort engagement* ($M = 0.03$; $SD = 0.96$, $min = -1.33$, $max = 1.80$): This factor captures the frequent engagement with news and political content through activities that require low effort from users on social media platforms, including exposure to content and basic reactions to it. It is measured by the respective factor scores.

Additionally, factor analysis (EFA and CFA, see Appendix C) showed that neither *public nor private sharing* fit well in either factor, nor do they form a coherent factor on their own due to being in parts conceptually opposed (public vs. private). Nevertheless, they offer an additionally level of insight into data donation behavior, since one could assume that frequent private sharing compared to public sharing make the social media data more sensitive to donors, and therefore impact data donation behavior. Therefore, both are kept as predictors and variables for predictive and bias analysis as a frequency scale (0–6; *public sharing*: $M = 2.01$; $SD = 1.87$; *private sharing*: $M = 2.17$; $SD = 1.85$).

Political attitudes

Political leaning was a strong predictor of donation behavior and non-response bias in previous studies, the same is true for political interest (Hase & Haim, 2024). Therefore, we included both, specifically since they are conceptually connected to social media engagement with political information and news. The latter is also the reason for the introduction of the variable perceived freedom of expression. With that we cover not just engagement with politically information online, political leaning and interest, but additionally the perceived comfort with sharing content which is perceived as contrary to the mainstream.

Perceived freedom of expression was measured by the agreement of participants with the statement "It is easy for me to express my opinion on platforms, even if I think many people might disagree with me." on a scale of zero to four ($M = 2.20$; $SD = 1.29$). 0 indicates no agreement at all and 4 indicates full agreement. *Political interest* was measured on a 10-point scale, with zero indicating no interest at all and nine being very interested ($M = 5.76$; $SD = 2.66$). *Political leaning* was measured on a 10-point scale (zero to nine) from left to right ($M = 4.44$; $SD = 1.97$).

Data donation behavior

For the target variables, first we consider whether participants indicated willingness to donate data for a platform based on the corresponding end of survey question: "For which social media platforms would you like to donate your data?" and second whether they donated data for the platform, measured based on the successful completion of a donation for a platform. Both target variables are binary.

However, a data donation can contain different amounts of data depending on user behavior (e.g., no shares, only few likes). Since participants are confronted with the data they are about to donate during the donation process, the richness of ones DDP could also impact the behavior. We account for this issue by conducting our analysis on two different criteria for completed donation: 1) we considered all donations that where non-empty as successful donations, no matter how sparse. 2) We tested whether results change if for a donation to be considered successful, it must include at least one entry per trace for at least 50% of the requested traces (e.g., like history, watch history, comments) for the respective platform. In theory, this requirement ensures that all participants counted as donors

were exposed to a sufficient variety of traces during donation. Notably, because of this criterion for this condition, 61 additional donations where treated as no donations. However, between the two different approaches to count a successful donation, results only marginally differed. We therefore report throughout the paper on the results based on a successfully data donation being any non-empty donation, but added the results for the condition, only counting non-sparse DDPs to the appendix (see Appendix G and H) and address the differences and their implications in the corresponding results sections and in the discussion.

Study design confounders

Although we could control that the answer options for the willingness question are proposed in random order to participants, this was not possible for the options in the donation interface (PORT). For each possible combination of platforms participants could indicated donation willingness for, we set up a dedicated PORT instance – where the order of that platform combination is static. To ensure that we do not miss possible study design confounders, we performed a descriptive analysis of willingness rates and donation rates based on self-reported platforms used by participants and the donation interface to which they were assigned based on their responses during the willingness question.

The analysis on the willingness stage (see Appendix D) shows higher rates of willingness if participants face more options (i.e., the number of self-reported platforms used). Therefore, we include the *number of platforms used* (TikTok, YouTube, Facebook, Instagram; $M = 2.48$; $SD = 1.03$) as a possible confounder in the predictive models for the analysis of willingness decisions.

For the impact of the donation interface (i.e., the order of platforms), donation rates do not appear to vary systematically (see Appendix E). Nevertheless, the static ordering of each platform combination may still exert a subtle influence on platform-specific donation rates (Beuthner et al., 2024). Therefore, we include the *platform position* for each user as a confounding variable in the predictive models of actual donation behavior. Importantly, this variable is not included in the TikTok model. Because TikTok was always displayed in position 1, the platform-position variable has no variation (it is constant). A predictor with zero variance cannot be estimated in a linear regression model, as it provides no information for explaining differences in the outcome.

Analysis strategy

To address our research questions about the willingness to donate data and actual donation rates (RQ1, RQ3, RQ4), we present two key sets of findings. First, we report on the willingness rates. This metric is defined as the fraction of eligible participants (those who completed the survey and used at least one of the four platforms of interest) who indicated willingness to donate. Second, we report actual donation rates, calculated as the fraction of participants who donated out of all eligible participants. All rates are reported in aggregation and broken down by platform.

For the predictive analysis, we used multiple logistic regression models to explore the predictors of platform users' decisions (RQ2, RQ5). We computed a separate model for each combination of platform (Facebook, Instagram, YouTube, TikTok) and stage (i.e. willingness and donation) with the same predictors. By building a separate model for each platform's user base, we effectively control for platform-specific differences in user behavior and demographics. Each model analyzes the willingness and donation behavior within that platform's user group, comparing those who proceeded with the data donation to those who did not. This approach isolates the predictive power of our variables on a per-platform basis and means that we control for platform use as a confounding variable through the models themselves. The comparative analysis is achieved by examining and contrasting the results of each separate model. Within each model, only the dependent variable changed with respect to which platform they refer to, while all other specifications remain the same. Results are reported as average marginal effects (AMEs) in full percentage points to facilitate a more intuitive interpretation. AMEs express how much the predicted probability of the outcome changes, on average,

when a predictor increases by one unit while holding all other variables constant. For example, if the AME for the binary variable gender (female) equals 5%, this indicates that women have, on average, a 5-percentage-point higher probability of donating compared to men. Similarly, if the AME for a continuous variable such as age is −1%, this means that each additional year of age is associated with a 1-percentage-point decrease in the probability of donating.

To explore the non-response bias for each platform (RQ6), we followed an established methodological pipeline using z-tests to compare the proportions of the respective variables between the two samples (Hase & Haim, 2024; Keusch et al., 2019; Struminskaya et al., 2021). Specifically, this test quantifies non-response bias by calculating the percentage point difference in a specific characteristic (e.g., gender (f), political orientation) between the sample of eligible participants and the subset of those who actually donated data. In addition to evaluating for statistical significance using p-values, we report on *practical significance* by examining whether the 95% confidence intervals maintained a consistent sign. We did this because, in our analysis, we sought to attribute significance to biases that, despite not achieving statistical significance, we argue have a meaningful real-world impact (Kirk, 1996). For this study, practical significance was considered supported when the entire confidence interval was consistently positive or consistently negative, as this indicates a meaningful and interpretable bias direction. For this part of the analysis, age (5 age groups), education (dummy variables for low and high), gender (female as dummy variable), location (rural vs. urbanized, see Sigismund, 2021), high and low effort engagement (above median), public and private sharing (high, upper three point on the 7-point scale), perceived freedom of expression (high, upper two points on the 5-point scale), political interest (high, upper three points on the 10-point scale) and political leaning (left-leaning as scores from 0 to 4 and right-leaning the remainder) were converted to binary variables.

Importantly, similar to previous research on data donation behavior (Hase & Haim, 2024; Struminskaya et al., 2021), we combine predictive (multivariate regression) and bias analysis to address two different sets of research questions. This dual approach allows us to: 1) identify the factors that predict an individual's willingness or success in donating data, and 2) assess and understand the potential biases resulting from this selective participation. A key distinction between these methods is their analytical accordance: while the multivariate regression models control for potential confounders to isolate the true drivers of behavior, the bivariate bias analysis reveals the unadjusted differences in our sample. For example, a gender bias found through bivariate analysis may be due to other underlying user characteristics (e.g., political leaning) that are more prevalent in one gender than in another, which the predictive model can help uncover. In contrast, a significant predictor in either donation phase might not lead to a biased sample due to other effects at play at other stages that negotiate this potential bias.

