User Guide

CDEI's Public Attitudes to Data and AI Tracker Survey, December 2021

Published by the Centre for Data Ethics and Innovation (CDEI), the Public Attitudes to Data and AI (PADAI) Tracker Survey offers data on public awareness, opinions and perceptions of the use of data and AI in addition to the factors shaping these attitudes. Key topics include the use of digital services, opportunities and fears around data use, levels of comfort with specific data and AI use-cases, perceptions of data practices and media representations of data use. It aims to build understanding in this space so that we can develop trustworthy approaches to data and AI governance that address the public's concerns and reflect their values.

The survey was undertaken with a large nationally representative sample of 4,250 individuals, alongside a booster sample of 200 digitally disengaged individuals who were contacted via telephone. This allows a wide range of views to be captured and to explore how attitudes differ across the population. Fieldwork for this study was conducted between November 2021 and January 2022.

This is the first wave of this survey which will be repeated in order to track how public attitudes in this space change over time. For each wave we will publish the data on the UK Data Service, making it freely available online alongside insights in line with our commitment to transparency. The accompanying report and summary tables can be found on the CDEI website[1]. For any further information please contact public-attitudes@cdei.gov.uk.

Documents

- Data for the online nationally representative sample in csv format including weighting variable
- Codebook for the online nationally representative sample explaining variables and answer coding in excel format.
- Data for the sample of digitally excluded individuals and codebook of variables included as a excel spreadsheet on separate tabs
- Questionnaire in pdf format

---

[1] https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey

Sample

The Public Attitudes to Data and AI Tracker Survey is designed to be a nationally representative sample on the UK adult population, aged 18 and over. However, we acknowledge the limitations of online panels which exclude those not using the internet, especially for researching topics such as attitudes towards data and AI. Therefore, we have complemented this online sample with telephone interviews specifically targeted at digitally excluded individuals.

The overall sample is split between an online nationally representative sample of 4,257 individuals and a booster sample of 200 digitally excluded individuals interviewed through Computer Assisted Telephone Interviewing (CATI). The CATI interviews mainly followed the same questionnaire as the online sample, however, the questions have been reduced slightly due to time constraints and adapted to fit the format of the telephone interview where identical questions are not possible.

The data for these two samples are included separately as the digitally excluded CATI sample is not nationally representative. It should also be noted that the different format of interviewing may have influenced responses and is therefore not always directly comparable.

Online Sample:
CDEI worked with Savanta to conduct the online survey, which participants received via an email invitation which directs them to the questionnaire. Quota sampling was used to ensure that the sample is nationally representative for UK adults. Interlocking quotas for age and gender, as well as quotas for NUTS1 UK region, ethnicity and social grade were set, using the latest ONS mid-year population estimates. Panellists are recruited from a diverse set of recruitment sources and through a variety of recruitment methods in order to avoid bias, including opt-in email, co-registration and e-newsletter campaigns as well as social media. Panellists have made a conscious decision to participate in online surveys through a double opt-in registration process, and the participants in the panel are closely monitored to track activity and engagement, and to ensure quality. Panellists receive a small monetary incentive for taking part in surveys.

There are seven individuals in the original sample who have been removed due to their age being recorded as 100. The low number of centenarians in the population means we believe this is due to input errors at the data collection stage.

Telephone Sample:
200 respondents with very low digital familiarity were contacted and interviewed via the telephone. This telephone booster sample captures the views of those who are excluded

from online surveys, and are likely to be adversely affected by digital issues in different ways to other groups. Respondents were recruited using screening questions based on the definition of very low digital skills outlined in [Lloyds UK Consumer Digital Index][2], which was adapted for recruitment for this survey. Specifically this included not using, or having low confidence in, at least 3 of the following 5 areas: using email, using online banking, preferring face to face banking, entering personal details online and getting help using the internet.

Quotas were not applied to this sample to best reflect the demographic makeup of respondents that lacked these digital skills. Due to differences in digital skills and the small sample size of this group, respondents were on average much older than the online survey sample, more often female, and were exclusively white ethnicity. There are other intersecting reasons that individuals may be digitally excluded that are potentially not captured by this sample due to difficulties contacting these groups. Further research that specifically captures the views of these groups would be of great benefit but was beyond the scope of this survey.

Weighting

The online sample was weighted to be nationally representative, based on official statistics concerning age, gender, ethnicity, region and socio-economic group in the UK. To ensure a final sample representative of the UK adult population, the data is weighted to correct discrepancies from quotas set for these demographics. Random Iterative Method (RIM) weighting was used for this study, such that the final weighted sample matches the actual population profile. The weighting variable is included in the dataset as **'w8'**.

The 'digitally excluded' respondents were not weighted, nor were any demographic quotas applied when interviewing.

Data

Data should be analysed in conjunction with the data codebook to align numeric responses with substantive answers to questions. Free text questions are removed from the dataset as they are more identifiable than other data but can be shared on request. The conjoint experiment data is also excluded and can be shared on request.

The free text questions include

---

[2]
https://www.lloydsbank.com/assets/media/pdfs/banking_with_us/whats-happening/210513-lloyds-consumer-digital-index-2021-report.pdf

- Q10. In a couple of sentences, please could you briefly tell us what the story you saw about data was about?
- Q22. Please type in one word that best represents how you feel about 'Artificial Intelligence'.
- 'Other' responses to questions throughout the survey e.g. ethnicity.