

# The political effects of X's feed algorithm

<https://doi.org/10.1038/s41586-026-10098-2>

Germain Gauthier<sup>1,5</sup>, Roland Hodler<sup>2,5</sup>, Philine Widmer<sup>3,5</sup> & Ekaterina Zhuravskaya<sup>3,4,5</sup>✉

Received: 16 December 2024

Accepted: 4 January 2026

Published online: 18 February 2026

Open access

 Check for updates

Feed algorithms are widely suspected to influence political attitudes. However, previous evidence from switching off the algorithm on Meta platforms found no political effects<sup>1</sup>. Here we present results from a 2023 field experiment on Elon Musk's platform X shedding light on this puzzle. We assigned active US-based users randomly to either an algorithmic or a chronological feed for 7 weeks, measuring political attitudes and online behaviour. Switching from a chronological to an algorithmic feed increased engagement and shifted political opinion towards more conservative positions, particularly regarding policy priorities, perceptions of criminal investigations into Donald Trump and views on the war in Ukraine. In contrast, switching from the algorithmic to the chronological feed had no comparable effects. Neither switching the algorithm on nor switching it off significantly affected affective polarization or self-reported partisanship. To investigate the mechanism, we analysed users' feed content and behaviour. We found that the algorithm promotes conservative content and demotes posts by traditional media. Exposure to algorithmic content leads users to follow conservative political activist accounts, which they continue to follow even after switching off the algorithm, helping explain the asymmetry in effects. These results suggest that initial exposure to X's algorithm has persistent effects on users' current political attitudes and account-following behaviour, even in the absence of a detectable effect on partisanship.

Social media platforms have fundamentally transformed human lives: a large and growing share of the global population connects with others, gets entertained and learns about the world through social media<sup>2</sup>. These platforms have also become increasingly important for political news consumption. A quarter of US adults report social media as their primary news source, and one half say they at least sometimes get news from these platforms<sup>3</sup>. Typically, platforms use feed algorithms to select and order content in personalized feeds for each user<sup>4</sup>. Before algorithms were introduced, users saw a simple chronological feed that displayed posts from followed accounts, with the most recent posts appearing at the top.

Public intellectuals and scholars have raised concerns about the potential adverse effects of social media, particularly feed algorithms, on social cohesion, trust and democracy<sup>5–8</sup>. These concerns arise from the spread of misinformation<sup>9–11</sup>, the promotion of toxic and inflammatory content<sup>12–14</sup> and the creation of 'filter bubbles' with increasingly polarized content<sup>15–18</sup>. There is substantial rigorous quantitative evidence that internet access and social media indeed have important negative effects<sup>19–22</sup>. Research on search engine rankings also shows that the order in which information is presented can influence user behaviour and political beliefs<sup>23</sup>. However, previous literature on the effects of social media feed algorithms reports zero political effects. A large study of Facebook and Instagram, conducted by academics in cooperation with Meta during the 2020 US election, found that experimentally replacing the algorithmically curated feed with a chronological feed did not lead to any detectable effects on users' polarization or political attitudes, despite causing a substantial change in political

content and lowering user engagement with the platforms<sup>1</sup>. Similarly, studies on Google's search engine and YouTube algorithms found little evidence of filter bubbles<sup>24–27</sup>. Studies of Meta platforms linking content to user behaviour and attitudes also found no impact, despite prevalent like-minded content and amplified political news<sup>28–30</sup>.

Yet, the fact that switching off a feed algorithm does not affect users' political attitudes does not mean that algorithms have no political impact. If the initial exposure to the algorithm has a persistent effect on political outcomes, switching off the algorithm might show no effects despite its importance. For instance, this could happen because people start following accounts suggested by the algorithm and continue following them when the algorithm is switched off. In addition, different platforms may have different effects, for instance, due to different informational environments or the different objectives of their owners<sup>31–34</sup>.

We study the effects of X's feed algorithm and find that switching the algorithm on substantially shifts attitudes on policies and current political news towards more conservative opinions without significantly affecting polarization or partisanship. We conducted a randomized experiment involving actual X users in the United States over a 7-week period in the summer of 2023.

Our study departs from the previous literature in several ways. First, we leverage the feature on X that allowed users in 2023 to choose between a chronological feed (the 'Following' tab) and an algorithmic feed (the 'For you' tab), where content was both added (showing posts from accounts not followed by the user) and reordered (prioritizing some posts while hiding others), compared with the chronological feed setting (as confirmed by our data). This feature enabled us to conduct

<sup>1</sup>Department of Social and Political Sciences, Bocconi University, Milan, Italy. <sup>2</sup>Department of Economics, University of St.Gallen, St.Gallen, Switzerland. <sup>3</sup>Paris School of Economics, Paris, France. <sup>4</sup>École des Hautes Études en Sciences Sociales, Paris, France. <sup>5</sup>These authors contributed equally: Germain Gauthier, Roland Hodler, Philine Widmer, Ekaterina Zhuravskaya. ✉e-mail: [ekaterina.zhuravskaya@psemail.eu](mailto:ekaterina.zhuravskaya@psemail.eu)

two experiments simultaneously: examining the effects of switching the feed algorithm on for users who previously used the chronological feed and of turning it off for those who were on the algorithmic feed before the experiment.

Second, this feature allowed us to conduct the experiment independently, without cooperation from X. Hence, we avoid potential concerns specific to studies conducted in partnership with platforms<sup>35,36</sup>. Recently, another independent study used a browser extension to show that re-ranking content within X's algorithmic feed by moving up or down posts expressing anti-democratic attitudes and partisan animosity influenced affective polarization<sup>37</sup>. In contrast, we randomize users' exposure to X's algorithmic feed as designed and implemented by X or to a chronological feed. We then quantify which posts the algorithm promotes and demotes, estimate its causal effects on users' attitudes and behaviour, and leverage behavioural responses to provide evidence on the mechanism.

Finally, in addition to affective polarization and partisanship, we also study outcomes such as policy priorities and attitudes towards current political events, which may be less rigid and, therefore, more easily influenced by exposure to different social media content.

Our experiment took place more than 6 months after Elon Musk's acquisition of Twitter, a few months following the publication of the platform's source code and shortly after Linda Yaccarino assumed the role of CEO, yet about 1 year before Musk's public endorsement of Donald Trump in July 2024. An earlier study examined changes in the content of users' feeds on Twitter, when the platform introduced the feed algorithm in 2016, well before Musk's takeover, and found that the algorithm already prioritized right-wing content<sup>38</sup>, despite different platform ownership. In addition to content analysis, our study focuses on the effects on real users' behaviour and political attitudes<sup>38</sup>.

## Design

We conducted an experiment with X users during July and September of 2023. This research resulted from a collaboration of academics independent of X. We obtained ethical approval from the Ethics Committee of the University of St. Gallen, Switzerland (see Supplementary Information section 1.1 for a detailed discussion of ethical considerations and the measures we implemented to uphold ethical integrity). We pre-registered the experiment with the American Economic Association's registry for randomized controlled trials (AEARCTR-0011464).

The experiment had several phases: recruitment, a pre-treatment survey (collecting baseline characteristics), randomization into feed settings, a treatment phase (with participants using their assigned feed setting) and a post-treatment survey to gather self-reported outcomes. We also collected data on the content of users' feeds and data on users' online behaviour. Participants were recruited through the survey company YouGov, drawing from US-based registered members of the YouGov panel. Participation was voluntary and compensated.

YouGov contacted 13,265 participants to enter the pre-treatment survey. Of these, 3,434 were screened out as they were not active X users. By design, we only admitted participants who self-reported being active on X at least 'several times a month'. A total of 8,363 participants provided informed consent and 6,043 completed the entire pre-treatment survey. During the survey, each participant was assigned randomly to a feed setting, which they were paid to stay on until completing the post-treatment survey. Admission to the pre-treatment survey occurred on a rolling basis in July 2023. At the end of August 2023, all participants were invited for the post-treatment survey, completed by the third week of September. The treatment phase duration varied slightly among participants, with the average time between entering and leaving the study being 7 weeks and 1 day, with s.d. of 3 days. During the treatment phase, participants received two reminders to stick to their assigned feed setting. A total of 4,965 participants completed the post-treatment survey, forming our main sample. Extended Data Fig. 1

presents the flowchart for the experiment, describing its structure and sample size at each stage. Supplementary Information sections 1.2–1.8 provide further details about the structure of the experiment, all collected data, compliance and attrition at each stage.

In the pre-treatment survey, we inquired about which feed setting the participants used before the experiment. Initially, 76% of participants were using the algorithmic feed, the default on X; and 24% were using the chronological feed. For the duration of the study, participants were assigned randomly with equal probability to either the algorithmic or the chronological feed.

The pre-treatment survey included questions about X usage, including purpose and frequency, as well as questions on life satisfaction, partisanship and feeling thermometers for Democrats and Republicans, which we use to calculate affective polarization. YouGov provided data on participants' socio-economic backgrounds. All pre-treatment characteristics were fully balanced, except for a slight imbalance in the initial feed setting. The share of participants initially on the algorithmic feed setting was two percentage points higher among those assigned randomly to the algorithmic feed (77% versus 75%; two-sided *t*-test:  $P = 0.08$ ; Supplementary Information section 1.7). In our analysis, we always control for the initial feed setting. Participants who had to switch their settings in either direction as a result of the experiment were more likely to drop out (22.7% versus 20.9%; two-sided *t*-test:  $P = 0.077$ ). As demonstrated in Methods, selective attrition during the treatment phase does not drive our results.

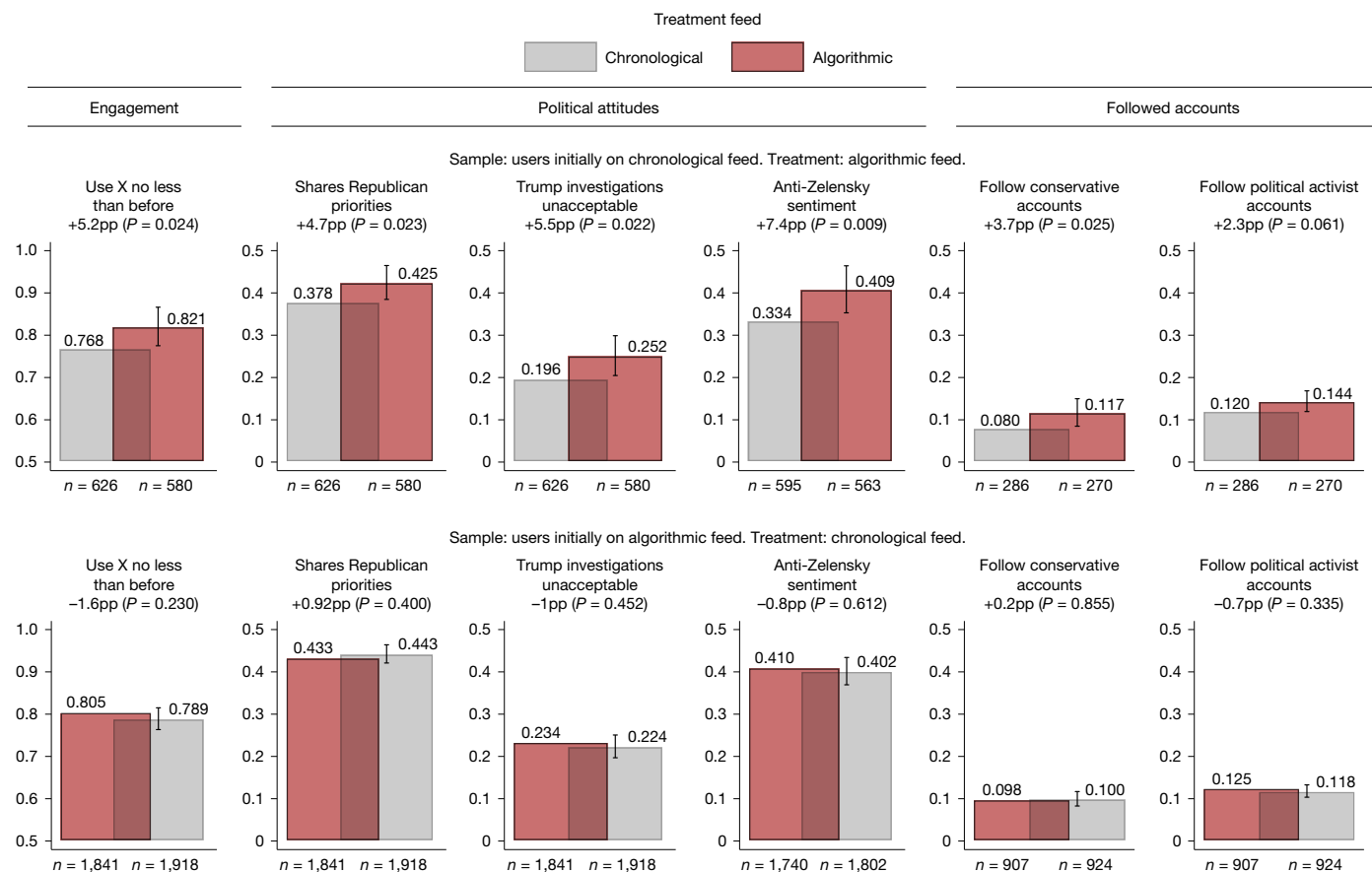
In the post-treatment survey, we collected several groups of outcomes. First, we repeated questions about X usage, partisanship, affective polarization and life satisfaction. Furthermore, we asked participants to rank policy areas by priority and to express their views on current political events, specifically their assessments of the criminal investigations into Donald Trump and the war in Ukraine, using a series of questions. For the baseline analysis, we aggregate these outcomes into seven composite measures by topic: user engagement, partisanship and affective polarization, conservative policy priorities, belief that the investigations into Trump are unacceptable, pro-Kremlin attitudes towards the war in Ukraine and life dissatisfaction. We also construct an overall measure that combines all questions related to policy attitudes and current political news. In the Supplementary Information, we present results for each individual outcome.