Results

Donation willingness and behavior rates

First, we report answers for research questions on willingness (**RQ1**) and (multi-platform) donation (**RQ3**, **RQ4**) rates across the platforms. Of the 2296 eligible participants, 51.18% ($n = 1{,}175$) were willing to donate and 28.94% ($n = 340$) of those willing ended up successfully donating at least one non-empty DDP. In total, we initially retrieved 503 complete data donations from 340 unique users (*donation rate* = 14.81%, against the number of eligible participants). The donating participants donated on average 1.48 DPPs: 9.28% ($n = 213$) of eligible users donated exactly one DDP, 2.57% ($n = 59$) donated two DDPs, 1.26% ($n = 29$) donated three DDPs and 0.35% ($n = 8$) donated DDPs for all platforms of interest.

Additionally, in Figure 2 we report the total number of users and the respective fractions against all survey participants who were (not) eligible, (not) willing, did (not) donate and donated empty data. Of interest here is that over six percent of the participants in an online panel indicated that they did not use either of the four requested platforms, and despite the absence of communicating specific rules for

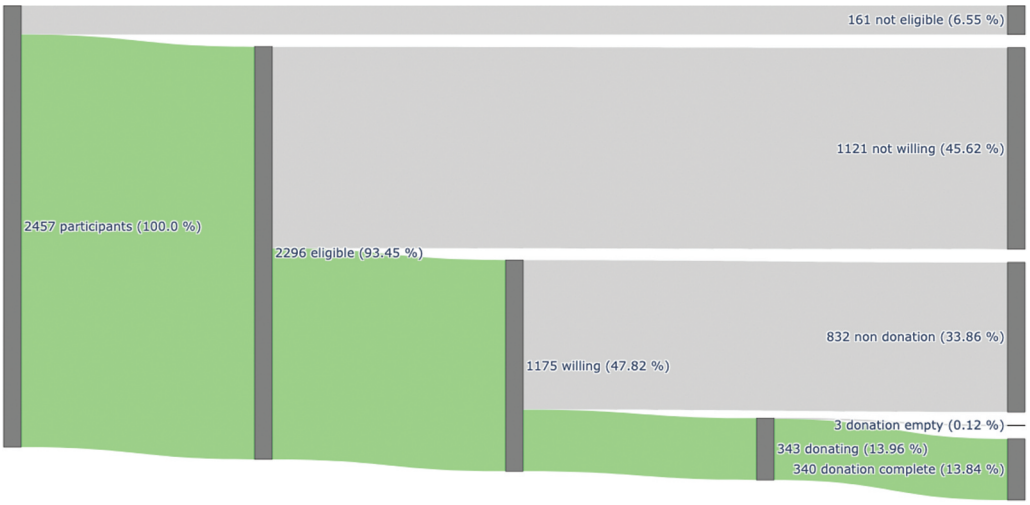

Figure 2. Donation success across the different stages of the data collection process. Eligible participants are those who use at least one of the four platforms of interest (YouTube, Facebook, Instagram, or TikTok). Percentages in parentheses refer to the total number of onboarded participants.

Table 2. Platform-level willingness and donation rates. Users may be active on multiple platforms, and rates are calculated relative to the number of eligible users per platform.

Platform	Eligible users	Willing users (%)	Donating users (%)	Donating users with filter (%)
Facebook	1,503	502 (33.34)	131 (8.72)	107 (7.12)
YouTube	1,799	591 (32.85)	174 (9.67)	165 (9.17)
Instagram	1,565	506 (32.33)	149 (9.52)	127 (8.12)
TikTok	828	250 (30.19)	54 (6.52)	51 (6.16)
Total unique users	2,296	937 (40.81)	340 (14.80)	309 (13.46)

what we count as a complete DDPs (e.g. minimum number of entries, specific time constraints; see Bechmann et al., 2025) only 2 donations were completely empty.

We also consider platform-by-platform willingness and donation rates (see Table 2). Here, we report the number of willing and donating users compared to the number of eligible participants in our sample, since platform usage is typically a screening question in related studies. Therefore, the total rates slightly diverge from Figure 2, which reports the rates in relation to all survey participants, not just eligible participants. In addition to being the platform used the least of the four requested by the sample ($n = 828$), TikTok is also the least likely to be donated ($n = 54$; 6.52%). Of the 1503 Facebook users, 8.38% ($n = 126$) donated their data. Instagram (149 out of 1565; 9.52%) and YouTube (174 out of 1799; 9.67%) showthe most substantial donations and donation rates. The willingness rates are more stable across platforms, again with TikTok marginally deviating by a lower willingness rate of 30.19% compared to the 32 to 33% of the other platforms. When applying the filter criterion for what counts as a complete data donation (in our case at least one data point for a minimum of 50% of requested data traces) the counts and fractions decrease most substantially for Facebook and Instagram (see Table 2, Donating users with filter (%)).

Predictors for data donation willingness

To explore what predictors lead to an increased probability of donation willingness (**RQ2**), we employed four multivariate linear regression models, one for each platform (see Figure 3). For detailed results, see Appendix F.

