

# Centre for Data Ethics and Innovation

## Public Attitudes to Data and AI

### Tracker Survey

Wave 1 (December 2021)  
Final report

# Contents

3	About us
6	Executive summary
10	Methodology
13	Findings
14	Section 1. Data for public good
23	Section 2. Uncertainty about data use
28	Section 3. Data sharing preferences
39	Section 4. Media's influence on attitudes towards data use
47	Section 5. Attitudes to AI
55	Section 6. Impacts of Covid 19
59	Contact details
60	Acknowledgements
61	Annex A: Deep dive sections (Trust in actors, Technology user segments, Digital behaviours)
93	Annex B: Details on methodology

# About Us



# About Us

The Centre for Data Ethics and Innovation (CDEI) leads the UK Government's work to enable trustworthy innovation using data and AI. Our mission is simple: to facilitate responsible and trusted innovation to improve the lives of citizens and support growth.

Engaging with the public is critical to realising this mission. The insights we generate through our research inform our work with partners, in the public sector and beyond, to develop trustworthy approaches to data and AI governance that address the public's concerns and reflect their values.

Our [Public Attitudes Insights Team](#) makes use of a variety of qualitative and quantitative methodologies when engaging with the public, to develop a deep understanding of public attitudes towards data and AI.

In line with our team's commitment to transparency and collaboration, we publish the findings and data from our public engagement research on [GOV.UK](#). We also lead the UK's first 'Public Attitudes to Data and AI' (PADAI) network for cross-Whitehall organisations involved in data policy and social research. This network allows insight and best practice to be shared across Government, maximising the value and impact of public engagement work.

# About the PADAI Tracker Survey

While many organisations are stepping up efforts to engage with the public on the use of technology, this survey is the first in this space that monitors how public attitudes towards data and AI change over time. It uses advanced analytics, including conjoint analysis, to assess and quantify the factors driving public attitudes towards data sharing and the use of data-driven technologies.

A large nationally representative sample of 4,000 individuals, alongside a booster sample of 200 digitally disengaged individuals who were contacted via telephone, has allowed us to capture a wide range of views, and explore how attitudes differ across the population.

We hope this survey will be a vital resource for those in Government, the wider public sector, civil society, industry and academia, who are seeking to understand public attitudes towards data and AI, and the conditions needed to create a trustworthy environment for data use.

There are, however, limitations to this research. While the survey has and will continue to provide important and novel insights about attitudes towards data and AI use, the findings should be contextualised and considered alongside other research. People do not have one singular attitude towards 'data': it is understood and perceived differently by different groups and is dependent on many contextual factors. It is important that we conduct in-depth public engagement exercises focused on specific use-cases, and utilise alternative methodologies to complement survey research - for example, conducting deliberative focus groups to hear from the public in their own words, or behavioural experiments to see how the public behave in practice. Data in this report is presented factually and neutrally. We do not seek to make recommendations or offer explanations beyond our research, and should be interpreted alongside other research in this space.

To stay up to date with future versions of the survey, please register your interest [here](#). We would welcome any further discussion on this survey and report. Please get in touch at [public-attitudes@cdei.gov.uk](mailto:public-attitudes@cdei.gov.uk).

# Executive summary



# Executive summary

2021 was a year of disruption and development for data use; the COVID-19 pandemic rapidly necessitated new ways of interacting with the digital world. This highlights the importance of understanding public opinion on data use, both in context and as something that can shift. Understanding public opinion enables data to be harnessed in a way that addresses public concerns and reflects public values.

In December 2021, we asked the public about their understanding, awareness and attitudes to data use. Below is a summary of the findings from the first wave of our PADAI Tracker Survey, which acts as a baseline for future waves to identify how attitudes change over time:

**Section 1: Data for the public good:** As digital activities have been widely adopted into everyday lives, both people's behaviours and reported attitudes indicate that they are broadly comfortable with their data being used in a variety of contexts. This includes data being used for personal benefit, and also the opportunity to use data to improve lives across society. COVID-19 and health are specific areas where people feel the use of data presents the greatest opportunities. However, it is important to recognise differences across the population, with some groups reporting low confidence using digital services and low knowledge of how data is used.