At the end of both surveys, we offered participants the option to install and run a Google Chrome extension on their desktop computer for additional compensation. This extension, designed for the experiment, captured the first 100 posts seen under each feed setting and saved them to a file on the participant's device, which they were asked to upload. A total of 784 participants ran the extension at the end of the pre-treatment survey, and 599 did so during the post-treatment survey.

For participants using the Google Chrome extension, we observed their actual X feed setting directly and, hence, their compliance with the assigned setting. The observed compliance rate in this sample was 89.15%. In addition, we asked all participants in the post-treatment survey whether they complied with the assigned feed setting during the study period. The self-reported compliance rate, that is, the share responding that they complied most of the time or always with the assigned setting, was 85.38% in the full sample (Extended Data Fig. 2).

We asked participants whether they agreed to share their X handles with us and grant permission to scrape their publicly available data, specifically, the names of the accounts they followed. All participants were informed about the purpose of our use of their X handles, and their decision to participate did not affect their compensation. We scraped the list of accounts followed by 2,387 participants after the post-treatment survey. Supplementary Information section 1.3 describes the construction of all variables and reports summary statistics.

Our experimental design allows us to evaluate the impact of switching the feed algorithm on and off. Intuitively, for respondents who were



**Fig. 1 | Raw data summary for the main outcomes by initial feed setting and treatment.** Means for selected outcomes by assigned treatment feed in the two samples: top, summarized outcomes for the sample of users initially on the chronological feed; bottom, summarized outcomes for the sample of participants initially on the algorithmic feed. The difference in the height of the two bars in each chart represents the unconditional ITT effect. Error bars, 95% CI for the ITT effect relative to the control mean (Control mean + ITT ± 1.96 × s.e.). The title of each chart presents the treatment-effect magnitude in percentage points (pp) and its *P* value (all tests are two-sided). The unit of observation

is respondent. Sample sizes for each subgroup are reported below the bars. All outcomes are indicator variables except 'Shares Republican priorities', for which the minimum, first quartile, median, third quartile and maximum are: 0, 0, 0.33, 0.67, 1 for initial-Chronological to Chronological subsample; 0, 0, 0.33, 0.67, 1 for initial-Chronological to Algorithmic; 0, 0, 0.33, 0.67, 1 for initial-Algorithmic to Algorithmic; and 0, 0.33, 0.33, 0.67, 1 for initial-Algorithmic to Chronological. Supplementary Information section 2.1 presents similar bar charts for all individual outcome variables. Supplementary Information section 2.4 presents regression results for all outcome variables.

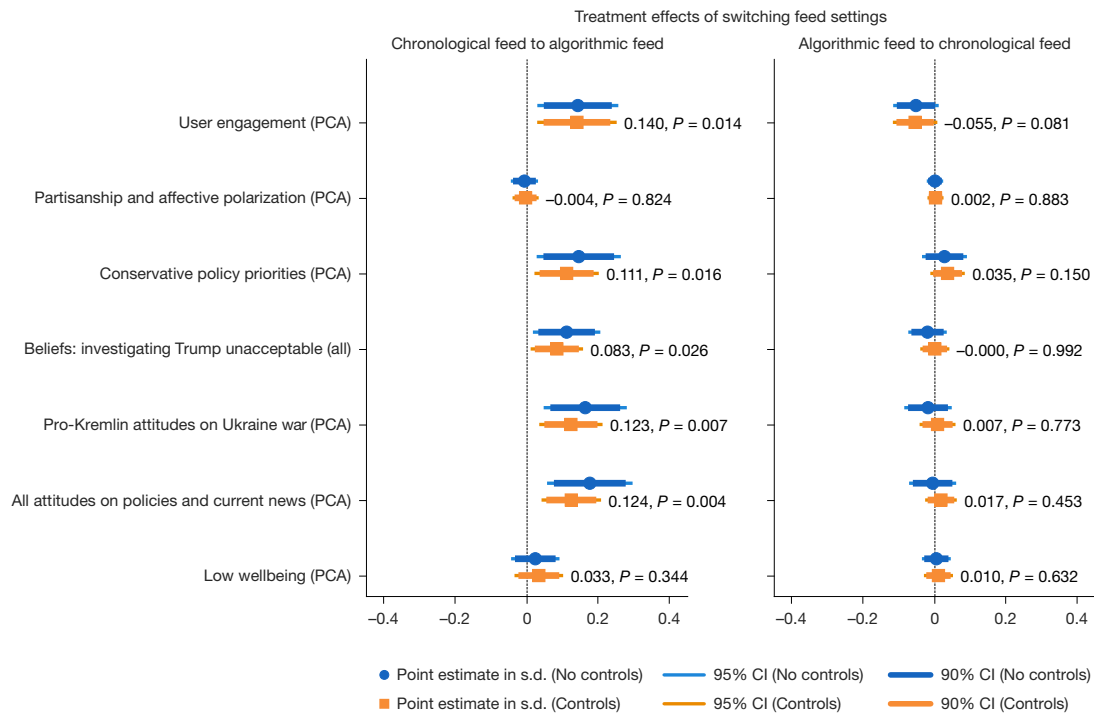
initially on the chronological feed, we compare the outcomes between those randomized to remain on the chronological feed and those randomized to switch to the algorithmic feed. Similarly, for respondents who were initially on the algorithmic feed, we compare the outcomes between those who remained randomly on the algorithmic feed and those who switched randomly to the chronological one. Our empirical strategy for baseline results is detailed in Methods, which also summarizes our analyses of attrition and compliance (discussed in detail in Supplementary Information section 1.8).

### Impact of feed algorithm on attitudes and behaviour

Figure 1 illustrates the main results by summarizing raw data means by treatment status for selected individual outcomes in a series of bar charts (all tests are two-sided, similar graphs for all individual outcomes are presented in Supplementary Information section 2.1). The top row presents data for participants who were initially on the chronological feed. Among these participants, those assigned to switch to the algorithmic feed were 5.2 percentage points less likely to reduce their X usage than those who remained on the chronological feed (95% confidence interval (CI): 0.7, 9.7; *P* = 0.024). They were 4.7 percentage points more likely to prioritize policy issues considered important by Republicans, such as inflation, immigration and crime (95% CI: 0.7, 8.7;

*P* = 0.023). They were also 5.5 percentage points more likely to believe that the investigations into Trump are unacceptable, describing them as contrary to the rule of law, undermining democracy, an attempt to stop the campaign and an attack on people like themselves (95% CI: 0.8, 10.2; *P* = 0.022). They were 7.4 percentage points less likely to hold a positive view of Ukrainian President Volodymyr Zelensky (95% CI: 1.8, 13.0; *P* = 0.009). Finally, they were 3.7 and 2.3 percentage points more likely to follow any conservative account (95% CI: 0.5, 7.0; *P* = 0.025) and any political activist account (95% CI: -0.1, 4.8; *P* = 0.061) on X, respectively. The bottom row of Fig. 1 presents the data summaries of the same outcomes for participants who were initially on the algorithmic feed. In sharp contrast to the top row, the outcome means in this group were not significantly affected by treatment assignment.

Regression results using aggregate outcomes are presented in Fig. 2, which displays the results of regression analysis for the composite survey outcomes. Average intention-to-treat (ITT) effects are reported (Methods). On the left, we show the standardized effects of switching the algorithm on for participants initially using the chronological feed; on the right are shown the effects of switching it off for those initially using the algorithmic feed. For each outcome, we report two sets of regression estimates: unconditional, controlling only for the initial feed setting and the pre-treatment value of the outcome variable when available; and conditional, controlling for all collected pre-treatment



**Fig. 2 | ITT estimates of feed-setting changes on engagement and political attitudes.** ITT effect estimates of switching the algorithm on and off (in s.d.). Left, effect of moving from the chronological to the algorithmic feed for users initially on the chronological feed. Right, effect of moving in the opposite direction for users initially on the algorithmic feed. For each outcome, the results of two specifications are reported. Blue, unconditional estimates with robust s.e., controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. Orange: conditional estimates, controlling for pre-treatment covariates using GRFs; 90% and 95% CIs are reported. Numerical

effect sizes and *P* values correspond to the conditional estimates (all tests are two-sided). The unit of observation is respondent. From top to bottom, sample sizes are  $n = 4,965$ ,  $n = 3,337$ ,  $n = 4,965$ ,  $n = 4,965$ ,  $n = 4,596$ ,  $n = 4,596$  and  $n = 4,850$ . Tests are described in Methods. Supplementary Information Table 2.16 reports the exact numerical point estimates, s.e., CIs and sample sizes for every specification. All outcomes are standardized. Additional results are presented in Supplementary Information section 2. PCA, first principal component from principal component analysis.

covariates, with the specification chosen using generalized random forests (GRFs)<sup>39</sup>.

First, we confirmed that the algorithmic feed is more engaging. Among participants initially on the chronological feed, the first principal component of the user engagement measures increased by 0.14 s.d. for those assigned to the algorithmic feed, compared with those who remained on the chronological feed (95% CI: 0.03, 0.25;  $P = 0.014$ ). For participants initially on the algorithmic feed, engagement declined by 0.06 s.d. for those assigned to the chronological feed, but this effect is not significant at the 5% level (95% CI: -0.12, 0.01;  $P = 0.081$ ). Second, we find precisely estimated null effects of the algorithm on partisanship and affective polarization, whether the algorithm was switched on or off.

Outcomes related to attitudes towards policies and current political events were affected significantly by turning the algorithm on. Participants who switched from the chronological to the algorithmic feed prioritized a conservative policy agenda over a liberal one by 0.11 s.d., compared with those who remained on the chronological feed (95% CI: 0.02, 0.20;  $P = 0.016$ ). The experiment also influenced views on the criminal investigations into Trump: those who switched to the algorithmic feed were 0.08 s.d. more likely to consider them completely unacceptable (95% CI: 0.01, 0.16;  $P = 0.026$ ). This effect is larger and more precisely estimated for aspects of this belief held by a smaller share of participants in the control group (as shown in Supplementary Information Fig. 2.4), indicating that the effect is concentrated among more loyal Trump supporters. Regarding Russia's invasion of Ukraine, for participants initially on the chronological feed, exposure to the algorithmic feed increased pro-Kremlin attitudes by 0.12 s.d. (95% CI: 0.03, 0.21;  $P = 0.007$ ). Finally, we address multiple

hypothesis testing concerns by aggregating all questions concerning policies and current political news into a single index. We find that this index is also 0.12 s.d. higher for those who were randomized to switch from the chronological to the algorithmic feed, compared with those who stayed on the chronological feed (95% CI: 0.04, 0.21;  $P = 0.004$ ).

Remarkably, none of the political attitudes were affected by the reverse switch from the algorithmic to the chronological feed, that is, by switching the algorithm off. This is consistent with the earlier findings of the Meta study<sup>1</sup>. We also found no significant impact of the experimental treatments on subjective wellbeing. As demonstrated by the similarity between the two specifications, unconditional versus conditional on pre-treatment covariates, the significant impact of switching the algorithm on, and the absence of any effect from switching it off, are robust to the choice of covariates and model specification. Supplementary Information Fig. 2.9 presents results for all individual outcomes in the same format as in Fig. 2. We also verify that the ITT estimates are very similar to the local average treatment effects (LATE) on self-reported compliers, presented in Extended Data Fig. 3 and discussed in detail in Methods.

The sample size is insufficient to reliably test for heterogeneities in these effects. However, the split of the sample by self-reported pre-treatment partisanship into Democrats versus Republicans and Independents (or other political parties) indicates strongly that the results are driven by Republicans and Independents, whereas Democrats' views are largely unaffected by the experiment (Extended Data Fig. 4).