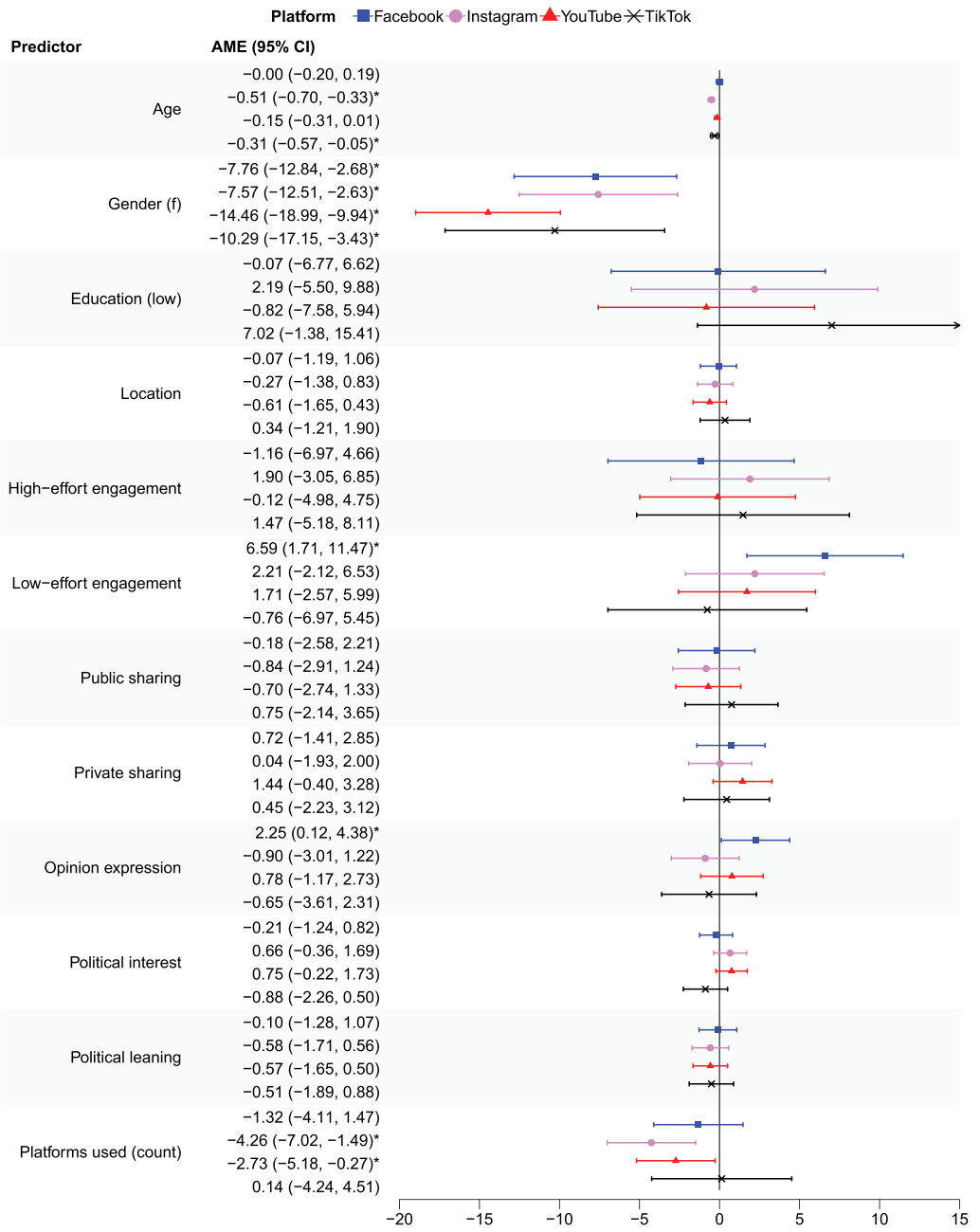

Figure 3. Results from multiple logistic regression models examining the willingness to donate data download packages. Average marginal effects (AMEs) and 95% confidence intervals are shown as full percentage points. Sample sizes: YouTube ($n = 1{,}799$), Facebook ($n = 1{,}503$), Instagram ($n = 1{,}565$), TikTok ($n = 828$). * indicates significance at $p \leq .05$.

For *age*, a significant negative effect was found for Instagram and TikTok users. Specifically, we observed that as the age of Instagram users increased by one year, their likelihood of indicating willingness to donate their data decreased by 0.51% (AME = −0.0051, $p \leq .001$). The likelihood of willingness of TikTok users to donate decreased by 0.31% with each year of age increase (AME = −0.0031, $p \leq .05$). Older users are hence less likely to donate their data. *Gender (f)* was a significant negative predictor across all platforms: Female users of the respective

platforms were less likely to be willing to donate their data by 7.76% (Facebook, $p \leq .01$), 7.57% (Instagram, $p \leq .01$), 14.46% (YouTube, $p \leq .001$) and 10.29% (TikTok, $p \leq .01$), respectively.

Regardless of age and gender, only two other predictors were significant predictors for participants' willingness to donate their data – both are so in the model for Facebook: First, with each one-unit shift in consuming and reaction behavior to news and politics (*low-effort social media engagement*), Facebook users were 6.59% more likely to indicate willingness to donate their data ($AME = .0659$; $p \leq 0.05$). Second, perceived freedom of expression of opinions online was a significant predictor for the willingness of Facebook users to donate data ($AME = 0.0225$; $p \leq .05$).

The number of self-reported platforms used was a significant negative predictor of the willingness to donate for Instagram ($AME = -.0426$; $p \leq .01$) and YouTube ($AME = -.0273$; $p \leq .05$). This implies that if Instagram and YouTube users had other options to choose at the willingness stage, they would be more likely to choose either of those or none at all, rather than YouTube or Instagram for donation.

Lower education, location, high-effort engagement (news/politics), public sharing, private sharing, political interest, and political leaning were insignificant across platforms in the willingness phase.

Predictors for data donation behavior

Next, we look at the predictors for the actual donation behavior (**RQ5**), by comparing the characteristics of donors and non-donors of the users who passed the willingness stage (see Figure 4 and Appendix G). Although *age* was a significant predictor for Instagram and TikTok users in the willingness stage, it was no longer so in the donation stage. However, we found a significant effect for YouTube users. Each additional year decreased the likelihood for YouTube users to donate their data by 0.29% (AME = −.0029, $p \leq 0.05$). *Gender (f)* was not a significant predictor for any platforms' user group.

Facebook remains unique when considering the participants location: With each increase in the level of urbanization (*location*) Facebook users' likelihood to donate decreased by 1.83% (AME = −0.0183 $p \leq .05$), meaning that users from urban areas are less likely to donate. For YouTube, Instagram, and TikTok donors, their place of residence did not play an significant role in their decision to donate.

The strongest significant effects in the donation stage were based on *high-effort social media use (news/politics)*, which indicates the frequency of self-reported posting, commenting, and public discussions online. Our results show that the more people self-reported such activities, the less likely they were to actually donate their Facebook and TikTok data. The likelihood to donate decreased by 12.81% (AME = 0.1281, $p \leq .001$, Facebook) and 11.53% (AME = 0.1153, $p \leq .05$, TikTok) for each one-unit increase in the respective factor score.

Political leaning was a significant predictor of YouTube, Instagram, and TikTok donation behavior. With each point toward self-reported right political leaning, the likelihood of Instagram ($AME = -.0274$; $p \leq .01$), YouTube ($AME = -.0289$; $p \leq .01$), and TikTok ($AME = -.0322$; $p \leq .05$) users to donate was decreasing. For all remaining predictors, no effects were found on actual donation behavior, including the position in which the platform was positioned in the data donation interface.

Considering the results with the stricter filter condition for a successful data donation applied, the overall findings remain stable, with the exception of the location effect for Facebook users (see Appendix G, Figure S9). As this effect was only marginally significant in the initial analysis, we attribute its disappearance to its limited robustness rather than to any systematic differences between low-quality donors and the filtered sample – especially since all stronger effects remain unchanged.

Model performance under both operationalizations is highly similar (see Appendix, Tables 16 and 17). Across all platforms, the logistic regression models clearly outperform their intercept-only baselines, and pseudo-R^2 values change only marginally between the two definitions. Log-likelihoods also shift in the expected direction without altering any substantive conclusions. These parallel patterns confirm that the disappearance of the marginal Facebook location effect reflects its limited robustness rather than systematic differences introduced by the filtering procedure.