**Section 2: Uncertainty about data use:** For the full potential of data to be realised, public concerns around data use must be addressed. People report feelings of uncertainty about current data practices and fairly limited knowledge regarding how data about them is used and collected in their day-to-day lives, demonstrating the opportunity and importance of meaningful transparency about data use by organisations. This uncertainty, alongside perceived risks around data security, data control and data accountability are barriers that must be overcome to build confidence in data use.

# Executive summary

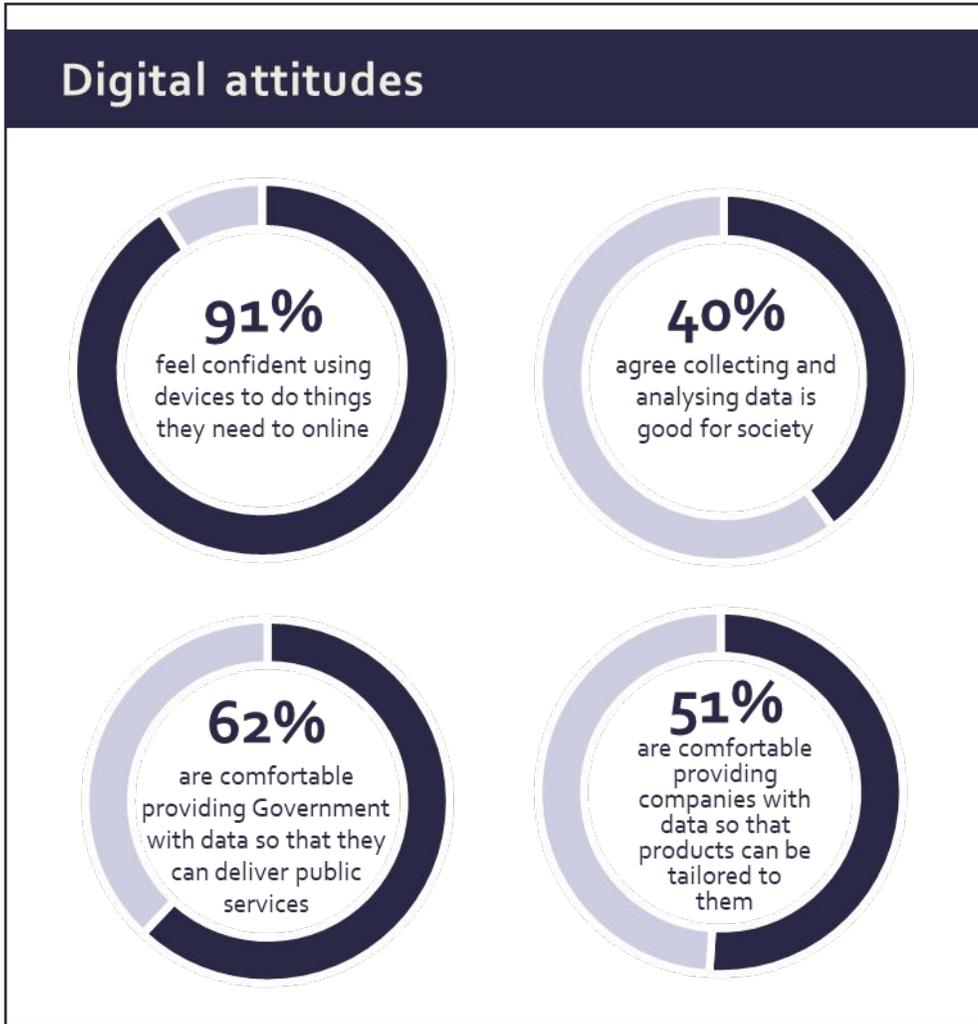
**Section 3: Data sharing preferences:** In practice, concerns about data use are strongly impacted by the extent to which individuals trust the organisations holding and using data about them, and the degree of data governance in place. Results from the conjoint experiment indicate that these were strong influencing factors on people's willingness to share data.

**Section 4: Media:** Our media analysis suggests that perceived opportunities and fears about data use are reflected in, and influenced by, media stories. Front-of-mind negative media stories about data use focused on data breaches and the misuse of data, mirroring fears about data security. Conversely, people reported seeing positive presentations of data use in the media about the reporting of, and response to, the COVID-19 pandemic, mirroring the view that this is an area where data use could bring benefits.