The initial feed setting is correlated with participants' socio-demographic characteristics, as it may reflect individual preferences and because selecting the chronological feed requires an active choice:

those on the chronological feed are more likely (based on two-sided *t*-tests; Extended Data Table 1) to be female (50% versus 48%;  $P = 0.15$ ), older (51.6 versus 50.1 years;  $P < 0.001$ ) and white (82% versus 77%,  $P < 0.001$ ); to have a college education (63% versus 56%;  $P < 0.001$ ); to use X more frequently (5.04 versus 4.85 on an ordinal frequency scale;  $P < 0.001$ ) and to use it more for hard news consumption (73% versus 68%,  $P < 0.001$ ). We verify that the difference in effects between switching the algorithm on and off is not driven by these socio-demographic differences. First, we show that the results are robust to re-weighting observations so that the socio-demographic characteristics of participants initially on the chronological feed match those of participants initially on the algorithmic feed, and vice versa (Extended Data Fig. 5). Second, we predict the probability of initially using the chronological feed based on pre-treatment covariates and show that there is no significant treatment heterogeneity with respect to the predicted initial feed setting, in contrast to the actual feed setting (details of these exercises are provided in Supplementary Information section 2.5). We also verified that the results are not driven by selective attrition by presenting Lee bounds on our estimates (Extended Data Fig. 6 and Supplementary Information section 1.8.2).

Although it is reassuring that the asymmetry between the political effects of turning the algorithm on and off is not explained by selection into the initial feed setting predicted by socio-demographic characteristics, selection on unobserved factors may contribute to this asymmetry. For example, conditional on socio-demographics, users with different traits—such as autonomy or desire for control—could both choose different initial feed settings and react differently to similar content changes.

Alternatively, the asymmetry of the effects could arise directly from an asymmetry in the changes to the content shown when feed settings are altered. In what follows, we examine how changes in feed settings affect the content shown to users and, in turn, how users adjust their online behaviour, while acknowledging that unobserved confounders, such as personality traits, cannot be ruled out.

To study the content, we used data from a Google Chrome extension capturing the first 100 posts each participant would see under both feed settings. We used natural language processing to annotate political leanings and content type (Supplementary Information section 1.5 describes the annotation process and human validation). Figure 3 presents raw summaries of the content shown to users under the two feed settings and reports estimates of the average content difference ( $\Delta$ ) from Poisson pseudo maximum likelihood and ordinary least squares regressions with individual and survey-wave fixed effects, together with  $P$  values from two-sided *t*-tests (Supplementary Information section 2.6 presents specifications and robustness to an alternative machine-learning classifier). First, the algorithm promotes more engaging content. On average, posts shown in the algorithmic feed receive 13,212 likes, 1,836 reposts and 722 comments, compared with 2,781 likes, 453 reposts and 145 comments for posts in the chronological feed. The conditional estimates of  $\Delta$  indicate increases of 480%, 408% and 508% in likes, reposts, and comments, respectively (all  $P < 0.001$ ).

Second, the algorithm promotes political content and, within that category, prioritizes conservative content. Regression results yield that posts annotated as conservative are 2.9 percentage points (19.9%) more likely to appear in the algorithmic feed ( $P < 0.001$ ), whereas posts annotated as liberal are 1.0 percentage point (3.1%) more likely ( $P = 0.043$ ). Restricting to political posts, conservative content is 2.5 percentage points (8%) more likely in the algorithmic feed ( $P < 0.001$ ).

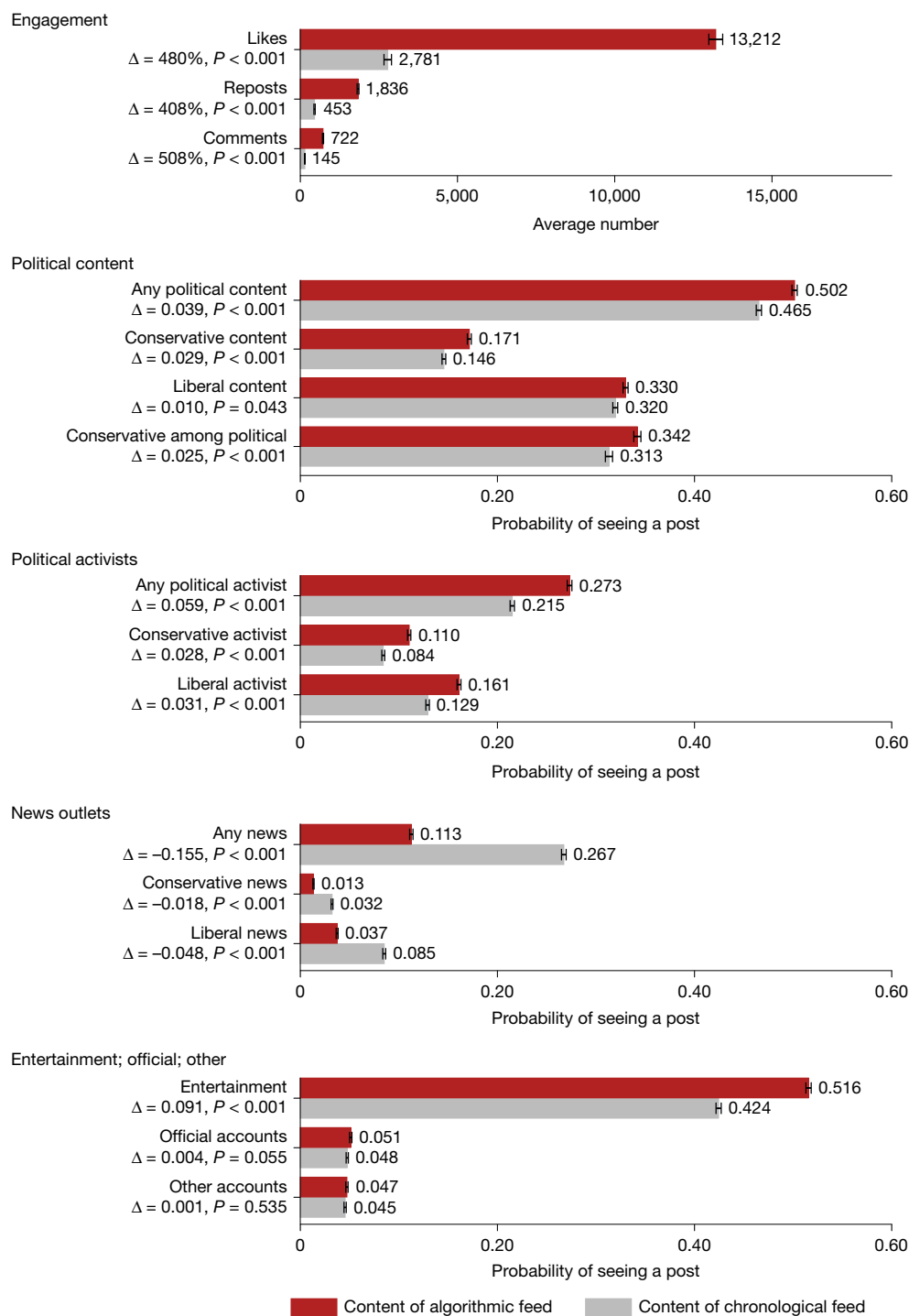
Third, the algorithm demotes accounts of traditional news media and promotes those of political activists. Posts from news organizations appear 15.5 percentage points (58.1%) less often in the algorithmic feed ( $P < 0.001$ ), whereas posts from political activists appear 5.9 percentage points more often (27.4%;  $P < 0.001$ ). Fourth, posts from entertainment accounts are 9.1 percentage points (21.5%) more likely in the algorithmic feed ( $P < 0.001$ ).

Supplementary Information Fig. 2.21 summarizes feed content separately for self-declared Democrats and for self-declared Republicans and Independents. As expected, partisans see substantially less cross-partisan than co-partisan content in their feeds. However, the overall share of conservative content among political content is significantly higher under the algorithmic feed than under the chronological feed for both groups (9.3% versus 5.4%,  $P < 0.001$  for Democrats; 54.2% versus 52.4%,  $P = 0.004$  for Republicans and Independents). This pattern does not hold for content generated by political activists: the algorithm promotes co-partisan activist content more strongly for both groups (Supplementary Information section 2.6).

Next, we use data collected from participants' accounts to analyse the effect of the feed setting on user behaviour. Figure 4 presents the regression results; raw data summaries are presented in Supplementary Information Fig. 2.7. Consistent with the results on political attitudes, switching the algorithm off did not lead to significant differences in the types of account users follow. In contrast, switching the algorithm on did affect the typology of followed accounts among participants initially on the chronological feed. Those exposed to the algorithmic feed as a result of the experiment were more likely to follow conservative accounts (by 0.17 s.d.; 95% CI: 0.03, 0.31;  $P = 0.015$ ), political activist accounts (by 0.13 s.d.; 95% CI: 0.01, 0.25;  $P = 0.036$ ) and, especially, conservative political activist accounts (by 0.18 s.d.; 95% CI: 0.05, 0.32;  $P = 0.010$ ). The entire effect on political activist accounts was driven by conservative political activists. There are no significant differences in the likelihood of following liberal accounts between treatment groups. Following news outlet accounts was also unaffected by the experimental change in the feed setting.

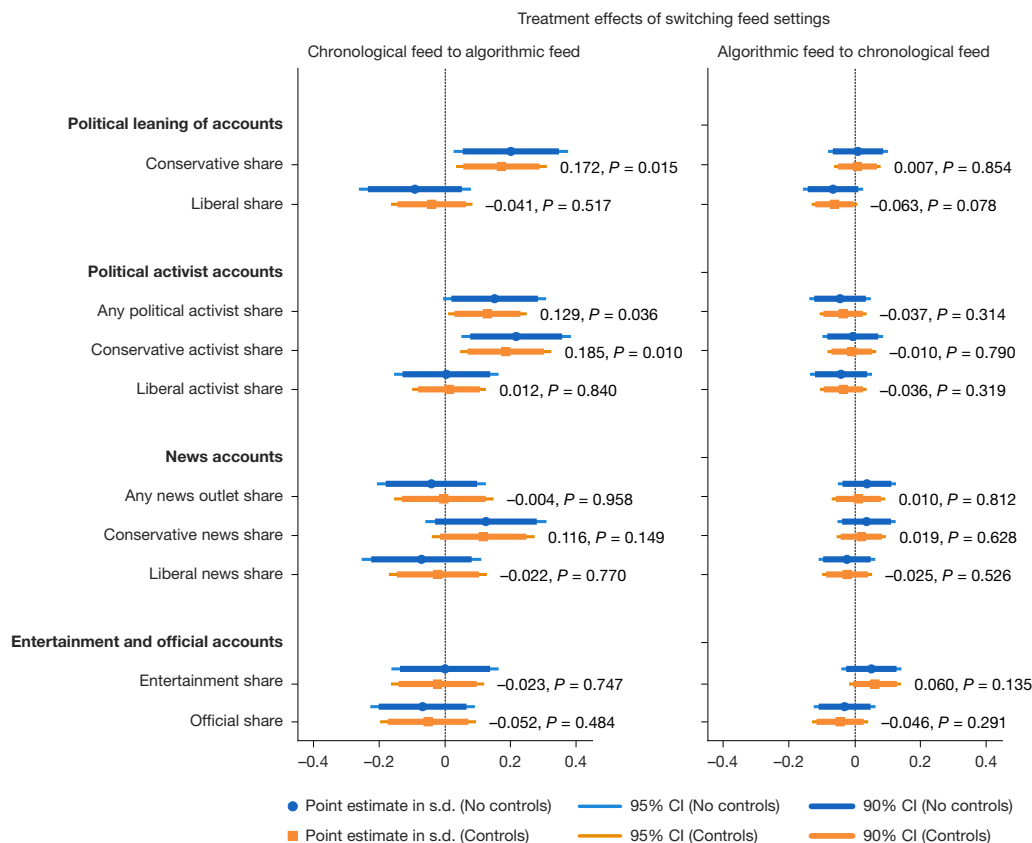
Finally, we provide suggestive evidence on how changes to the feed setting, and the resulting choices of which accounts to follow, shape the content shown to users. Extended Data Fig. 7 reports ITT estimates for the average feed characteristics of the 599 participants who ran the Chrome extension during the post-treatment survey; these results should be interpreted with caution given the small sample. The share of conservative content rose by 0.35 s.d. (95% CI: 0.06, 0.64;  $P = 0.018$ ) when the algorithm was switched on, but did not change significantly when it was switched off, whereas liberal content was unaffected by either switch. This pattern reflected shifts in both activist and news posts. Switching the algorithm on increased posts by conservative activists (0.37 s.d.; 95% CI: 0.05, 0.68;  $P = 0.024$ ) and reduced posts from liberal news outlets ( $-0.43$  s.d., 95% CI:  $-0.71$ ,  $-0.16$ ;  $P = 0.002$ ), whereas turning the algorithm off reduced liberal activists' posts ( $-0.23$  s.d.; 95% CI:  $-0.38$ ,  $-0.08$ ;  $P = 0.004$ ) and increased posts from both liberal (0.50 s.d.; 95% CI: 0.32, 0.67;  $P < 0.001$ ) and conservative news outlets (0.28 s.d.; 95% CI: 0.10, 0.46;  $P = 0.002$ ).