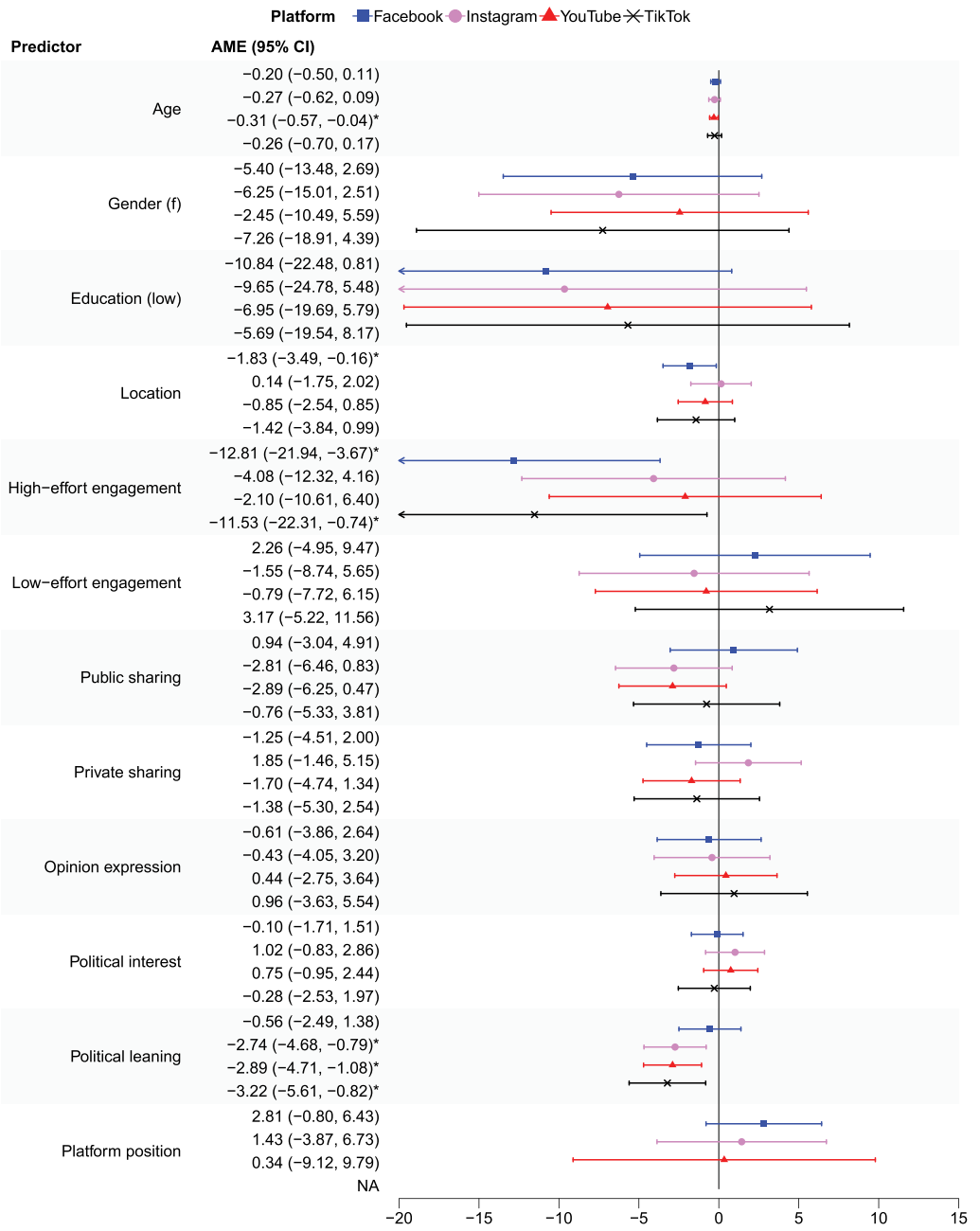

Figure 4. Results from multiple logistic regression models examining the actual data donation behavior. Average marginal effects (AMEs) and 95% confidence intervals are shown as full percentage points. Sample sizes: YouTube ($n = 591$), Facebook ($n = 502$), Instagram ($n = 506$), TikTok ($n = 250$). * indicates significance at $p \leq .05$.

Non-response bias

Lastly, to better understand how the final data donation samples for each platform are biased, we compare the characteristics of donors and the respective eligible platform users (see Figure 5 and Appendix H). In general, we found non-response biases for age, gender (f), education (low), political

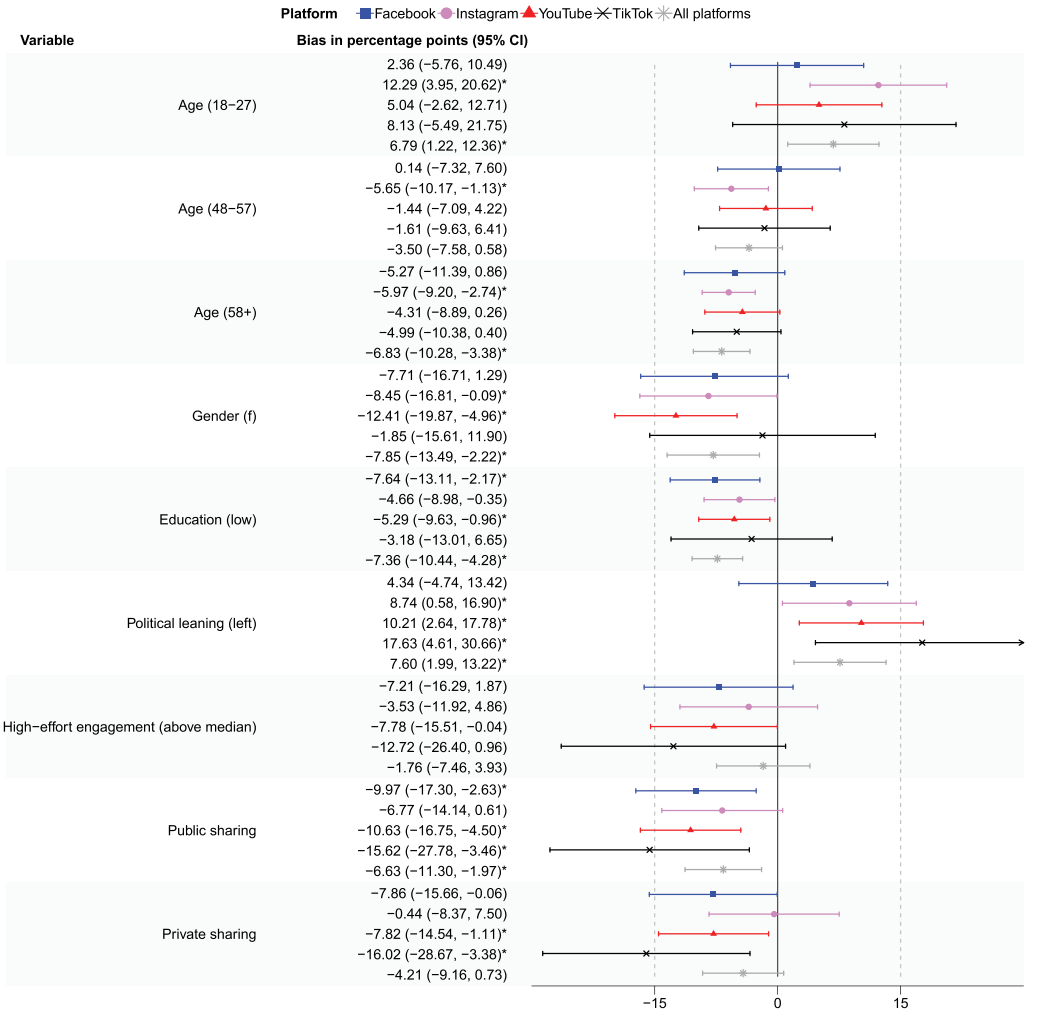

Figure 5. Non-response biases comparing the sample that donated data with the eligible sample, both per platform and in aggregate. The following levels/variables are omitted to highlight significant biases: age (28–37), age (38–47), education (high), location (urban), perceived freedom of opinion expression online (high), political interest (high), and low-effort engagement (above median). A full overview of all variables is provided in Appendix H. * indicates significance at $p < .05$.

leaning, high-effort engagement, public sharing, and private sharing. Biases are reported as percentage points (pp). We report on the biases platform by platform.

The Facebook sample contained fewer participants with lower education ($bias = -7.64$ pp, $p \leq .05$). Additionally, in the Facebook donor sample users who self-reported sharing frequently news or political information online publicly were statistically significantly underrepresented ($bias = -9.97$ pp, $p \leq .05$). A significant practical bias against users self-reporting frequent private sharing was also observed ($bias = -7.86$ pp, $p = .067$).

The sample of Instagram donors had a bias toward the age group 18 to 27 ($bias = +12.29$ pp; $p \leq .01$) and against those aged 48 to 57 ($bias = -5.65$ pp; $p \leq .05$) or older than 58 ($bias = -5.97$ pp; $p \leq .05$). The sample also contained fewer female participants than the sample of eligible Instagram users ($bias = 8.45$ pp; $p \leq .05$). We furthermore found a practical bias for Education (low), meaning that lower educated participants were on average less prevalent by 4.66% points among Instagram donors compared to all eligible Instagram users. The Instagram-donor sample was also biased with respect to their political lean toward the left by 8.74%.