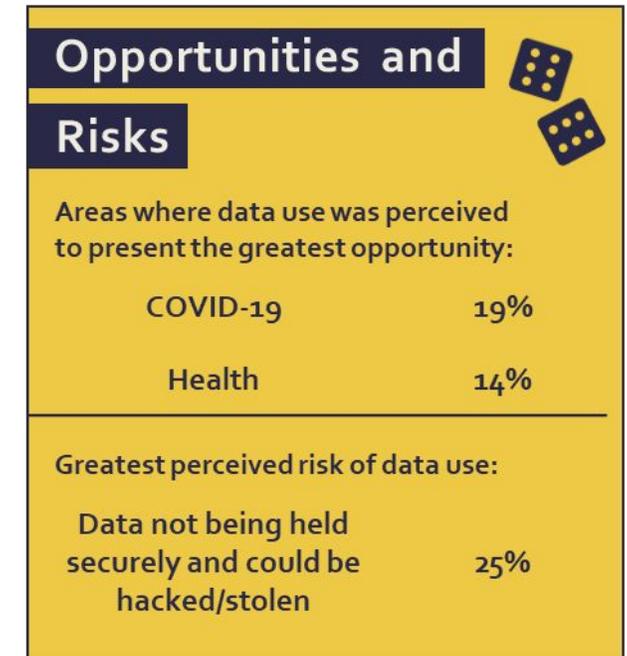
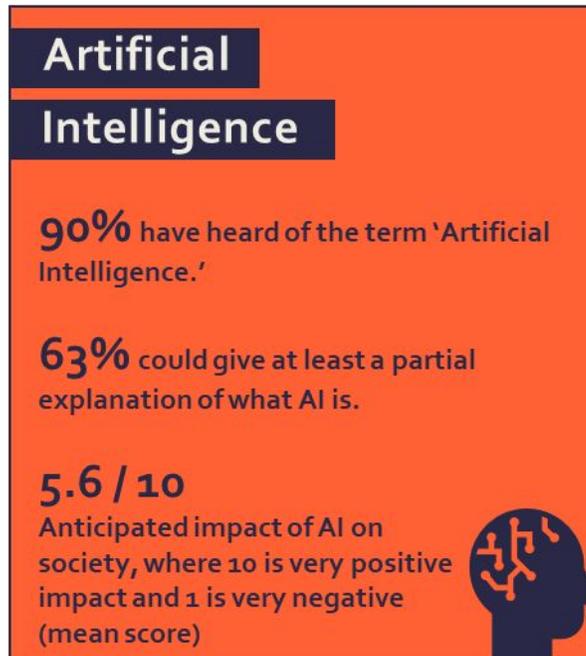
**Section 5: Attitudes to AI:** When it comes to Artificial Intelligence, the public highlighted the potential concern that the benefits of AI would not be felt equally across society. Feelings of worry and fear about the impacts of AI are also more pronounced amongst those who have lower digital familiarity.

**Section 6: Impacts of COVID-19:** With the COVID-19 pandemic changing so many aspects of our day-to-day lives, there were split views on how effectively data has been used in the response to COVID-19. With rapid and continuous developments in the ways in which data is being used and perceived by the public, it will be important to continue monitoring its impact on public attitudes.

# Dashboard of summary statistics



Trust in organisations	(Central) Government	NHS	Social media companies	Big technology companies
To act in your best interests	39%	89%	36%	60%
Keep data about you safe	47%	75%	33%	49%



# Methodology



# Methodology

## Sample

A total of 4,257 online surveys were completed by a demographically representative sample of UK adults (18+).\*

A further 200 'digitally excluded' UK adults were interviewed via Computer Assisted Telephone Interviewing (CATI) to ensure the views of those with low digital familiarity were included.



## Fieldwork dates

The online survey ran from 29<sup>th</sup> November 2021 to 20<sup>th</sup> December 2021.

The telephone interviews with the 'digitally excluded' sample were conducted between 15<sup>th</sup> December 2021 and 14<sup>th</sup> January 2022.



## Additional analysis

CDEI conducted a segmentation on the online sample according to attitudes to data to help tailor insights and policy.