These treatment effects on feed content reflect both what the algorithm promotes (Fig. 3) and changes in the accounts users follow (Fig. 4). Users may be particularly responsive to the latter because these changes arise from their own preferences and behaviour. To isolate the effect of followings, we compare the content users saw, or would have seen, on their chronological feeds depending on treatment assignment in Extended Data Table 2 (methodological details in Supplementary Information section 2.7). Controlling for the intensity of X use before the experiment, the chronological feeds of those exposed to the algorithmic feed setting as a result of the experiment contain 9.0 percentage points (60%) more posts by conservative accounts (95% CI: 2.6, 15.5;  $P = 0.006$ ) and 6.1 percentage points (28%) more posts by conservative political activists (95% CI: 1.0, 11.3;  $P = 0.020$ ) compared with the chronological feeds of those participants who were randomized to stay on the chronological feed. Meanwhile, we found no significant effects on these content characteristics of chronological feeds after switching the algorithm off. We also ran a placebo test, verifying that, in the pre-treatment survey before users were exposed to new content or could follow new accounts, there was no relationship between chronological



**Fig. 3 | Content the algorithm promotes.** Average content shown to users in each feed setting: grey, chronological feed, red, algorithmic feed; 95% CIs around subsample means ( $\text{mean} \pm 1.96 \times \text{s.e.}$ ) are reported. The unit of observation is X post. The number of observations for each feed setting is  $n = 268,532$  for all variables, except for 'Conservative' among 'Political content', for which  $n = 129,790$  X posts in each feed setting. All variables are indicator variables except Likes, Reposts and Comments, which are heavily right-skewed. For the algorithmic feed, the minimum, first quartile, median, third quartile and maximum are 0, 174, 933, 8,197 and 1,000,000 for Likes; 0, 9, 137, 1,363 and 316,700 for Reposts and 0, 5, 56, 384 and 247,200 for Comments. For the chronological feed, the minimum, first quartile, median, third quartile and maximum are 0, 4, 40, 342 and 1,700,000 for Likes; 0, 1, 8, 66 and 572,000 for

Reposts and 0, 0, 3, 21 and 256,100 for Comments. Corresponding summary statistics are presented in Supplementary Information Table 1.8. On the left, for each outcome,  $\Delta$  (the estimated difference in content between the algorithmic and chronological feed from regressions with user and survey-wave fixed effects) and the corresponding  $P$  value are reported. For Likes, Reposts and Comments,  $\Delta$  is expressed in percent and estimated using a Poisson pseudo maximum likelihood model; for all other outcomes,  $\Delta$  is in percentage points and estimated using linear ordinary least squares. Because fixed effects are included,  $\Delta$  does not exactly match the raw ratio or difference in means across feed settings shown in the bar chart. See Supplementary Information section 2.6 for the specification, full regression results and robustness; all tests are two-sided with s.e. values clustered at the respondent level.



**Fig. 4 | ITT estimates of feed-setting changes on accounts that users follow.** ITT effect estimates of switching the algorithm on and off (in s.d.). Left, effect of moving from the chronological to the algorithmic feed for users initially on the chronological feed. Right, effect of moving in the opposite direction for users initially on the algorithmic feed. For each outcome, the results of two specifications are reported. Blue, unconditional estimates with robust s.e. values, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. Orange, conditional estimates, controlling

for pre-treatment covariates using GRFs. Point estimates and 90% and 95% CIs are reported. Numerical effect sizes and *P* values correspond to conditional estimates (all tests are two-sided). The unit of observation is respondent. The sample size is  $n = 2,387$  for all estimates. Tests are described in Methods. Supplementary Information Table 2.16 reports the exact numerical point estimates, s.e. values, CIs and sample sizes for every specification. All outcomes are standardized. Additional results are presented in Supplementary Information section 2.

feed content and treatment assignment (Supplementary Information section 2.7).

### Discussion

In addition to promoting entertainment, X's feed algorithm tends to push more conservative content to users' feeds. Seven weeks of exposure to such content in 2023 shifted users' political opinions in a more conservative direction, particularly with regard to policy priorities, perceptions of the criminal investigations into Trump and views on the war in Ukraine. The effect is asymmetric: switching the algorithm on influenced political views, but switching it off did not reverse users' perspectives on policy priorities or current political issues.

We show that exposure to algorithmically curated content led users to follow conservative activist accounts. In contrast, when the algorithmic feed was switched off, users continued to follow the accounts they had engaged with previously. This indicates that exposure to feed algorithms has a lasting impact on users' feeds and their political attitudes.

This mechanism helps explain why previous research found no political effects of switching the algorithm off on Meta platforms, even though it altered political content<sup>1</sup>. Once users follow accounts promoted by the algorithm that they find engaging and persuasive—such as those of conservative political activists—they continue to see posts from these accounts after turning the algorithm off. Hence, users' political attitudes may remain unchanged, as the content they engage

with most persists despite observed shifts in overall political content. Earlier work also showed that users' feeds are shaped largely by the accounts they follow, even under algorithmic curation<sup>40</sup>. Our findings demonstrate that the algorithm significantly influences whom users choose to follow, indicating that algorithms shape content exposure more than believed previously.

This mechanism may also help to explain why social media deactivation experiments, which completely disconnect users from such content, produce political effects, unlike merely switching off the algorithm while keeping users active on the platform. Facebook deactivation in 2018 reduced political polarization<sup>41</sup>, whereas deactivation during the 2020 US elections reduced self-reported net votes for Trump, although the effect did not meet the authors' pre-registered significance threshold<sup>42</sup>. The latter finding aligns with the political effects observed in our experiment. It is important to note that potential treatment heterogeneity between respondents who initially selected different feed settings may also contribute to the asymmetry in switching the algorithm on and off. Although we control for observed socio-demographics, these groups may still differ along unobserved dimensions.

Additional considerations help clarify the broader implications of our findings. In contrast to policy attitudes and views on current events, partisanship and affective polarization were unaffected by our experimental treatments, consistent with the Meta Study<sup>1</sup>. This may reflect the rigidity and salience of these measures, as they are less susceptible to change over short periods<sup>43</sup>. Despite this rigidity, researcher-designed re-ranking of feeds can affect affective polarization<sup>37</sup>.

Politicized content was a large part of the public conversation on Twitter even before it became X<sup>38</sup> and, in 2024, X stood out among social media platforms as a place people turn to for politics<sup>44</sup>. Further evidence is needed on the effects of algorithm activation on other platforms. Our study has limitations common to research on social media and feed algorithms. The effects are specific to the platform and the time period. For example, they depend on the platform owner's preferences. We find that, relative to the chronological feed, X's algorithm promotes conservative content to both Democrats and to Republicans and Independents, with suggestive evidence that persuasion is stronger for Republicans and Independents, who are more positively predisposed to conservative messages. This indicates that, within the context studied, X's algorithm affects users mainly on one side of the political spectrum—a result that may not hold for algorithms that amplify content from both ideological sides. Another limitation, common to most studies, is that platforms can modify their algorithms at any time, including during the study, without researchers being aware of such changes<sup>33,38,45</sup>. Finally, because we focus on active X users, effect sizes may not generalize to the broader user base, as they are probably smaller for irregular or inactive users, who experience lower treatment intensity.

We conducted the experiment independently of the platform, which helps address additional external validity concerns<sup>35</sup>. However, relative to what could be achieved with the platform's cooperation, our study's independence came at the cost of relying on self-reported compliance for most of the sample, which limited our ability to measure treatment intensity directly, and of working with a smaller sample, which constrained our ability to test finer-grained treatment heterogeneity. Nevertheless, the sample was sufficiently large to detect significant and robust political effects from switching the algorithm on.

Overall, we conclude that social media feed algorithms can play an important role in shaping political opinions and online behaviour.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-026-10098-2>.

1. Guess, A. M. et al. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* **381**, 398–404 (2023).
2. Kemp, S. Digital 2023: Global Overview Report. *DataReportal* <https://datareportal.com/reports/digital-2023-global-overview-report> (2023).
3. Social Media and News Fact Sheet. *Pew Research Center* <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/> (2024).
4. Persily, N. & Tucker, J. A. (eds) *Social Media and Democracy: The State of the Field, Prospects for Reform* (Cambridge Univ. Press, 2020).
5. O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Books, 2016).
6. Castells, M. *Rupture: the Crisis of Liberal Democracy* (Polity, 2018).
7. Guriev, S., Melnikov, N. & Zhuravskaya, E. 3G Internet and confidence in government. *Q. J. Econ.* **136**, 2533–2613 (2021).
8. Rose-Stockwell, T. *Outrage Machine: How Tech Amplifies Discontent, Disrupts Democracy—And What We Can Do About It* (Hachette Books, 2023).
9. Allcott, H. & Gentzkow, M. Social media and fake news in the 2016 election. *J. Econ. Perspect.* **31**, 211–236 (2017).
10. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
11. Bandy, J. & Diakopoulos, N. Curating quality? How Twitter's timeline algorithm treats different types of news. *Social Media Soc.* <https://doi.org/10.1177/205633051211041648> (2021).
12. Castells, M. *Networks of Outrage and Hope: Social Movements in the Internet Age* (Polity, 2015).
13. Brady, W. J., McLoughlin, K., Doan, T. N. & Crockett, M. J. How social learning amplifies moral outrage expression in online social networks. *Sci. Adv.* **7**, eabe5641 (2021).

14. Beknazar-Yuzbashev, G., Jiménez Durán, R., McCrosky, J. & Stalinski, M. Toxic content and user engagement on social media: evidence from a field experiment. *SSRN Res. Paper* <https://doi.org/10.2139/ssrn.4307346> (2022).
15. Pariser, E. *The Filter Bubble: What the Internet is Hiding from You* (Penguin, 2011).
16. Flaxman, S., Goel, S. & Rao, J. M. Filter bubbles, echo chambers, and online news consumption. *Public Opin. Q.* **80**, 298–320 (2016).
17. Settle, J. E. *Frenemies: How Social Media Polarizes America* (Cambridge Univ. Press, 2018).
18. Einav, G., Allen, O., Gur, T., Maaravi, Y. & Ravner, D. Bursting filter bubbles in a digital age: opening minds and reducing opinion polarization through digital platforms. *Technol. Soc.* **71**, 102136 (2022).
19. Zhuravskaya, E., Petrova, M. & Enikolopov, R. Political effects of the Internet and social media. *Annu. Rev. Econ.* **12**, 415–438 (2020).
20. Kubin, E. & von Sikorski, C. The role of (social) media in political polarization: a systematic review. *Ann. Int. Commun. Assoc.* **45**, 188–206 (2021).
21. Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C. & Sternisko, A. How social media shapes polarization. *Trends Cogn. Sci.* **25**, 913–916 (2021).
22. Aridor, G., Jiménez Durán, R., Levy, R. & Song, L. The economics of social media. *J. Econ. Lit.* **62**, 1422–1474 (2024).
23. Epstein, R. & Robertson, R. E. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proc. Natl Acad. Sci. USA* **112**, E4512–E4521 (2015).
24. Robertson, R. E. et al. Auditing partisan audience bias within Google Search. *Proc. ACM Hum-Comput Interact* **2** (CSCW) <https://doi.org/10.1145/3274417> (2018).
25. Courtois, C., Slechten, L. & Coenen, L. Challenging Google Search filter bubbles in social and political information: disconfirming evidence from a digital methods case study. *Telemat. Inform.* **35**, 2006–2015 (2018).
26. Hosseinmardi, H. et al. Causally estimating the effect of YouTube's recommender system using counterfactual bots. *Proc. Natl Acad. Sci. USA* **121**, e2313377121 (2024).
27. Ulrich, M. & Hodler, R. Web search personalization during the US 2020 election. *Am. Econ. Rev. Insights* **7**, 516–533 (2025).
28. González-Bailón, S. et al. Asymmetric ideological segregation in exposure to political news on Facebook. *Science* **381**, 392–398 (2023).
29. Guess, A. M. et al. Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science* **381**, 404–408 (2023).
30. Nyhan, B. et al. Like-minded sources on Facebook are prevalent but not polarizing. *Nature* **620**, 137–144 (2023).
31. Garrett, R. K. & Resnick, P. Resisting political fragmentation on the Internet. *Daedalus* **140**, 108–120 (2011).
32. Helberger, N. in *Algorithms, Automation, and News* (eds Thurman, N. et al.) 14–33 (Routledge, 2021).
33. Knudsen, E. Modeling news recommender systems' conditional effects on selective exposure: evidence from two online experiments. *J. Commun.* **73**, 138–149 (2023).
34. Mattis, N., Masur, P., Möller, J. & van Atteveldt, W. Nudging towards news diversity: a theoretical framework for facilitating diverse news consumption through recommender design. *New Media Soc.* **26**, 3681–3706 (2024).
35. Wagner, M. W. Independence by permission. *Science* **381**, 388–391 (2023).
36. Thorp, H. H. & Vinson, V. Context matters in social media. *Science* **385**, 1393–1393 (2024).
37. Piccardi, T. et al. Reranking partisan animosity in algorithmic social media feeds alters affective polarization. *Science* **390**, eadu5584 (2025).
38. Huszár, F. et al. Algorithmic amplification of politics on Twitter. *Proc. Natl Acad. Sci. USA* **119**, e2025334119 (2022).
39. Athey, S., Tibshirani, J. & Wager, S. Generalized random forests. *Ann. Statist.* **47**, 1148–1178 (2019).
40. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
41. Allcott, H., Braghieri, L., Eichmeyer, S. & Gentzkow, M. The welfare effects of social media. *Am. Econ. Rev.* **110**, 629–676 (2020).
42. Allcott, H. et al. The effects of Facebook and Instagram on the 2020 election: a deactivation experiment. *Proc. Natl Acad. Sci. USA* **121**, e2321584121 (2024).
43. Zaller, J. R. *The Nature and Origins of Mass Opinion* (Cambridge Univ. Press, 1992).
44. How Americans navigate politics on TikTok, X, Facebook and Instagram. *Pew Research Center* <https://www.pewresearch.org/internet/2024/06/12/how-americans-navigate-politics-on-tiktok-x-facebook-and-instagram/> (2024).
45. Bagchi, C. et al. Social media algorithms can curb misinformation, but do they? *Science eLetter* <https://www.science.org/doi/10.1126/science.abp9364#lettersSection> (2024).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