The YouTube sample was negatively biased against female participants (*bias* = −12.41 pp; $p \leq .01$), and those with lower education (*bias* = −5.29 pp; $p \leq .05$). The YouTube sample also showed a bias toward politically left-leaning participants (*bias* = 10.21 pp; $p \leq .05$). In terms of social media engagement for the YouTube sample, we found a practical bias against high-effort social media engagement with news and political content (*bias* = −7.78 pp; $p \leq .05$), and against users self-reporting frequent public (*bias* = −10.63 pp; $p \leq .05$) and private sharing (*bias* = −7.82 pp; $p \leq .05$) of news and political content.

For TikTok, the sample did not show gender, age, or educational biases. However, the TikTok sample was biased in terms of politically left-leaning participants (*bias* = 17.63 pp; $p \leq .05$) and against frequent public (*bias* = −15.62 pp; $p \leq .05$) and private sharing (*bias* = −16.02 pp; $p \leq .05$) of news and political information.

For all non-response biases, we can see that whether a bias is found differs by platform, no bias applies across platform samples alike as the aggregated (gray, Figure 5) bias condition for comparison shows.

When comparing the results for data-donation success without any filter to those obtained under the stricter condition – requiring at least one entry for 50% of all data traces – the overall findings largely hold (see Appendix H, Figures 20 and 21). Importantly, the stricter definition weakens the education-related biases, removing the effect for Instagram donors, while it makes the biases related to social media engagement (high-effort/low-effort engagement; public/private sharing) more pronounced. Specifically, for Facebook we now find a negative bias for self-reported high-effort engagement (*bias* = 10.76 pp; $p \leq .05$), and for YouTube a negative bias for low-effort engagement (*bias* = 8.01% pp; $p \leq .05$). The weaker education bias likely reflects that this effect was not very strong to begin with. In contrast, the stronger engagement-related biases suggest that the amount of donated data after all matters: when limiting successful donations to those that contain more substantial data, the models show clearer negative effects linked to social media engagement variables.

Discussion

The literature on what drives data donation behavior and the resulting biases is fragmented and provides only selective information to DDP researchers when planning a data donation-based study. This paper systematically explored the donation behavior of a German non-probability quota sample from three perspectives: the willingness and donation rates, the predictors that might explain the donation behavior of the participants, and the non-response biases introduced throughout the data collection. Our findings show differences in the willingness and donation behavior between platforms and the resulting non-response biases. Importantly, for all platforms, we find non-response biases that need to be considered when conducting data donation studies – either pre-hoc or post-hoc. And for a subset of those, our predictive analysis also provides context on where and partly how those biases occur. Thus, we reveal genuinely between-platform differences and similarities.

With 14.81% of eligible participants donating data from at least one platform, the data donation rate of this study is comparable to other multi-platform studies in Germany through convenience samples (Hase & Haim, 2024). However, this aggregated rate, as well as our platform-specific rates, are significantly lower than the 37.79% reported by Keusch et al. (2024) in their 2021 Facebook data donation collection among Germans. This disparity can be partially attributed to their study using third-wave panel participants who already had a metering app installed on their devices. It can be assumed that these participants are more familiar and less concerned with the sharing of data for scientific purposes than the participants in our study, who were likely not yet exposed to passive data collection or data download packages. In addition, our incentive for survey participation was not tied to a successful data donation, in a scenario where only the combination of survey participation and data donation yields a reward, the donation rate would be likely higher.

In terms of incentives, previous literature has found that an increase in monetary incentive does not result in higher donation rates (Pfiffner, Reiss, et al., 2024). Moreover, previous studies with lower incentives report higher donation rates (2 Euro, Keusch et al., 2024), while those with more uncertain incentives (lottery for gift card, Hase & Haim, 2024) report similar donation rates. This points toward monetary incentives playing not as much of a role as one might assume intuitively. A promising future approach to improve donation rates could be to provide non-monetary incentives such as instant personalized reports on donated data (Groot Kormelink et al., 2025). Different projects of this kind have recently started (*Dein Feed, Deine Wahl,* 2025; What shapes your TikTok Feed?, 2025).

Main findings

Our analysis finds that multiple variables are statistically significant in predicting the willingness to donate and donation behavior. In addition, we find consistent biases in the final samples. Crucially, biases and predictors differ between platforms.

For demographic variables, we only see evidence of non-response biases for the Facebook, Instagram, and YouTube donor samples. Our TikTok DDP sample is free from demographic bias. Our Instagram donor sample is biased toward younger users. The Instagram and YouTube samples over represent male users. Both biases are introduced during the willingness stage. So, while being confronted with the hypothetical thought of donating the data decreased participation by older and female users of Instagram and YouTube, being confronted with the actual effortful process is not further impacting those demographic factors. This refines previous findings for Germany that did not find such aggregated bias across four collected platforms (Hase et al., 2024) and supports similar non-findings for Facebook donors (Keusch et al., 2024).

Crucially, Gender (f) is, for all platforms, found to be a significant and strong negative predictor of the willingness to donate data. So regardless of whether it results in a non-response bias in the final donor samples, female participants seem to be more hesitant to participate in data donation studies. This effect does not result in a Gender (f) non-response bias for the final Facebook and TikTok donor samples, but seems to be negotiated during the subsequent stages of the donation process. It is likely that unknown confounders are at play here.

Although lower education levels did not emerge as a significant predictor of donation behavior, we observed an underrepresentation of formally lower-educated users within the final data donation samples for Facebook, Instagram, and YouTube. This discrepancy suggests that a distinct set of participant characteristics, beyond education alone, may be responsible for this drop-off. One plausible explanation is that less-educated participants may perceive themselves as less technically proficient, which could hinder their ability to successfully complete the data donation process. This aligns with prior research that has identified both technical skill and higher education as key drivers of data donation behavior (Hase & Haim, 2024; Ohme et al., 2021; Strycharz et al., 2024). Further improvements in the ease of the donation process are necessary, as lower-educated users already count as a hard-to-reach group (Schepers et al., 2017).

The mechanisms and biases found underscore the need for targeted recruitment strategies (e.g., quota) to ensure demographic balance in data donation samples to counter the potential to exclude marginalized groups.

The study also highlights the importance of considering political orientation when analyzing data donation behavior. Our findings indicate that donors on all platforms except Facebook tend to be politically left-leaning. This bias could potentially influence the interpretation of research findings, particularly in studies that examine political topics. The bias is introduced during the donation phase – while political leaning does not impact the willingness step, people of right political leaning drop out once they have to actually donate their data. This finding is consistent with previous work (Hase et al., 2024; Welbers et al., 2024). Although general distrust in research can lead participants on the political right (Cologna et al., 2025) to deny their final donation, their initial participation in the survey suggests that

their curiosity outweighed their skepticism, at least until the point of a concrete final commitment. It is possible that being confronted with the actual data they were about to donate causes this bias, potentially due to social-desirability motivations.

In terms of social media engagement with news and political content, we find an interesting pattern. We found a non-response bias against users who frequently share news and political information publicly and privately on TikTok, YouTube, and Facebook, a bias not present in the Instagram sample. When stricter criteria for what counts as a successful data donation are applied, the biases related to sharing behavior and to high- and low-effort engagement became even more pronounced. Put differently, once users with very limited activity in their donated files are excluded and the analysis focuses only on those providing richer datasets, the underrepresentation of highly engaged users becomes stronger. This pattern suggests that donors who submitted minimal data tended to self-report higher engagement with news and political content than those whose richer data donations were retained. In contrast, users who contributed more substantial datasets report lower levels of news and political engagement in the survey, yet their donation files contain more overall activity – presumably reflecting engagement with other, nonpolitical content types.