A conjoint analysis was also conducted to identify data sharing preferences.



## 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.

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

### Digitally excluded telephone booster 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 based on 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.

### Statistical significance and interpretation

Findings marked with  arrows are statistically significant, with a 95% confidence level. That is, arrows indicate a statistically significant difference from all other subgroups within the same category. For instance, if there is an  arrow attached to the results for a particular question for those identifying as male, this means that this figure is higher than the one for females *and* that there is a 95% chance this difference would persist if we repeated the fieldwork under the same conditions. Where arrows are not shown, statistical significance may exist with other members of that subgroup but not all. For example, the youngest age category may be statistically significant from older age categories but not from the second youngest age band. Where 'significant' is used throughout this report, statistical significance is meant.\*\*

\*The online sample contains 7 interviews which recorded ages of 100. These have been left in the total sample, but do not have an impact on the overall findings.

11 \*\*Please note that Slides 99-100 includes a different set of arrows which are to be interpreted differently to the above description. This is explained fully on those slides. All other instances of statistical significance arrows in this report are to be interpreted as detailed above.

# Digital familiarity sample profiles

A proxy measure for **digital familiarity** is used in this report to compare responses to other questions broken down by the digital familiarity of respondents. The very low digital familiarity band is comprised of the 'digitally excluded' telephone sample. The low-high familiarity bands are based on responses to questions about respondents' frequency of online activities and confidence using devices online\*. Respondents were given a composite score out of 18 and categorised into bands based on their score. The demographic differences of these bands reflect demographic differences in digital familiarity, for example the older age profile of groups with lower digital familiarity.

	Number of respondents	Average age	Gender Female / Male	Ethnicity (showing % > 1%)	Socio-economic group***
<b>Very low digital familiarity</b> the 'digitally excluded' telephone (CATI) booster sample	200	71	63% / 36%	100% white**	Higher (ABC1): 49% Lower (C2DE): 52%
<b>Low digital familiarity</b> (Score of 0-10/18)	1,476	61	43% / 56%	93% White, 4% Asian	Higher (ABC1): 46% Lower (C2DE): 54%
<b>Medium digital familiarity</b> (Score of 11-14/18)	1,697	47	51% / 48%	89% White, 6% Asian, 3% Black	Higher (ABC1): 46% Lower (C2DE): 54%
<b>High digital familiarity</b> (Score of 15-18/18)	1,084	36	61% / 39%	82% White, 10% Asian, 4% Black, 2% Mixed	Higher (ABC1): 59% Lower (C2DE): 41%

# Findings



# 1. Data for public good



# 1. Data for public good: key insights

There are clear opportunities for data to be used for the public good:

