Social Data Collaboratory at NORC

Online Sports Betting Hits Social Feeds—But Not Evenly, New NORC Study Finds

These surveys were conducted by NORC at the University of Chicago's Social Data Collaboratory in collaboration with the AmeriSpeak® Omnibus team.

Responses were collected via three waves of cross-sectional online surveys using the AmeriSpeak® probability-based panel, designed to be representative of the U.S. household population.

- Wave 1 (August 8-12, 2024): 1,137 completed interviews
- Wave 2 (December 18-21, 2025): 1,042 completed interviews
- Wave 3 (January 22-26, 2026): 1,029 completed interviews

The margin of sampling error for the August 2024 wave is +/-3.8 percentage points, and the margin of sampling error across the combined December 2025 and January 2026 waves is +/-2.9 percentage points, all at the 95 percent confidence level, including the design effect. The margin of error may be higher for subgroups.

NOTE: All results show percentages among all respondents, unless otherwise labeled. For measures of gambling exposure, we include all content exposures up to a week prior to each respondent's date of donation.

Study Methodology

This study was conducted by NORC's Social Data Collaboratory in collaboration with the AmeriSpeak® Omnibus team and was supported by a NORC internal investment and an external grant.

Recruitment

Data were collected using the AmeriSpeak® Omnibus, a bi-monthly multi-client survey using NORC's probability-based panel designed to be representative of the U.S. household population. The survey was part of a larger study that included questions about other topics not included in this report. During the initial recruitment phase of the panel, randomly selected U.S. households were sampled with a known, non-zero probability from the NORC National Sample Frame and then contacted by mail, email, telephone, and field interviewers (face-to-face). The panel provides approximately 97 percent coverage of the U.S. household population. Those excluded from the sample include people with P.O. Box-only addresses, some addresses not listed in the USPS Delivery Sequence File, and some newly constructed dwellings.

Three waves of cross-sectional online surveys were conducted between August 8-12, 2024, December 18-21, 2025, and January 22-26, 2026. In each wave, respondents were randomly drawn from the AmeriSpeak panel of adults ages 18 and over, representing the 50 states and the District of Columbia. Respondents were offered a small monetary incentive for completing the survey.

In August 2024, 1,137 respondents completed the survey; in December 2025, 1,042; and in January 2026, 1,029. The final stage completion rate is 23.0 percent for the August 2024 wave and 29.8 percent across the combined December 2025 and January 2026 waves. The margin of sampling error for the August 2024 wave is +/-3.8 percentage points and the margin of sampling error across the combined December 2025 and January 2026 waves is +/-2.9 percentage points, all at the 95 percent confidence level, including the design effect. The sampling error margins may be higher for subgroups. Sampling error is only one of many potential sources of error, and there may be other unmeasured errors in this or any other survey.

Quality assurance checks were conducted to ensure data quality. In total, 141 interviews were removed for nonresponse to at least 50 percent of the questions, for completing the survey in less than one-third the median interview time for the full sample, or for straight-lining all grid questions. These interviews were excluded from the data file prior to weighting.

Once the sample was selected and fielded, and all study data were collected and finalized, a poststratification process was used to adjust for survey nonresponse and for noncoverage or under- and oversampling resulting from the study-specific sample design.

Poststratification variables included age, gender, census division, race/ethnicity, and education. Weighting variables for all three waves were obtained from the 2024 Current Population Survey. The weighted data reflect the U.S. adult population ages 18 and over.

Additional information on the AmeriSpeak Panel methodology is available at: https://amerispeak.norc.org/about-amerispeak/Pages/Panel-Design.aspx.

Eligibility for Data Donation

In the August 2024, wave, respondents who reported they had Facebook, Instagram, or both (n=869/76.4%) were considered eligible to donate data. In the December 2025 and January 2026 waves, respondents who reported they had Facebook, Instagram, and/or TikTok (n=1,688/81.5%) were considered eligible to donate data.

Eligible respondents were shown a screen that explained that NORC is interested in learning more about the effects of social media on society and asking respondents to share their own social media data with NORC for research purposes. A detailed description of the steps involved in data donation was presented. In addition, respondents were assured that "All social media data shared with NORC will be stored securely, and personally identifiable information (PII) will be removed. Any results from analysis will be reported in aggregate, with no PII disclosed."