Although descriptive biases in sharing behavior emerge, sharing news itself is not a significant predictor of either willingness to donate or successful donation. Instead, other forms of engagement with news and political content play a more substantial role. Low-effort engagement with news and political content (e.g., watching or reacting) is a positive and significant predictor of willingness to donate among Facebook users, whereas high-effort engagement (e.g., posting, commenting, discussing) is a negative predictor of donation behavior on both Facebook and TikTok. These patterns may be linked to platform-specific affordances. The negative relationship between high-effort engagement and data donation on Facebook and TikTok likely reflects increased perceptions of risk among users who frequently post or comment. On Facebook, high-effort activities are closely tied to one's social identity and personal network, which can make high-engagement users more aware of the sensitivity of their data and therefore more hesitant to share it for third-party analysis. On TikTok, high-effort engagement often involves content creation and personal expression; users who create or comment extensively may regard their data as particularly personal, leading to stronger privacy concerns when asked to donate it. By contrast, the positive relationship between low-effort engagement and willingness to donate on Facebook suggests that more passive users – those who mainly watch or react – perceive fewer risks. Their online footprint is less revealing, and they may therefore feel that they have less to lose by agreeing to a data donation. For Instagram, social media engagement is not related to either willingness or successful donation, nor do we observe a comparable bias. One possible explanation is that Instagram use is already highly public and visually oriented, potentially fostering lower privacy concerns. Moreover, Instagram content, in general, is less centered on news and political topics, which may reduce the perceived sensitivity of the data. Taken together, these findings indicate that platform-specific logics shape users' perceptions of data sensitivity and risk, which in turn influence their willingness to donate their social media data.

Previous research has found that general (Hase et al., 2024) and specific (Gilbert et al., 2021) privacy concerns result in reduced donation behavior. Our findings may reflect a related mechanism: highly active social media users could experience stronger privacy concerns simply because they have "more to lose," making them less inclined to share their data for research purposes despite reporting higher engagement. To better understand the emergence of this engagement bias, future research should adopt a more fine-grained approach to the distinct sub-steps of the donation process – request, download, and actual donation – as demonstrated by previous research in an analysis of sociodemographic effects (Welbers et al., 2024). Moreover, a combined analysis of social media engagement patterns and privacy concerns is needed to clarify whether heightened privacy awareness among highly active users drives their lower donation rates, or whether other mechanisms are at play.

Limitations

This study is subject to certain limitations. First, generalizability is limited due to the focus on a German non-probability quota sample, which may not fully capture population diversity. However, by employing quotas along education, age, and gender, the resulting sample provides a coarse representation of the target population. Second, the study relies on self-reported data, which can be prone to bias themselves and may not cover the full range of potential predictors that influence donation behavior. Finally, the small sample size of TikTok donors ($n = 54$) reduces our statistical power and makes it difficult to draw reliable conclusions related to this platform, highlighting the need for further research to understand its unique user base. Our study identifies TikTok users as especially hard to reach for a data donation study. The smaller sample size of TikTok donors is not just due to the smaller number of users in the general population, but also to the lower donation rate compared to other platforms. This might be due to the multi-platform choice environment of this study. However, our results show that at least for the willingness step, that here users tend to choose against moving on with YouTube and Instagram donations. Why participants still seem more likely to donate to a different platform than TikTok remains to be explored in future research.

Due to the study's design limitations, we were unable to differentiate the specific reasons for participant attrition during the donation stage. It remains unclear whether participants initiated the data request, encountered technical issues with data upload, or withdrew upon reviewing the specific requirements of the donation process. Therefore, any conclusions about dropout motivations at this stage are only speculative and require further investigation with a more refined study design in future research.

Another limitation of our study is that we measure social media engagement exclusively through interactions with news and political content. This means that our findings on the relationship between user participation and donation behavior may not fully generalize to other forms of social interaction, such as participation in entertainment, lifestyle content, or personal communication Although this focus aligns with a central interest of the communication science field, it presents a clear boundary on the applicability of our results.

Practical implications and future research

This research contributes to a growing body of literature on data donation behavior and provides valuable information for researchers looking to use this increasingly relevant data collection method. By highlighting platform-specific differences in donation behavior, this study underscores the need for tailored recruitment strategies and careful consideration of potential biases when designing and interpreting data donation studies.

The following core implications are the result of our findings for further DDP research:

- **Underlying motivations**: DDP samples are biased, and the reasons for this lie in systemic non-response errors throughout the data donation collection pipeline. To fully explain and thus counteract these biases, future research must explore the underlying motivations behind these observed platform-specific differences in donation behavior. Understanding why certain user groups are more or less likely to donate their data on specific platforms could inform the development of more effective recruitment strategies and help mitigate potential biases in data donation samples.
- **Oversampling**: Our findings point to the need for oversampling of female users for DDP studies on YouTube and Instagram; as well as an oversampling of lower educated users for Facebook, Instagram, and YouTube. It furthermore remains a challenge to reach politically right leaning participants for data donation studies.

- **Impact of bias on research**: Specifically, the engagement and political biases are highly relevant for typical research topics based on DDPs for political communication or media use effects. Current data donation samples run danger not just to be biased in terms of gender, age, or education, but to also consist of left leaning users that tend to use social media less intensively than the average user.

Conclusion

In conclusion, this study provides a comprehensive cross-platform analysis of data donation behavior, revealing significant disparities in willingness and actual donation rates across samples from YouTube, Facebook, Instagram, and TikTok. Notably, we observed that the likelihood of donating decreases with each additional platform, highlighting the increasing threshold for multi-platform data sharing. Additionally, donation rates are lower for TikTok than for YouTube, Instagram, and Facebook. Compared to single platform studies, the resulting donation rates are fairly low, which points to a research design with optional multi-platform donations not being very effective.

Our findings underscore the importance of considering platform-specific differences, as demographic biases and significant predictors varied substantially. While we identified gender (f) as a strong negative predictor for all platforms at the willingness stage, political leaning was only significant at the donation stage for all platforms except Facebook. Furthermore, only for Facebook and TikTok donors, frequent posting and commenting was a negative predictor during the donation stage. In terms of non-response biases, only the Instagram donor samples are biased toward younger users, while the other platforms are biased toward users who share news and political content less frequently online, both publicly (YouTube, TikTok, Facebook) and privately (YouTube, TikTok).

This research contributes valuable information for researchers using data donation methodologies, emphasizing the importance of developing tailored recruitment strategies and careful consideration of potential biases to ensure the robustness of their findings. Future studies should delve deeper into the underlying motivations driving these platform-specific behaviors, specifically in relation to social media engagement, and explore effective strategies to mitigate identified biases, thus improving the reliability and validity of data donation-based research.

Acknowledgments

To enhance the clarity and grammatical precision of this manuscript, the authors utilized Gemini 2.0 Flash by prompting and Writefull 2.2 within Overleaf for language improvement. The use of these AI tools was limited to language editing and did not involve the generation or manipulation of research data, figures, or original research content.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Federal Ministry of Research, Technology and Space (BMFTR) under Grant 16DII141 (Weizenbaum-Institut).

Notes on contributors

Lion Wedel is a research associate at the Weizenbaum-Institut in Berlin, in the research group "Digital News Dynamics". After studying Sociology and Economics at Martin–Luther–Universität Halle–Wittenberg, he earned his M.Sc. in Computational Social Systems at RWTH Aachen University. His research focuses on applying computational methods to social-scientific questions — especially how digital platform structures steer communication and information flows, and thereby shape public discourse.

Jakob Ohme is Research Group Lead of the "Digital News Dynamics" group at the Weizenbaum Institute, Germany, and Fellow at the Digital Communication Methods Lab at the University of Amsterdam, The Netherlands. His research interests center around the impact of digital and mobile communication on news exposure and political behavior in digital democracies. He is further interested in developing digital methods in political communication and journalism research, and in platform data access under the EU's Digital Services Act.

Anna-Theresa Mayer (M.A., LMU Munich) is a research associate in the research group "Digital News Dynamics" at the Weizenbaum Institute in Berlin and a doctoral candidate at the Freie Universität Berlin. Her research interests include digital journalism, public value creation, and news on social media.