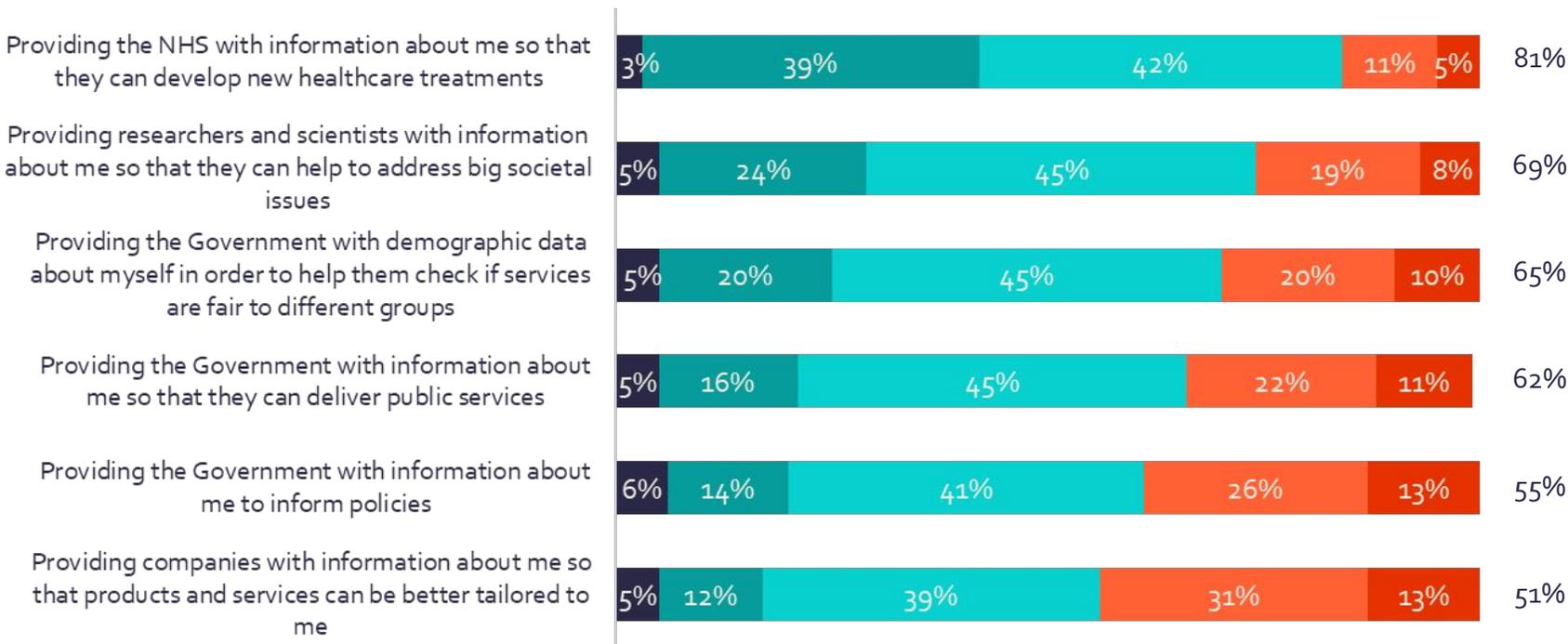
- **The vast majority of survey respondents reported that they frequently create and use data and data-driven technology;** 93% reported using the internet most days or every day.
- **Although, there is a risk that a group of people will be excluded from using digital services due to low knowledge, skill or access;** only 45% of the digitally excluded telephone sample reported that they were confident using devices online.
- **Most respondents reported being comfortable with data being used in a variety of different contexts;** 81% of respondents were comfortable providing personal data about themselves to the NHS to develop new healthcare treatments, 62% were comfortable providing personal data for the Government to deliver public services, and 51% were comfortable providing companies with personal data to tailor products and services.
- **COVID-19, healthcare and the economy were identified by as the top priorities for where data use could bring benefits.**
- **Both the personal and societal benefits of data use were widely recognised by respondents,** however, they were less certain whether the benefits were felt equally across society.

# Respondents were generally comfortable with personal data being used by organisations for a variety of purposes. They were most comfortable with data being used to develop new healthcare treatments and address large societal issues.

**How comfortable or uncomfortable are you with providing personal information about yourself in the following instances? (overall, excluding CATI)**

■ Don't Know ■ Very comfortable ■ Somewhat comfortable ■ Somewhat uncomfortable ■ Very uncomfortable