## Methods

### Participants

Participants in our experiment were recruited by YouGov—a professional online polling firm with a large panel of respondents in the United States. All study participants gave their informed consent to participate. Participants were included in the study only if they self-reported using X at least several times a month. The sample is 78% white, 52% male and relatively well educated, with 58% having completed at least 4 years of university. In terms of political affiliation, 46% identify as Democrats and 21% as Republicans. Among the participants, 66% use X at least once a day and 94% at least once a week. As for posting activity, 27% post at least once a day and 53% at least once a week. See Supplementary Information section 1.4 for detailed summary statistics and the comparison of our sample with Twitter users from the nationally representative Social Media Study by the American National Election Studies.

### Experimental design

Participants were randomized into an algorithmic or chronological feed for the duration of the experiment in exchange for compensation. Supplementary Information section 1.2 provides information on the compensation scheme and on how different elements of the experiment appeared on the X platform.

The randomization procedure was effective: participants assigned to different feed conditions exhibited no systematic differences in demographic characteristics or social media usage beyond what would be expected by chance. The only notable imbalance relates to the initial feed setting. The proportion of participants who were already using the algorithmic feed at baseline was two percentage points higher among those assigned to the algorithmic group (77% compared with 75%). All analyses account for the initial feed setting. Supplementary Information section 1.7 provides further details on balance and reports summary statistics by treatment group and initial feed setting.

### Data on outcomes

Outcome data on attitudes come from the post-treatment survey. The coding of the survey-based outcome variables is presented in Supplementary Information section 1.3. To analyse the content of users' feeds and the accounts they follow, we applied natural language processing methods. The Llama 3-based classification of feed content, collected using a Google Chrome extension, categorizes posts by political leaning—either conservative or liberal—and by type, distinguishing between posts from political activists, entertainment accounts and news media outlets. Details are provided in Supplementary Information section 1.5. The Llama 3-based classification of followed accounts is based on data collected through users' X handles. This procedure is described in Supplementary Information section 1.6.

We conduct several validation exercises for the Llama 3 annotations in Supplementary Information sections 1.5 and 1.6. These include comparisons with machine-learning classifiers based on word frequencies, as well as evaluations conducted by human annotators.

### ITT effects estimates

Given the experimental design, the ITT effect estimates were obtained by comparing mean outcomes between respondents assigned randomly to the algorithmic feed and those assigned to the chronological feed, conditional on their initial feed setting. Our main specification estimates the effect of switching the algorithm on for users who initially had it off, and switching it off for users who initially had it on. Specifically, we estimate the following model:

$$Y_i = \alpha + \beta_1 \text{Initial Chrono}_i \times \text{Treatment Algo}_i + \beta_2 \text{Initial Algo}_i \times \text{Treatment Chrono}_i + \beta_3 \text{Initial Algo}_i + \mathbf{X}'_i \boldsymbol{\Gamma} + \epsilon_i \quad (1)$$

where  $i$  indexes respondents.  $Y_i$  represents outcome variables (for example, post-treatment political attitudes of respondents). Initial  $\text{Algo}_i$  is a dummy variable for the initial feed setting; it equals one for respondents who had been using the algorithmic feed before our intervention and zero for those who had been using the chronological feed. Treatment  $\text{Algo}_i$  is a dummy variable for being assigned to the algorithmic feed setting during the treatment phase; it equals 1 for respondents assigned to the algorithmic feed and 0 for those assigned to the chronological feed. Note that, by construction, Initial  $\text{Chrono}_i = 1 - \text{Initial Algo}_i$  and Treatment  $\text{Chrono}_i = 1 - \text{Treatment Algo}_i$ .  $\mathbf{X}_i$  is a vector of control variables discussed below.

The coefficients of interest are  $\beta_1$  and  $\beta_2$ , which estimate the ITT effects of switching the algorithm on and switching it off, respectively. Because we control for Initial  $\text{Algo}$ , those switching from the chronological to the algorithmic feed are compared with users who initially had the chronological feed and remained on it, whereas those switching from the algorithmic to the chronological feed are compared with users who initially had the algorithmic feed and remained on it.

We consistently report two sets of estimates. First, we report unconditional estimates, namely, controlling only for the initial feed setting and, when available, the pre-treatment value of the outcome variable. The unconditional model is justified by the balance in observables across treatment groups. Second, we also report estimates that flexibly control for pre-treatment covariates. Our baseline approach relies on GRFs—a machine-learning method that identifies the most relevant covariates and adjusts for them non-parametrically<sup>39</sup>. This approach limits the researcher's discretion in selecting covariates and functional form, including mitigating concerns about model misspecification<sup>46</sup>. We use the full set of available pre-treatment covariates as input to the GRFs. The covariates include initial feed setting, gender, age, indicators of X use, all categories of educational attainment and race, frequency of X use and posting, political affiliation, the device used to complete the pre-treatment survey (such as a laptop or smartphone), life satisfaction, happiness and affective polarization. Technical details are provided in Supplementary Information section 2.2.1. Baseline ITT estimates are reported in Fig. 2.

Throughout the analysis (in the main text and Extended Data Figs. 4, 5 and 7 and Supplementary Information), we present ITT estimates in graphical form using coefficient plots. Estimates from the unconditional specification are shown in blue, whereas those from the specification controlling for pre-treatment covariates are shown in orange. For clarity, we refer to  $\beta_1$  estimates as 'Chrono to Algorithm' and  $\beta_2$  estimates as 'Algorithm to Chrono' in these coefficient plot figures.

In Supplementary Information section 2.4, we report results for all outcome variables using a range of linear specifications, progressively adding the following controls to equation (1): demographic characteristics (gender, age, indicators for being white and for high educational attainment), pre-treatment X use (frequency of use, frequency of posting and self-reported purpose of using X), and political affiliation (Republican, Independent, Democrat or other). We also present results from a specification that includes the predicted probability of having an algorithmic initial feed, obtained using LASSO on all pre-treatment characteristics. In addition, we report results from a specification that controls for post-treatment X use. Although post-treatment use is endogenous, robustness to including this control indicates that the estimated effects are not solely driven by changes in X usage.

In Supplementary Information section 2.4, we also present tests for the symmetry of the effects of switching the algorithm on and off. Specifically, we test whether  $\beta_1 = -\beta_2$ , and whether 7 weeks of algorithm exposure, induced by treatment assignment for users initially on the chronological feed, leads to outcome levels similar to those of users who were initially on the algorithmic feed and remained on it, that is, whether  $\beta_1 = \beta_3$ .

## Compliance and LATE estimates

Due to minimal non-compliance with assigned treatments (Extended Data Fig. 2 and Supplementary Information section 1.8.1), the ITT estimates do not strictly represent average treatment effects. We estimate LATE for compliers, that is, for respondents who adhered to their assigned feed setting, using instrumental variables regressions, with the assigned feed setting used as an instrument for the feed setting actually used.

We define ‘actual’ feed use during the treatment phase using treatment assignments and respondents’ self-reported compliance from the post-treatment survey. Respondents are classified as compliant if they answered ‘always’ or ‘most of the time’ to the question, ‘Did you stick to the assigned feed?’ Based on this, we construct two variables: Actual Algo<sub>*i*</sub> and Actual Chrono<sub>*i*</sub>.

Actual Algo<sub>*i*</sub> equals 1 if respondent *i* was compliant and assigned to the algorithmic feed, or non-compliant and assigned to the chronological feed, implying they used the algorithmic feed. Conversely, Actual Chrono<sub>*i*</sub> equals 1 if the respondent was compliant and assigned to the chronological feed, or non-compliant and assigned to the algorithmic feed. By construction, Actual Chrono<sub>*i*</sub> = 1 – Actual Algo<sub>*i*</sub>.

We estimate two-stage least squares regressions, with the second stage specified as follows:

$$Y_i = \alpha + \gamma_1 \text{Hat}(\text{Initial Chrono}_i \times \text{Actual Algo}_i) + \gamma_2 \text{Hat}(\text{Initial Algo}_i \times \text{Actual Chrono}_i) + \gamma_3 \text{Initial Algo}_i + \mathbf{X}'_i \boldsymbol{\Gamma} + \epsilon_i \quad (2)$$

Hat() denotes the fitted value from the corresponding first-stage regression. The two first-stage regressions predict the following:

Hat(Initial Chrono<sub>*i*</sub> × Actual Algo<sub>*i*</sub>) with Initial Chrono<sub>*i*</sub> × Treatment Algo<sub>*i*</sub> and Hat(Initial Algo<sub>*i*</sub> × Actual Chrono<sub>*i*</sub>) with Initial Algo<sub>*i*</sub> × Treatment Chrono<sub>*i*</sub>. With a compliance rate of 85.38%, the instruments are extremely strong: in all specifications, the *F*-statistics for the excluded instruments in the first stage exceed 1,000.

In equation (2),  $\gamma_1$  represents the LATE estimate of the effect of switching the algorithm on, and  $\gamma_2$  the corresponding effect of switching it off. The main LATE estimates are reported in Extended Data Fig. 3. Given the high compliance rate, the ITT and LATE estimates are very similar.

In Supplementary Information Fig. 2.13, we also report LATE estimates using a more conservative definition of self-reported compliance, in which a respondent is considered compliant only if they answered ‘always’ to the question ‘Did you stick to the assigned feed?’ For those who answered ‘almost always’, we assume that they complied with the assigned setting at any moment only with probability 0.5. In this case, the instruments remain very strong and the results are very similar.

The results are also robust to restricting the analysis to respondents whose compliance was confirmed through observation, that is, those who installed the Google Chrome extension and remained on their assigned feed setting (Supplementary Fig. 2.12).

## Attrition and Lee bounds estimates

Significant attrition was observed between the pre-treatment and post-treatment surveys, common to studies like ours. Attrition rates did not differ across treatment groups. We formally assess whether selective attrition could bias our results using Lee bounds and conclude that attrition does not meaningfully affect our findings. Details are provided in Supplementary Information section 1.8.2, and results are reported in Extended Data Fig. 6.