Felix Gaisbauer studied Physics and Philosophy at LMU Munich and earned his PhD at the Max Planck Institute for Mathematics in the Sciences in 2022. He is part of the research group "Digital News Dynamics" at Weizenbaum Institute, where he studies the use of computational methods to capture news-driven public debate online. His core research areas are network theory, modelling of complex systems and quantitative analysis of platform communication, e.g. on Twitter.

Yangliu Fan is a postdoctoral researcher at the Weizenbaum Institute, Berlin. She is a fellow at the Institute for Media and Communication Studies at the Free University of Berlin. Before, she studied for her PhD in Social Data Science at the Copenhagen Center for Social Data Science (SODAS) at the University of Copenhagen in Denmark. She uses digital trace data to explore and model digital news dynamics. Besides, she is broadly interested in social data analysis, complex networks, and the science of science.

References

Bechmann, A., Kroman Brems, M., Olesen, M. K., Walter, J. G., & Wegmann, D. (2025). *Data donation as a method for investigating trends and challenges in digital media landscapes at national scale: The Danish population's use of YouTube as an illustrative case*. Nordicom, University of Gothenburg. Retrieved May 26, 2025, from https://urn.kb.se/resolve?urn=urn:nbn:se:norden:org:diva-13593

Behre, J., Hölig, S., & Möller, J. (2023). *Reuters Institute Digital News Report, 2023: Ergebnisse für Deutschland*. Verlag Hans-Bredow-Institut. Hamburg. https://doi.org/10.21241/ssoar.86851

Beuthner, C., Weiß, B., Silber, H., Keusch, F., & Schröder, J. (2024). Consent to data linkage for different data domains - the role of question order, question wording, and incentives. *International Journal of Social Research Methodology*, *27*(4), 375–388. https://doi.org/10.1080/13645579.2023.2173847

Blassnig, S., Mitova, E., Pfiffner, N., & Reiss, M. V. (2023). Googling referendum campaigns: Analyzing online search patterns regarding Swiss direct-democratic votes. *Media and Communication*, *11*(1), 19–30. https://doi.org/10.17645/mac.v11i1.6030

Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, *4*(2), 388–423. https://doi.org/10.5117/CCR2022.2.002.BOES

Boeschoten, L., de Schipper, N. C., Mendrik, A. M., van der Veen, E., Struminskaya, B., Janssen, H., & Araujo, T. (2023). Port: A software tool for digital data donation. *Journal of Open Source Software*, *8*(90), 5596. https://doi.org/10.21105/joss.05596

Bosch, O. J., Sturgis, P., Kuha, J., & Revilla, M. (2024). Uncovering digital trace data biases: Tracking undercoverage in web tracking data. *Communication Methods and Measures*, 1–21. https://doi.org/10.1080/19312458.2024.2393165

Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (2020). The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships. *New Media & Society*, *22*(11), 2058–2080. https://doi.org/10.1177/1461444820924622

Breuer, J., Breuer, J., Kmetty, Z., Kmetty, Z., Haim, M., Haim, M., Stier, S., & Stier, S. (2022). User-centric approaches for collecting Facebook data in the 'post-API age': Experiences from two studies and recommendations for future research. *Information, Communication & Society*, *26*(14), 1–20. https://doi.org/10.1080/1369118x.2022.2097015

Carrière, T. C., Boeschoten, L., Struminskaya, B., Janssen, H. L., De Schipper, N. C., & Araujo, T. (2025). Best practices for studies using digital data donation. *Quality & Quantity*, *59*(S1), 389–412. https://doi.org/10.1007/s11135-024-01983-x

Christner, C., Urman, A., Adam, S., & Maier, M. (2021). Automated tracking approaches for studying online media use: A critical review and recommendations. *Communication Methods and Measures*, 1–17. https://doi.org/10.1080/19312458.2021.1907841 MAG ID:3152635690.

Cologna, V., Mede, N. G., Berger, S., Besley, J. Brick, C. Joubert, M. Maibach, E. Mihelj, S. Oreskes, N. Schöfer, M. S. van der Linden, D. S (2025). *Trust in scientists and their role in society across 68 countries*. *Nat Hum Behav*. *9*. https://doi.org/10.1038/s41562-024-02090-5

Dein Feed, Deine Wahl. (2025). Retrieved August 28, 2025 from, https://dein-feed-deine-wahl.de/

De Vreese, C. H., Boukes, M., Schuck, A., Vliegenthart, R., Bos, L., & Lelkes, Y. (2017). Linking survey and media content data: Opportunities, considerations, and pitfalls. *Communication Methods and Measures*, *11*(4), 221–244. https://doi.org/10.1080/19312458.2017.1380175

Fan, Y., Ohme, J., & Wedel, L. (2025, May 24). Exploring temporal dynamics in digital trace data: Mining user-sequences for communication research. *arXiv: 2505.18790 [cs]*. https://doi.org/10.48550/arXiv.2505.18790

Gilbert, S., Vitak, J., & Shilton, K. (2021). Measuring Americans' comfort with research uses of their social media data. *Social Media + Society*, *7*(3), 20563051211033824. https://doi.org/10.1177/20563051211033824

Gil-López, T., Christner, C., de León, E., Makhortykh, M., Urman, A., Maier, M., & Adam, S. (2023). Do (not!) track me: Relationship between willingness to participate and sample composition in online. *Social Science Computer Review*, *41*(6), 2274–2292. https://doi.org/10.1177/08944393231156634

Groot Kormelink, T., Houwing, F., Struminskaya, B., Boeschoten, L., de Schipper, N., & Welbers, K. (2025). Meaningful informed consent? How participants experience and understand data donation. *Information, Communication & Society*, 1–18. https://doi.org/10.1080/1369118X.2025.2540915

Hase, V., Ausloos, J., Boeschoten, L., Pfiffner, N., Janssen, H., Araujo, T., Carrière, T., Vreese, C. D., Haßler, J., Loecherbach, F., Kmetty, Z., Möller, J., Ohme, J., Schmidbauer, E., Struminskaya, B., Trilling, D., Welbers, K., & Haim, M. (2024). Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies? *Internet Policy Review*, *13*(3). https://policyreview.info/articles/analysis/fulfilling-data-access-obligations

Hase, V., & Haim, M. (2024). Can we get rid of bias? Mitigating systematic error in data donation studies through survey design strategies. *Computational Communication Research*, *6*(2), 1. https://doi.org/10.5117/CCR2024.2.2.HASE

Heiss, R., Schmuck, D., & Matthes, J. (2019). What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society*, *22*(10), 1497–1513. https://doi.org/10.1080/1369118X.2018.1445273

Kahlert, P. (2024, April 17). *YouTube recommends . . .*. STS@ENS. Retrieved October 6, 2024, from, https://medium.com/sts-ens/youtube-recommends-148b5608635a

Keusch, F., Pankowska, P. K., Cernat, A., & Bach, R. L. (2024). Do you have two minutes to talk about your data? Willingness to participate and nonparticipation bias in Facebook data donation. *Field Methods*, *36*(4), 279–293. https://doi.org/10.1177/1525822X231225907

Keusch, F., Struminskaya, B., Struminskaya, B., Antoun, C., Antoun, C., Couper, M. P., Couper, M. P., Kreuter, F., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, *83*(1), 210–235. https://doi.org/10.1093/poq/nfz007 MAG ID:2963617970.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746–759. https://doi.org/10.1177/0013164496056005002

Kmetty, Z., Stefkovics, Á., Számely, J., Deng, D., Kellner, A., Pauló, E., Omodei, E., & Koltai, J. (2024). Determinants of willingness to donate data from social media platforms. *Information, Communication & Society*, 1–26. https://doi.org/10.1080/1369118X.2024.2340995