Total:  
Comfortable

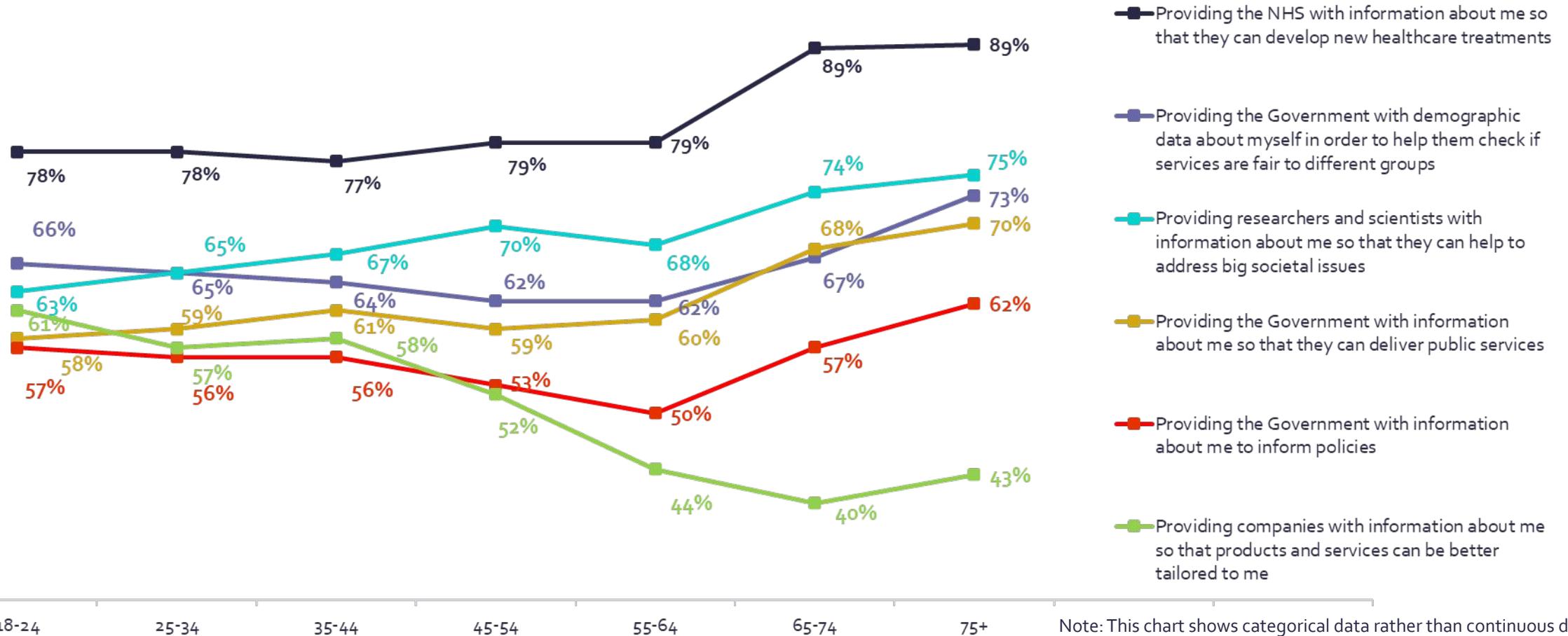


84% of those who trust the NHS to act in their best interests reported being comfortable with providing personal information to the NHS to develop new healthcare treatments. Comparatively, just 51% of those who do not trust the NHS reported feeling comfortable providing personal data in this scenario.

Black respondents were significantly more likely than any other ethnic group to report being comfortable with sharing personal data for the purposes of checking that services are fair to individual groups (72%).

Older respondents reported more polarised views on data being used for different purposes than younger respondents. They reported higher levels of comfort with data being used to develop healthcare treatments and lower for the personalisation of services.