## Ethics statement

Ethical approval for this research was granted by the Ethics Committee of the University of St. Gallen, Switzerland. All procedures involving

human participants were performed in accordance with the institutional guidelines and regulations. Informed consent was obtained from all participants at the time of their enrolment. Participants received compensation for their time: they were compensated in points—a currency used within our implementation partner’s (YouGov) platform, with an exchange rate of 1,000 points equalling US\$1.00. For the pre-treatment survey, they received 500 points (US\$0.50), with clear information that they would earn 2,500 points (US\$2.50) for using the assigned feed setting between the two surveys and returning for the post-treatment survey. Furthermore, participants could earn an extra 2,000 points (US\$2.00) for using the Chrome extension during the pre-treatment survey. After completing the post-treatment survey, they were paid the 2,500 points (US\$2.50) that had been announced during the pre-treatment survey. Again, they could earn additional compensation for running the Chrome extension, this time 2,500 points (US\$2.50). Therefore, participants could earn up to 10,000 points (US\$10.00) if they completed both surveys and used the Chrome extension both pre- and post-treatment. Sharing the X handle was not incentivized.

The experimental interventions were limited to the choice of feed setting, which is freely available to every user on the X platform. No artificial manipulation of users’ content was introduced. Additional information on ethical safeguards and considerations is provided in Supplementary Information section 1.1.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The replication data are available at Figshare (<https://doi.org/10.6084/m9.figshare.28033772>)<sup>47</sup>. Due to privacy concerns and Institutional Review Board requirements, the following identifiable information has been excluded: the respondents’ X handles, the text of the posts in their feeds and the lists of accounts they follow. The repository contains anonymized survey data, all variables derived from identifiable information, as well as all codes and protocols necessary for replication. Source data are provided with this paper.

## Code availability

This study uses publicly and commercially available software: Python (open source), R (open source) and Stata (available through StataCorp LLC with a license). Details about the code used are provided at Figshare (<https://doi.org/10.6084/m9.figshare.28033772>)<sup>47</sup>.

46. King, G. & Roberts, M. E. How robust standard errors expose methodological problems they do not fix, and what to do about it. *Polit. Anal.* **23**, 159–179 (2015).

47. Gauthier, G., Hodler, R., Widmer, P. & Zhuravskaya, E. The political effects of X’s feed algorithm. *Figshare* <https://doi.org/10.6084/m9.figshare.28033772> (2026).

**Acknowledgements** We thank J. Reinhold for outstanding research assistance. This research was funded by the Swiss National Science Foundation (SNSF), grant no. 215554.

**Author contributions** G.G., R.H., P.W. and E.Z. jointly conceived the study, designed the experiment, developed the methodology and acquired funding. G.G., R.H., P.W. and E.Z. contributed equally to data collection, analysis and interpretation of results. G.G., R.H., P.W. and E.Z. participated in writing the original draft and critically revised the manuscript. G.G., R.H., P.W. and E.Z. proofread and approved the final version.

**Competing interests** The authors declare no competing interests.

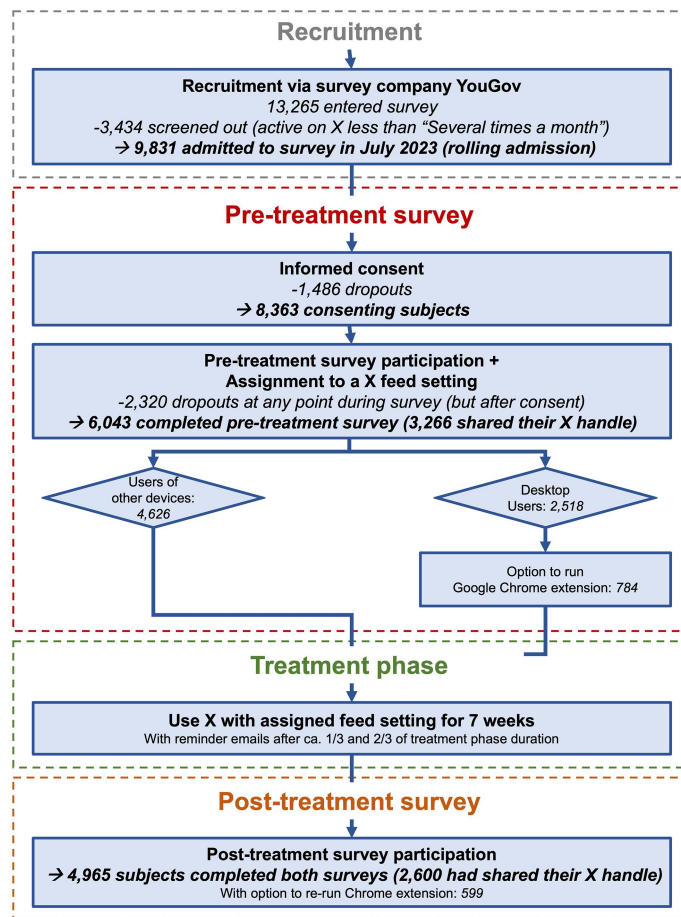
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-026-10098-2>.

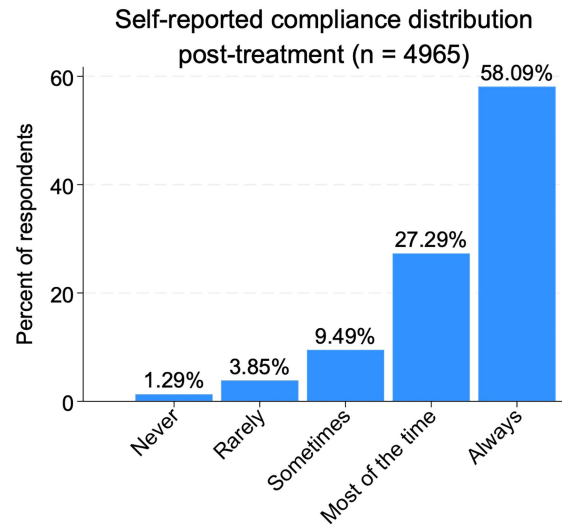
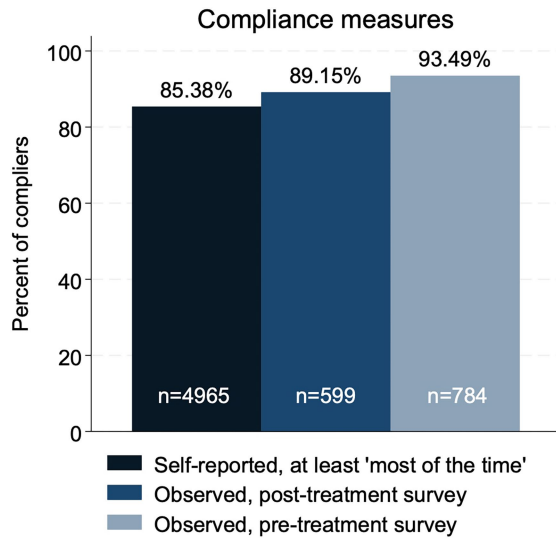
**Correspondence and requests for materials** should be addressed to Ekaterina Zhuravskaya.

**Peer review information** Nature thanks Ceren Budak, Erik Knudsen, Jennifer Pan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



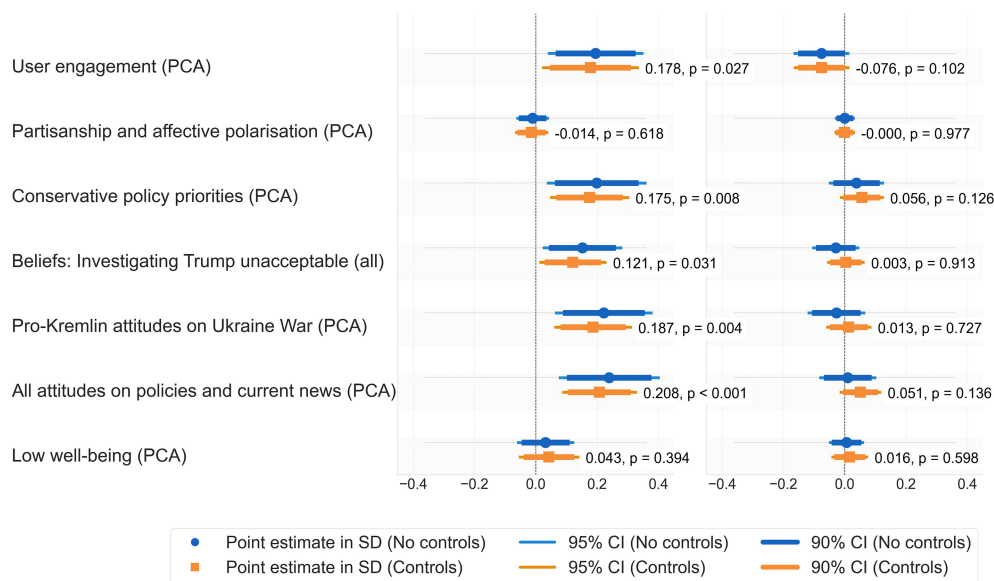
Extended Data Fig. 1 | Flow Diagram for the Experiment.



**Extended Data Fig. 2 | Compliance.** On the left, mean compliance by measure of observed and self-reported compliance. Observed compliance is based on the Chrome extension, which provides information on which feed setting respondents are on when using the extension. Self-reported compliance is based

on a question in the post-treatment survey asking the respondents whether they complied with the assigned treatment. On the right, distribution of values for self-reported compliance. Sample size indicated on the figure. The unit of observation is respondent.

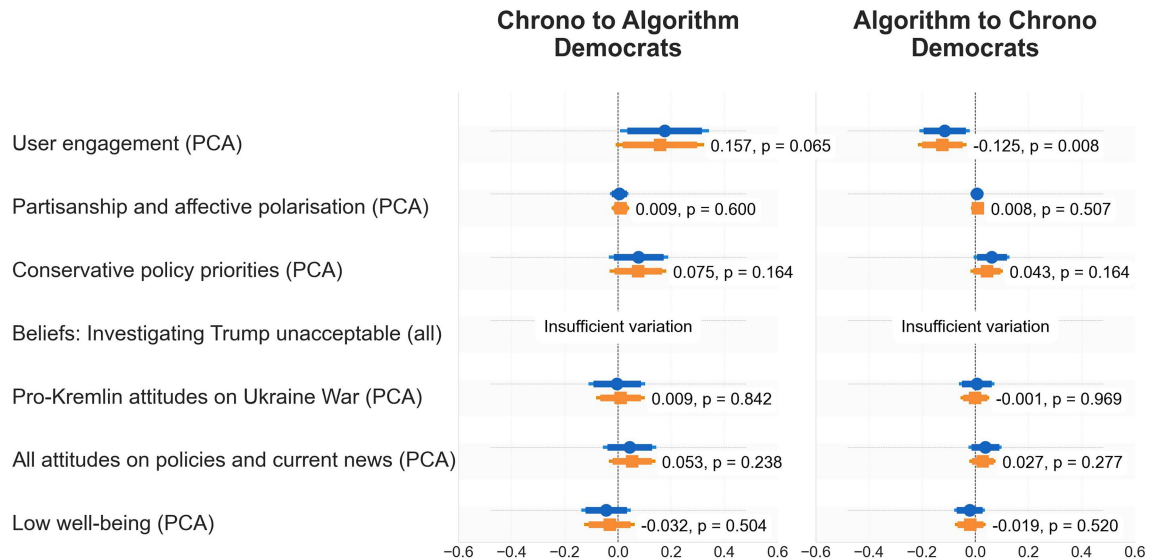
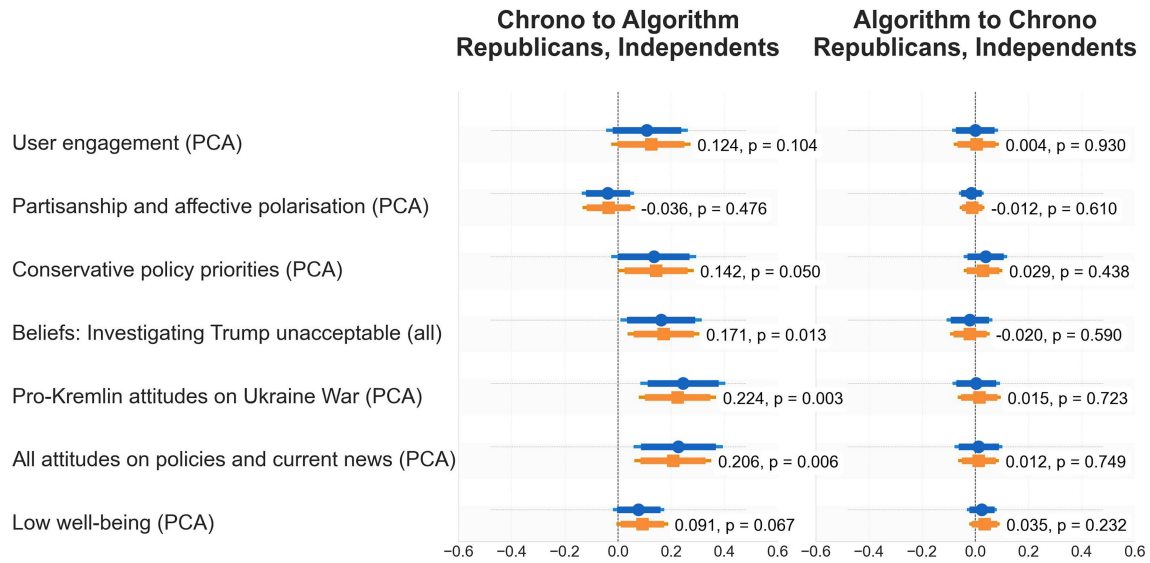
Treatment Effects of Switching Feed Settings  
 Chrono to Algorithm      Algorithm to Chrono  
 Local Avg. Treatment Effect      Local Avg. Treatment Effect



**Extended Data Fig. 3 | Engagement and Political Attitudes. LATE Estimates by Initial Feed Setting.** Local Average Treatment Effect (LATE) estimates of the effect of switching the algorithm on (left panel) and switching it off (right panel) in SD. The left panel: the effect of moving from the chronological to the algorithmic feed for users initially on the chronological feed. Right panel: the effect of moving in the opposite direction for users initially on the algorithmic feed. “Chrono” and “Algorithm” refer to the chronological and algorithmic feeds, respectively. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment

outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). 90% and 95% CIs are reported. Numerical effect sizes and *p*-values printed on the figure correspond to the conditional estimates (all tests are two-sided). The unit of observation is respondent. From top to bottom, sample sizes are *n* = 4965, *n* = 3337, *n* = 4965, *n* = 4965, *n* = 4596, *n* = 4596, and *n* = 4850. Methods describe the tests. SI Table S2.16 reports the exact numerical point estimates, standard errors, confidence intervals and sample sizes for every specification. All outcomes are standardized.