Loecherbach, F., Moeller, J., Trilling, D., & van Atteveldt, W. (2024). What is news? Mapping the diversity of news experiences in digital trace data. *Journalism*, 14648849241303115. https://doi.org/10.1177/14648849241303115

Ohme, J., & Araujo, T. (2022). Digital data donations: A quest for best practices. *Patterns*, *3*(4), 100467. https://doi.org/10.1016/j.patter.2022.100467

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B., & Robinson, T. N. (2024). Digital trace data collection for social media effects research: APIs. *Data Donation, and (Screen) Tracking*, *18*(2), 124–141. https://doi.org/10.1080/19312458.2023.2181319

Ohme, J., Araujo, T., Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS screen time function. *Mobile Media & Communication*, *9*(2), 293–313. https://doi.org/10.1177/2050157920959106

Ortega, A. G., Bourgeois, J., Hutiri, W. T., & Kortuem, G. (2023). Beyond data transactions: A framework for meaningfully informed data donation. *AI & Soc*, *40* (8), 30. https://doi.org/10.1007/s00146-023-01755-5

Ortega, A. G., Bourgeois, J., & Kortuem, G. (2022). Reconstructing intimate contexts through data donation: A case study in menstrual tracking technologies. In Nordic Human-Computer Interaction Conference (NordiCHI '22), New York, NY, USA. Association for Computing Machinery. https://doi.org/10.1145/3546155.3546646

Otto, L. P., Loecherbach, F., & Vliegenthart, R. (2024). Linkage analysis revised - linking digital traces and survey data. *Communication Methods and Measures*, *18*(2), 186–204. https://doi.org/10.1080/19312458.2023.2257595

Pak, C., Cotter, K., & Thorson, K. (2022). Correcting sample selection bias of historical digital trace data: Inverse probability weighting (IPW) and type II Tobit model. *Communication Methods and Measures*, *16*(2), 134–155. https://doi.org/10.1080/19312458.2022.2037537

Pfiffner, N., & Friemel, T. N. (2023). Leveraging data donations for communication research: Exploring drivers behind the willingness to donate. *Communication Methods and Measures*, *17*(3), 227–249. https://doi.org/10.1080/19312458.2023.2176474

Pfiffner, N., Reiss, M. V., Mitova, E., & Blassnig, S. (2024). Comparing online recruitment strategies for data donation studies. *Computational Communication Research*, *6*(2), 1. https://doi.org/10.5117/CCR2024.2.8.PFIF

Pfiffner, N., Witlox, P., & Friemel, T. N. (2024). Data donation module: A web application for collecting and enriching data donations. *Computational Communication Research*, *6*(2), 1. https://doi.org/10.5117/CCR2024.2.4.PFIF

Pouwels, J. L., Araujo, T., van Atteveldt, W., Bachl, M., & Valkenburg, P. M. (2023). Integrating communication science and computational methods to study content-based social media effects. *Communication Methods and Measures*, *18* (2), 1–9. https://doi.org/10.1080/19312458.2023.2285766

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., Chiatti, A., Cho, M. J., Roehrick, K., Yang, X., Gagneja, A., Brinberg, M., Muise, D., Lu, Y., Luo, M., Fitzgerald, A., & Yeykelis, L. (2021). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human-Computer Interaction*, *36*(2), 150–201. https://doi.org/10.1080/07370024.2019.1578652

Schepers, W., Juchtmans, G., & Nicaise, I. (2017). *Reaching out to hard-to-survey groups among the poor.survey protocols, statistical issues and research design (milestone 20.5)*. inGRID.

Semenova, A., Degeling, M., & Hess, G. (2024, August 26). *Understanding TikTok's for you feed - auditing TikTok*. Retrieved October 6, 2024, from https://tiktok-audit.com/blog/2024/For-You-Feed/

Sigismund, M. (2021, December 9). *Regiostar regional statistical spatial typology for mobility and transport research*. Federal Ministry of Transport and Digital Infrastructure (BMVI). Retrieved October 11, 2024, from https://bmdv.bund.de/SharedDocs/DE/Anlage/G/regiostar-raumtypologie-englisch.pdf?blob=publicationFile

Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., Siegers, P., Stier, S., & Weiß, B. (2022). Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behaviour. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Supplement_2), S387–S407. https://doi.org/10.1111/rssa.12954

Statistisches Bundesamt. (2022). *Internationale Bildungsindikatoren im Ländervergleich* (Tabellenband no. Ausgabe 2022). Statistische Ämter des Bundes und der Länder. Retrieved October 18, 2024, from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Publikationen/Downloads-Bildungsstand/bildungsindikatoren-1023017227004.html.

Struminskaya, B., Lugtig, P., Toepoel, V., Schouten, B., Giesen, D., & Dolmans, R. (2021). Sharing data collected with smartphone sensors: Willingness, participation, and nonparticipation bias. *Public Opinion Quarterly*, *85*(S1), 423–462. https://doi.org/10.1093/poq/nfab025

Strycharz, J., Meppelink, C., Zarouali, B., Araujo, T., & Voorveld, H. (2024). The blind spot in data donations: Who is (not) willing to donate digital data in social scientific research. *Computational Communication Research*, *6*(2), 1. https://doi.org/10.5117/CCR2024.2.3.STRY

Toth, R. (2023). One app to assess them all. *Publizistik*, *68*(2), 281–290. https://doi.org/10.1007/s11616-023-00788-6

van Driel, I. I., Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and pitfalls of social media data donations. *Communication Methods and Measures*, *16*(4), 266–282. https://doi.org/10.1080/19312458.2022.2109608

van Hoof, M., Trilling, D., Meppelink, C., Möller, J., & Loecherbach, F. (2024). Googling politics? Comparing five computational methods to identify political and news-related searches from web browser histories. *Communication Methods and Measures*, *19*(1), 1–27. https://doi.org/10.1080/19312458.2024.2363776

Verbeij, T., Beyens, I., Trilling, D., & Valkenburg, P. M. (2024). Happiness and sadness in adolescents' Instagram direct messaging: A neural topic modeling approach. *Social Media + Society*, *10*(1), 20563051241229655. https://doi.org/10.1177/20563051241229655

Wedel, L., Mayer, A.-T., Fan, Y., Gaisbauer, F., & Ohme, J. (2026, January). *Multi-platform social media data donation behavior dataset*. Weizenbaum Library. https://doi.org/10.34669/wi.rd/6

Wedel, L., Ohme, J., & Araujo, T. (2024). Augmenting data download packages - integrating data donations, video metadata, and the multimodal nature of audio-visual content. *Methods, Data, Analyses*, 32. https://doi.org/10.12758/mda.2024.08

Welbers, K., Loecherbach, F., Lin, Z., & Trilling, D. (2024). Anything you would like to share: Evaluating a data donation application in a survey and field study. *Computational Communication Research*, *6*(2), 1. https://doi.org/10.5117/CCR2024.2.5.WELB

What Shapes Your TikTok Feed?. (2025). Retrieved August 28, 2025, from https://www.foryouparticipate.net/tiktok/

Wojcieszak, M., Menchen-Trevino, E., Clemm Von Hohenberg, B., De Leeuw, S., Gon¸calves, J., Davidson, S., & Gon¸calves, A. (2024). Non-news websites expose people to more political content than news websites: Evidence from browsing data in three countries. *Political Communication*, *41*(1), 129–151. https://doi.org/10.1080/10584609.2023.2238641

Wu-Ouyang, B., & Chan, M. (2023). Overestimating or underestimating communication findings? Comparing self-reported with log mobile data by data donation method. *Mobile Media & Communication*, *11*(3), 415–434. https://doi.org/10.1177/20501579221137162

Xenos, M. A., Macafee, T., & Pole, A. (2017). Understanding variations in user response to social media campaigns: A study of Facebook posts in the 2010 US elections. *New Media & Society*, *19*(6), 826–842. https://doi.org/10.1177/1461444815616617