Total: Comfortable with providing personal data in different instances – by age

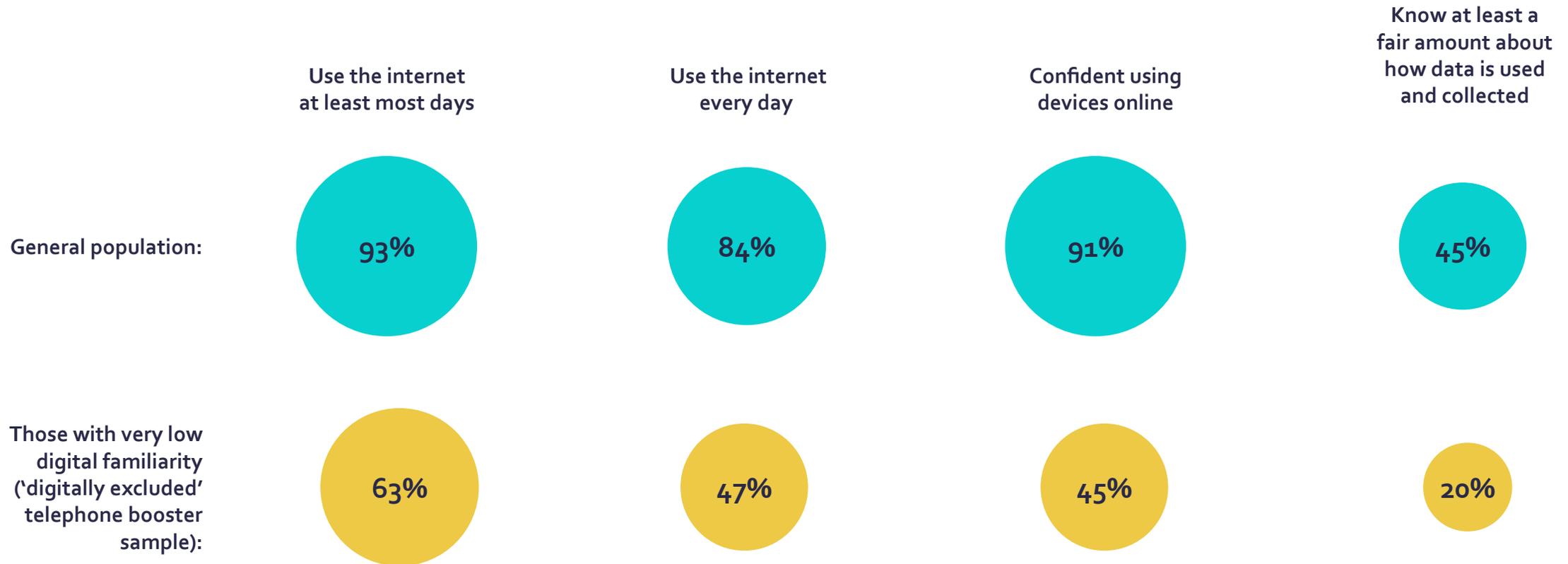


Note: This chart shows categorical data rather than continuous data.

Source: Q13. Please indicate how comfortable or uncomfortable you are with providing personal information about yourself in the following instances.

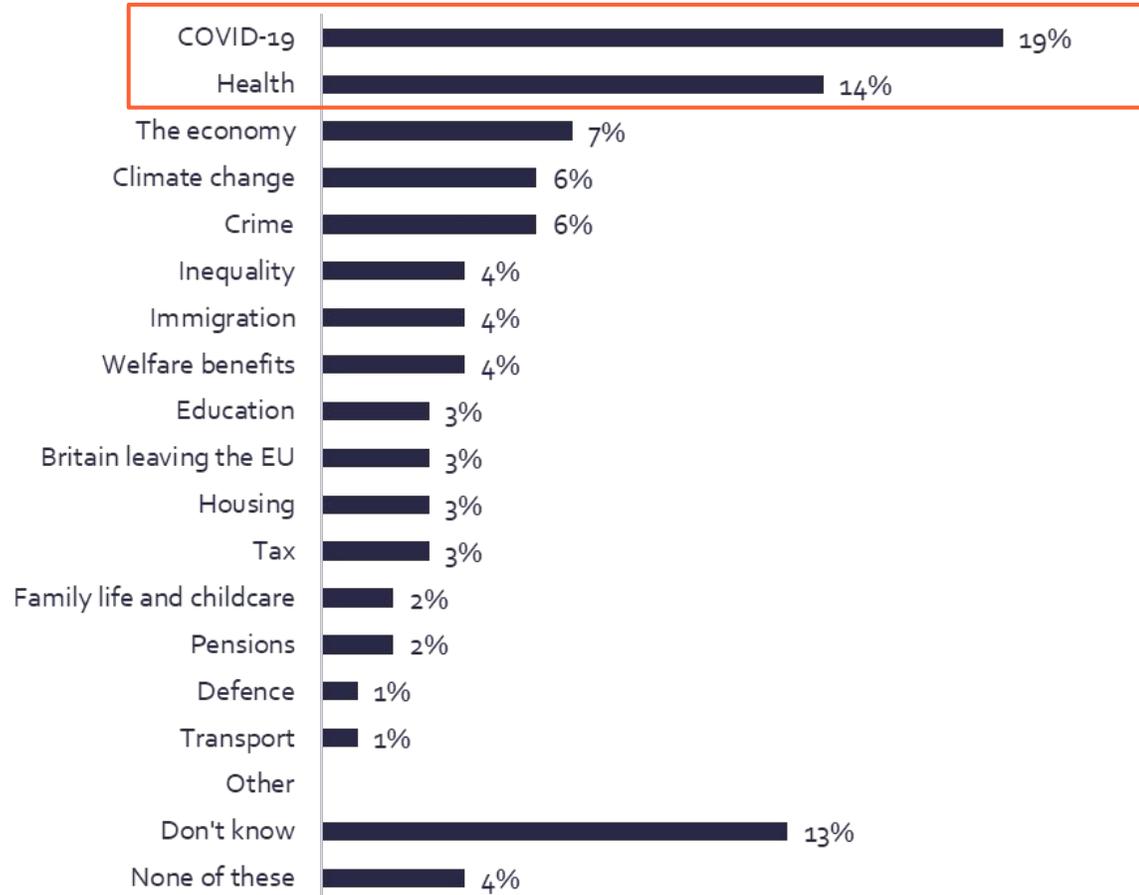
Base: all respondents, excl. CATI (n=4,257)