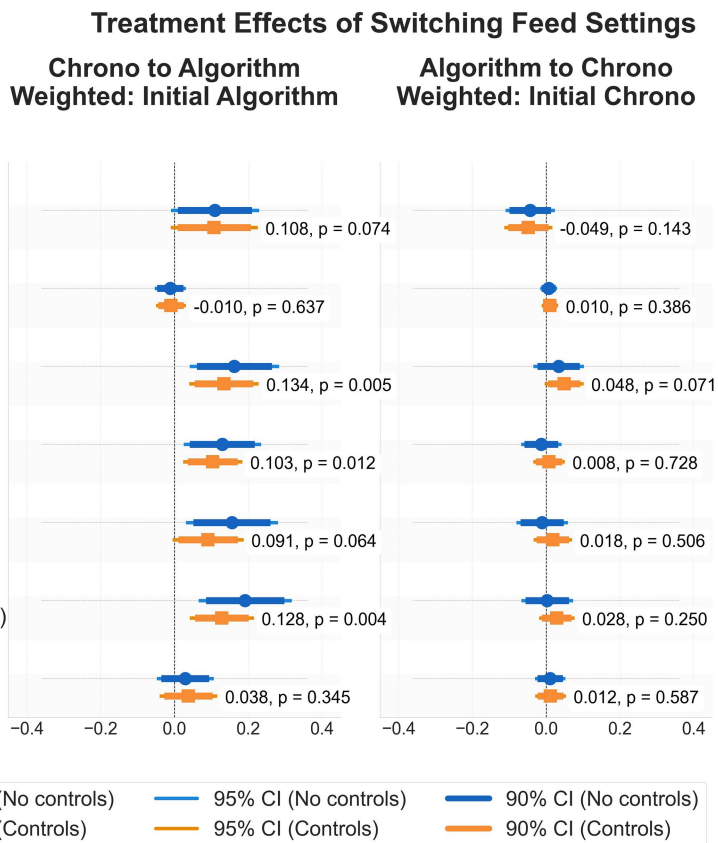
## Treatment Effects of Switching Feed Settings



● Point estimate (No controls)    — 95% CI (No controls)    — 90% CI (No controls)  
■ Point estimate (Controls)    — 95% CI (Controls)    — 90% CI (Controls)

**Extended Data Fig. 4 | Engagement and Political Attitudes. ITT Estimates by Initial Feed Setting: Separately for Republicans and Independents vs. Democrats.** Replication of main-text Fig. 2 separately in the subsamples of (1) self-declared Republicans and Independents and (2) self-declared Democrats. Intention-to-treat (ITT) effect estimates in SD. Left panel: ITT estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. “Chrono” and “Algorithm” refer to the chronological and algorithmic feeds, respectively. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and, where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). 90% and 95% CIs are reported.

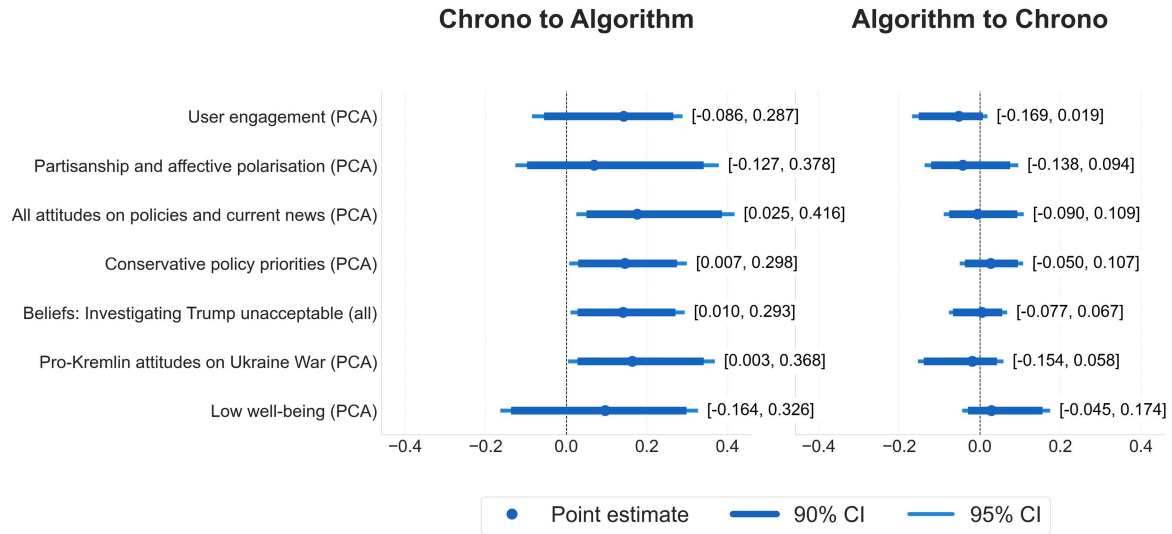
Numerical effect sizes and *p*-values printed on the figure correspond to the conditional estimates (all tests are two-sided). The unit of observation is respondent. From top to bottom, sample sizes are  $n = 2659$ ,  $n = 1059$ ,  $n = 2659$ ,  $n = 2659$ ,  $n = 2452$ ,  $n = 2452$ , and  $n = 2596$  for Republican and Independent respondents;  $n = 2306$ ,  $n = 2278$ ,  $n = 2306$ ,  $n = 2306$ ,  $n = 2144$ ,  $n = 2144$ , and  $n = 2255$  for Democratic respondents. Methods describe the tests. SI Table S2.16 reports the exact numerical point estimates, standard errors, confidence intervals and sample sizes for every specification. Extremely few Democrats think that investigating Trump is unacceptable. All outcomes are standardized. ITT effect estimates separately for Republicans and Independents are shown in SI Fig. S2.11.



**Extended Data Fig. 5 | Engagement and Political Attitudes. ITT Estimates by Initial Feed Setting: Re-weighted to Address Selection into Initial Feed.** Replication of main-text Fig. 2 with subsample of respondents initially on the chronological feed re-weighted to mimic the subsample initially on the algorithmic feed, and vice versa. Intention-to-treat (ITT) effect estimates in SD. Left panel: ITT estimates of switching the algorithm on. Right panel: ITT estimates of switching the algorithm off. “Chrono” and “Algorithm” stand for the chronological and algorithmic feeds, respectively. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting and,

where applicable, pre-treatment outcome levels. In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). 90% and 95% CIs are reported. Numerical effect sizes and *p*-values printed on the figure correspond to the conditional estimates (all tests are two-sided). The unit of observation is respondent. From top to bottom, sample sizes are *n* = 4965, *n* = 3337, *n* = 4965, *n* = 4965, *n* = 4596, *n* = 4596, and *n* = 4850. Methods describe the tests. SI Table S2.16 reports the exact numerical point estimates, standard errors, confidence intervals and sample sizes for every specification. All outcomes are standardized.

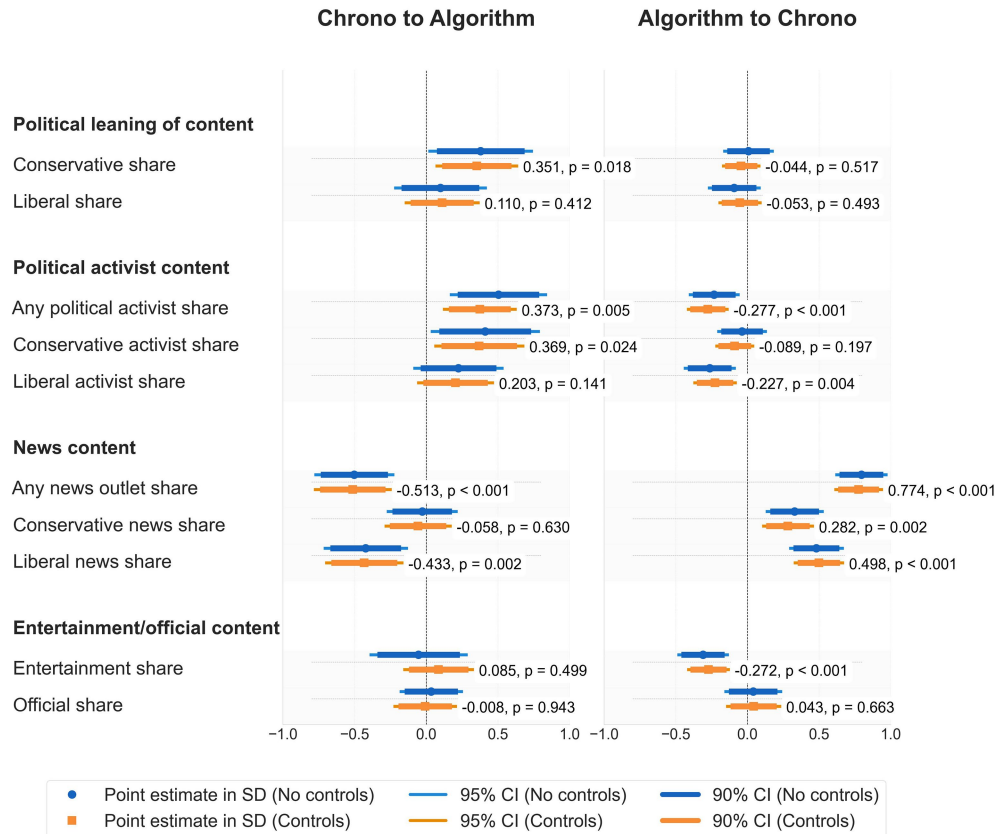
## Treatment Effects of Switching Feed Settings: Lee Bounds



**Extended Data Fig. 6 | Engagement and Political Attitudes. Selective Attrition Does Not Drive Results: Lee Bounds of ITT Estimates by Initial Feed Setting.** Lee bounds for the unconditional point estimates of switching the algorithm on (left panel) and switching it off (right panel). The left panel: estimates for respondents with chronological feed as an initial setting and who were assigned to the algorithm. The right panel: estimates for respondents with the algorithm as an initial feed setting and who were assigned to the chronological feed. “Chrono”

and “Algorithm” stand for the chronological and algorithmic feeds, respectively. 90% and 95% CIs are reported graphically and 95% CIs printed numerically. The unit of observation is respondent. From top to bottom, sample sizes are  $n = 4965$ ,  $n = 3337$ ,  $n = 4965$ ,  $n = 4965$ ,  $n = 4596$ ,  $n = 4596$ , and  $n = 4850$ . Methods describe the tests. SI Table S2.16 reports the exact numerical estimates and sample sizes for every specification. All outcomes are standardized. A similar figure for the accounts followed by users is presented in SI Fig. S1.15.

Treatment Effects of Switching Feed Settings



**Extended Data Fig. 7 | Content that Users See: ITT Estimates by Initial Feed Setting.** Intention-to-treat (ITT) effect estimates of switching the algorithm on (left panel) and switching it off (right panel) in SD. Left panel: the effect of moving from the chronological to the algorithmic feed for users initially on the chronological feed. Right panel: the effect of moving in the opposite direction for users initially on the algorithmic feed. “Chrono” and “Algorithm” refer to the chronological and algorithmic feeds, respectively. For each outcome, the results of two specifications are reported. In blue: the unconditional estimates with robust standard errors, controlling only for the initial feed setting.