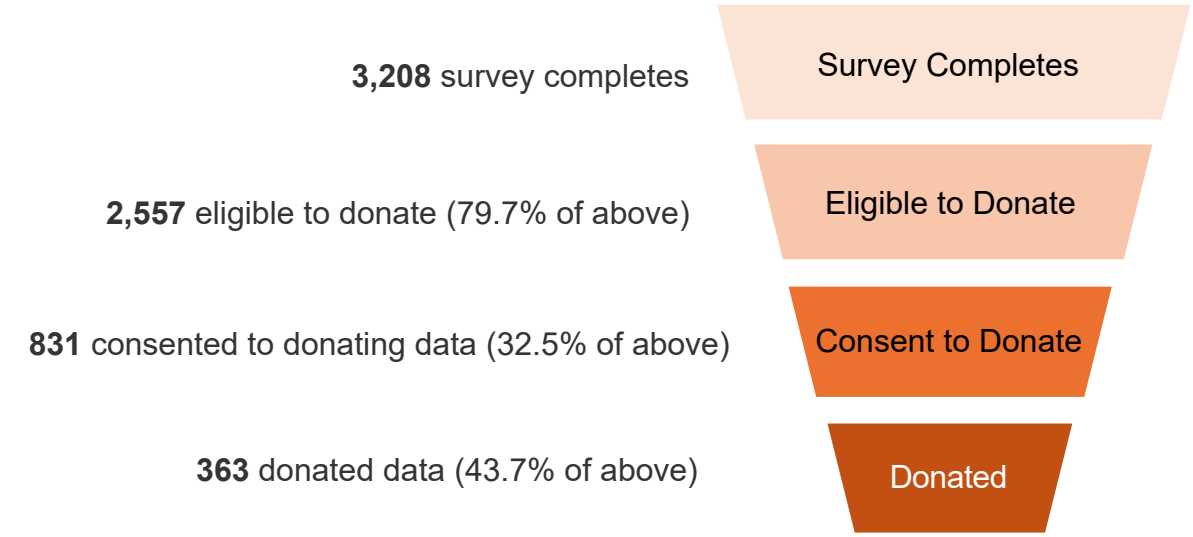
Opt-in and Consent to Data Donation

After seeing the screen with the description of data donation, all eligible respondents were asked:

"Would you consider sharing or 'donating' your social data from any platform to NORC to use for research, if provided adequate compensation?" Response categories included Yes, Maybe, and No.

Immediately following this question, all eligible respondents were informed that NORC would provide an additional incentive per platform as a thank-you for their time and for contributing their data. Figure 1 represents the flowchart showing eligibility, consent, and data donation.

Figure 1. Data Donation Process Across Three Waves



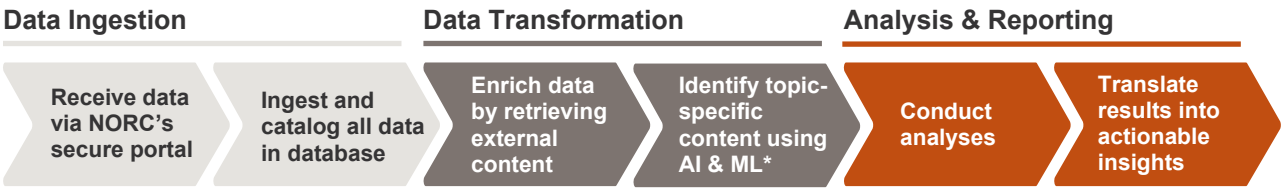
Source: NORC at the University of Chicago

Social Media Data Analysis

The social media data provided by the platforms to respondents arrived in raw JSON or HTML format in a zip folder consisting of several files. The process of creating measures from this raw data involved several steps. Figure 2 summarizes our data ingestion, transformation, and analysis pipeline.

Figure 2. Data Process Pipeline: Ingestion, Transformation, Analysis & Reporting

Our resource-intensive workflow first catalogs interactions from complex data archives, enabling complete download and artificial intelligence (AI) analysis of every associated URL.



*AI = artificial intelligence; ML = machine learning

Source: NORC at the University of Chicago

First, the data were ingested into a database, in which every interaction, search, and app they used was sorted and catalogued. Files ingested into the database included data on time-stamped views of paid advertisements, along with videos and posts viewed, respondent-specific ad targeting, searches conducted, purchases, and other consumer product/media account logins using Meta account credentials. For the current project, we analyzed the Facebook ads,

videos, and posts, the TikTok videos, and the Instagram accounts viewed by each respondent a week prior to their donation.

The files we received typically contained a mix of text data, including web addresses pointing to specific content viewed. Next, we enriched the data by retrieving the content of linked web addresses for TikTok and Facebook exposures. To transform the raw data into measures of exposure, we developed—with assistance from large language models (LLMs)—a set of nearly 200 keyword criteria to identify online gambling, casino, and prediction market exposure from the text of individual TikTok and Facebook content and from the usernames of each Instagram account.

To improve the precision of our matches, we included a set of approximately 20 exclusion criteria. We validated our approach using a sample of 100 random exposures flagged for online gambling and 100 random non-flagged exposures, verifying that our search criteria accurately distinguished between the two. We repeated this process for benchmark content categories such as e-cigarettes/vaping, beauty, and fitness.

About the Social Data Collaboratory

NORC's Social Data Collaboratory (SDC) brings together a diverse team of social science, data science, and communication experts. With decades of experience, cutting-edge technology, and a commitment to being responsible data custodians, the SDC offers unparalleled expertise in collecting, processing, and interpreting social media data. The SDC can access and analyze data from a wide range of platforms, including Facebook, Instagram, YouTube, TikTok, Reddit, and X (formerly Twitter), and we continuously add new platforms to our portfolio. We employ advanced computational methods, including natural language processing, machine learning, and AI capabilities, to provide unbiased research and actionable insights.

Our portfolio of work is agnostic to subject matter, and we are agile in our capability to collect and analyze social media data from varying platforms. We have conducted comprehensive analyses on social, health, and political topics. Our partners trust us to harness the potential of social media data for transformative research, informed decision-making, and insights into societal trends.