In orange: the conditional estimates, controlling for pre-treatment covariates using Generalized Random Forests (GRFs). 90% and 95% CIs are reported. Numerical effect sizes and *p*-values printed on the figure correspond to the conditional estimates (all tests are two-sided). The unit of observation is respondent. The sample size is *n* = 599 for all outcomes. Methods describe the tests. SI Table S2.16 reports the exact numerical point estimates, standard errors, confidence intervals and sample sizes for every specification. All outcomes are standardized.

## Extended Data Table 1 | Summary Statistics by Initial Feed Setting

Variable	Initial (Pre-Treatment) Feed Setting		
	Chronological	Algorithm	
	Mean (Standard Deviation)		
<b>General Information</b>			
Gender	Share of Males	0.50 (0.50)	0.52 (0.50)
Race	Share of Whites	0.82 (0.39)	0.77 (0.42)
Education	Share of 4-year college or above	0.63 (0.48)	0.56 (0.50)
Age		51.62 (14.25)	50.06 (15.14)
<b>X Use</b>			
X use frequency		5.04 (1.18)	4.85 (1.24)
X posting frequency		2.89 (2.07)	2.94 (2.07)
News consumption	Share of soft news consumption	0.63 (0.48)	0.61 (0.49)
	Share of hard news consumption	0.73 (0.44)	0.68 (0.47)
<b>Well-being</b>			
Life Dissatisfaction		4.53 (2.11)	4.61 (2.11)
Unhappiness		2.05 (0.70)	2.04 (0.72)
<b>Political Attitudes</b>			
Affective polarisation		34.46 (31.01)	38.67 (31.58)
Partisanship	Share of Democrats	0.47 (0.50)	0.46 (0.50)
	Share of Independents	0.30 (0.46)	0.28 (0.45)
	Share of Republicans	0.19 (0.39)	0.22 (0.42)
	Share of Some other party	0.03 (0.18)	0.03 (0.18)
<b>Number of Observations</b>		1206	3759

Means and SD (in parentheses) for pre-treatment characteristics by initial feed settings.

# Article

## Extended Data Table 2 | Accounts Shown in Chronological Feed: Llama 3 Annotations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Probability of seeing account in chronological feed that is:								
	Conservative			Political Activist			Conservat. Polit. Activist		
$\beta_1$ : Initial Chrono	0.082**	0.090***	0.066**	0.031	0.046	0.039	0.054**	0.061**	0.047*
× Treatment Algo	(0.034)	(0.033)	(0.030)	(0.034)	(0.030)	(0.030)	(0.027)	(0.026)	(0.025)
$\beta_2$ : Initial Algo	0.017	0.024	0.020	-0.009	0.003	0.000	0.012	0.018	0.016
× Treatment Chrono	(0.022)	(0.021)	(0.018)	(0.020)	(0.018)	(0.017)	(0.015)	(0.015)	(0.013)
Placebo $\delta_1$ : Init. Chrono	-0.002	-0.020	-0.009	0.039	0.009	0.010	-0.007	-0.020	-0.013
× Treat. Algo Pre-Treat	(0.040)	(0.038)	(0.030)	(0.034)	(0.028)	(0.025)	(0.027)	(0.026)	(0.020)
Placebo $\delta_2$ : Init. Algo	0.008	0.002	-0.001	0.010	-0.002	-0.003	0.011	0.007	0.005
× Treat. Chrono Pre-Treat	(0.021)	(0.021)	(0.018)	(0.019)	(0.018)	(0.017)	(0.015)	(0.015)	(0.013)
Initial Algo	0.023	0.024	-0.001	0.002	-0.004	-0.007	0.000	0.001	-0.016
	(0.03)	(0.03)	(0.025)	(0.027)	(0.023)	(0.021)	(0.021)	(0.021)	(0.018)
Mean dependent variable	0.15	0.15	0.15	0.22	0.22	0.22	0.22	0.22	0.22
Observations	134706	134706	134706	134706	134706	134706	134706	134706	134706
X Use Controls		√	√		√	√		√	√
Demogr. & Polit. Controls			√			√			√

Regression estimates of the intention-to-treat (ITT) and placebo effects on the chronological feed's political content measured using Llama 3 annotations (see SI Section S1.5.1 for details of measures and SI Section 2.7 for the specification, namely, equation 10). The unit of observation is X post. Sample size is reported in the Table. Col. 1, 4, and 7 report estimates without controls, examining effects on Conservative content (1), political activist content (4), and Conservative political activist content (7). Col. 2, 5, and 8 add controls for pre-treatment X use frequency and purpose. Col. 3, 6, and 9 further include X use, demographic, and political controls. All specifications control for respondents' initial feed settings. For each outcome, coefficients  $\beta_1$  and  $\beta_2$  capture content changes in the chronological feed between pre- and post-treatment periods, while  $\delta_1$  and  $\delta_2$  serve as placebo estimates validating the randomization. The sample is restricted to posts shown in the chronological feed setting. Standard errors are clustered by respondent. \*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% level, respectively (all tests are two-sided).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** For the survey data, no software was used by the researchers. We worked with the survey company YouGov. To collect the X data, we used a custom Chrome extension and Python 3.12.3.

**Data analysis** We use Python 3.12.3, R4.3.3 (with the package grf version 2.4.0), and Stata 19 (with the package leebounds version 1.5). All Python and R packages used are open-source and available under their respective licenses. Stata 19 is a proprietary software from StataCorp LLC under a commercial license. The full list of packages used for each software is provided in the replication folder.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The replication data is available in the Figshare repository (<https://doi.org/10.6084/m9.figshare.28033772>). We notably provide all anonymized survey, newsfeeds,

and followings data from our experiment. To comply with IRB directives and respect the privacy of our respondents, we deleted all identifiable information. We notably deleted: respondents' X (formerly Twitter) handles, the text of the posts and the name of the accounts shown to them in the newsfeeds, and the lists of the accounts they followed. In addition to our own data collection, we used five external datasets: a dataset of all US cities, states, and counties (Source: <https://github.com/grammakov/USA-cities-and-states>); a dataset of news media bias (Source: R. E. Robertson, et al., Auditing partisan audience bias within Google Search. Proceedings of the ACM on Human-Computer Interaction 2 (CSCW), 148 (2018), doi:10.1145/3274417); a dataset containing the universe of posts by congressmen on X between January 2023 and July 2023 (Source: <https://alexlitel.github.io/congresstweets/>); a dataset containing the universe of X handles of news media outlets (Source: U. Matter, P. Widmer, Who Owns the Online Media?, Tech. rep., SSRN (2021).1484); and the ANES Social Media Study Dataset (Source: <https://electionstudies.org/data-center/2020-2022-social-media-study/>). These external datasets are also included in the replication repository.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

### Reporting on sex and gender

This study collected self-reported gender information through YouGov's survey panel (male, female, or prefer not to disclose). All respondents identified as either male or female. According to the summary statistics, approximately 52% of participants identified as male across both initial feed settings (50% for those initially on chronological feed and 53% for those on algorithmic feed). The study uses 'gender' rather than 'sex' as the data comes from self-reporting. No biological sex data was collected. Gender was included as a control variable in robustness checks but was not a primary focus of analysis. The study did not perform specific gender-based analyses of the algorithm's effects. We conducted our study independently of a social media platform, which means that our sample is smaller than what might otherwise have been possible, restricting our options to do heterogeneity analyses for reasons of statistical power (as described in the manuscript). The full dataset with individual-level gender information was collected with participant consent and is available in the repository, with appropriate privacy protections for identifiable information.

### Reporting on race, ethnicity, or other socially relevant groupings

The study used several socially relevant variables: gender, race, education, and political partisanship, all of them self-reporting through the YouGov survey platform. For details on the gender information, see the previous question. For linear regressions, some variables were dichotomized: Race was collected with eight categories in the YouGov survey, but due to small sample sizes in individual non-white categories, was dichotomized by the researchers into 'White' (78% of participants) and 'Other' categories for statistical analysis. Education was collected on a six-point scale (from 'No High School' to 'Post-graduate'), which researchers dichotomized into '4-year college degree or above' versus 'below 4-year college degree'. This corresponds to a dichotomization along our sample's median. Political partisanship was self-identified as Democrat, Republican, Independent, or Other party. These demographic variables were not used as proxies for other social constructs. They were included as controls in robustness checks, but the main results presented in the paper also show specifications without demographic controls, relying solely on the experimental randomization. When examining treatment effect heterogeneity using LASSO, these variables were included to predict initial feed choice.

### Population characteristics

The sample consisted of 4965 active, US-based, X users, of which approximately (up to rounding) 52% were males, 58% had a four-year college degree or more, 78% were white, 46% declared themselves as Democrats, 32% as Independents, and 21% as Republicans (percentages are rounded and may not sum to 100%). The average age in the sample was 50 years old.

### Recruitment

Participants were recruited through the survey company YouGov, drawing from U.S.-based registered members of the YouGov panel. Of 13,265 panel members contacted, 3,434 were screened out as they were not active X users (defined as using X less than 'several times a month'). The study design only admitted participants who self-reported being active on X at least 'several times a month' and provided informed consent. Of the 9,831 eligible participants, 8,363 provided consent, and 6,043 completed the pre-treatment survey. The final sample consisted of 4,965 participants who completed both pre- and post-treatment surveys. Several potential sources of bias should be noted. First, the sample is limited to YouGov panel members who are active X users, which may not be representative of the general U.S. population or all X users. Second, there was attrition between surveys -- approximately 21.8% of pre-treatment respondents did not complete the post-treatment survey. However, attrition rates were similar across treatment groups (21.88% for chronological and 21.72% for algorithmic,  $p = 0.87$ ), and Lee bounds analysis was conducted to assess the robustness of results to selective attrition.

### Ethics oversight

We obtained ethical approval from the Ethics Committee of the University of St. Gallen, Switzerland.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Study description

This study is a quantitative field experiment that randomized X (formerly Twitter) users to different feed settings (algorithmic vs.

Study description	chronological) for seven weeks in 2023. The study collected quantitative data through pre- and post-treatment surveys, computational analysis of feed content, and analysis of accounts followed by participants.
Research sample	The research sample consists of 4,965 active U.S.-based X users recruited through the YouGov survey panel. The sample demographics include: mean age of approximately 49 years (with standard deviation of about 15 years), 52% male, 78% white, and 58% with a college degree or higher. Prior to the experiment, 76% of participants used the algorithmic feed (X's default setting) and 24% used the chronological feed. The sample is not representative of the general U.S. population as it is limited to active X users (defined as using X at least 'several times a month'). However, this selection was necessary for the objective of understanding the effects of X's recommender algorithm on users' political attitudes.
Sampling strategy	The sample was recruited from YouGov's U.S. panel of respondents through rolling admission in July 2023. All panel members were contacted and screened for eligibility (being active on X at least 'several times a month'), resulting in 6,043 pre-treatment survey completions and 4,965 post-treatment survey completions. No formal sample size calculations were performed prior to data collection. The approximate sample size was given by our budget (we remunerated participants). The achieved sample size of nearly 5,000 participants provided sufficient statistical power to detect significant treatment effects, as demonstrated by the precise estimates reported in the study.
Data collection	Data were collected through these methods: online surveys administered through YouGov's platform at the beginning and end of the seven-week experiment period, including the option to run a custom Google Chrome extension that captured feed content from participants who opted to install it, and public X account data for participants who shared their handles. The study was conducted entirely online without direct interaction between researchers and participants. Treatment assignment (algorithmic vs. chronological feed) was randomized through the YouGov platform. The researchers were blind to individual participants' treatment assignments during data collection. Participants were informed about the general purpose of studying X's newsfeed but were not aware of specific hypotheses.
Timing	Pre-treatment survey data were collected on a rolling basis in July 2023. Post-treatment survey data were collected from the end of August through September 2023. The average duration between entering and leaving the study was seven weeks.
Data exclusions	Participants were recruited through the survey company YouGov, drawing from U.S.-based registered members of the YouGov panel. Of 13,265 panel members contacted, 3,434 were screened out as they were not active X users (defined as using X less than 'several times a month'). The study design only admitted participants who self-reported being active on X at least 'several times a month' and provided informed consent. No participants were excluded after.
Non-participation	Of the 9,831 eligible participants (see above), 8,363 provided consent, and 6,043 completed the pre-treatment survey. The final sample consisted of 4,965 participants who returned to also fill out the post-treatment survey. This corresponds to a 50.5% response rate.
Randomization	Participants were randomized into the algorithmic or chronological feed setting through a random generator on the YouGov survey platform.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Plants

---

Seed stocks

Not used.

Novel plant genotypes

Not used.

Authentication

Not used.