While the majority of respondents reported frequent internet use, there is a risk that some people will not be able to access digital services due to low skill or knowledge.

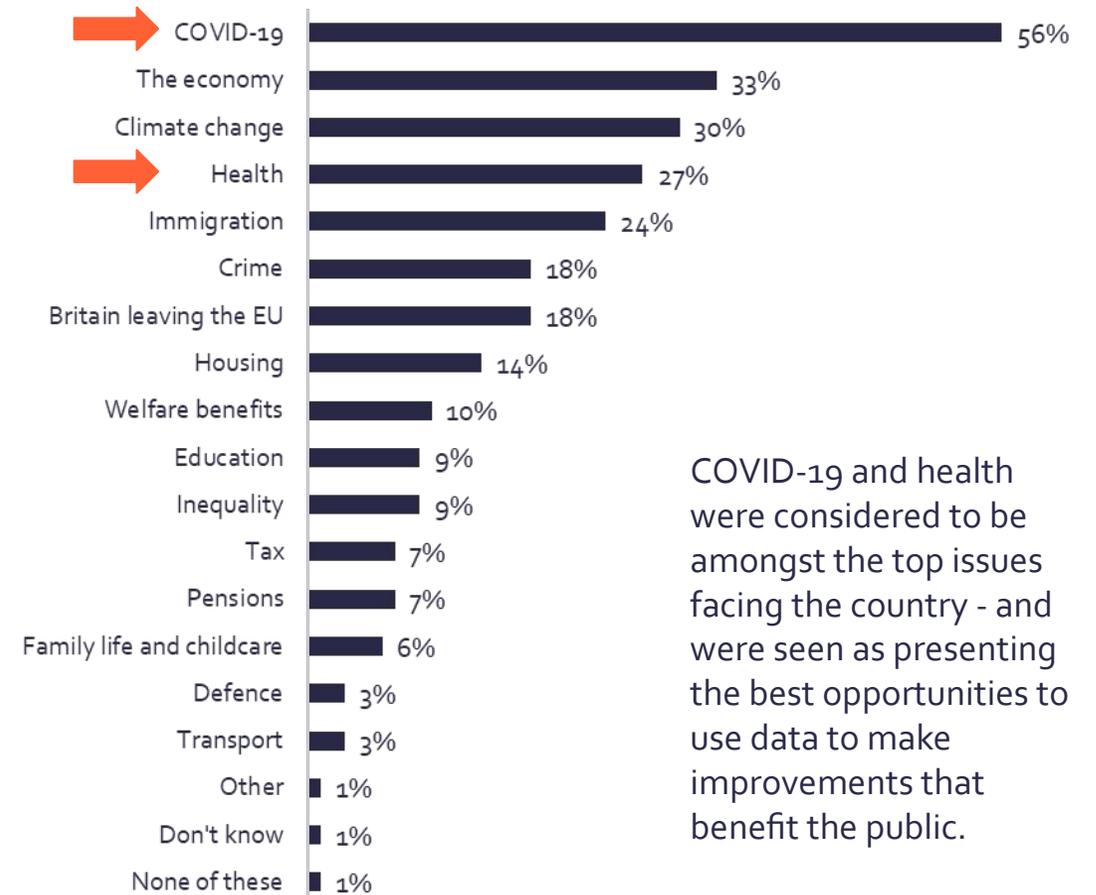


# Respondents reported that they believe data presents the greatest opportunities to make improvements to the COVID-19 response and health.

Issues where the use of data presents the greatest opportunity for making improvements that benefit the public



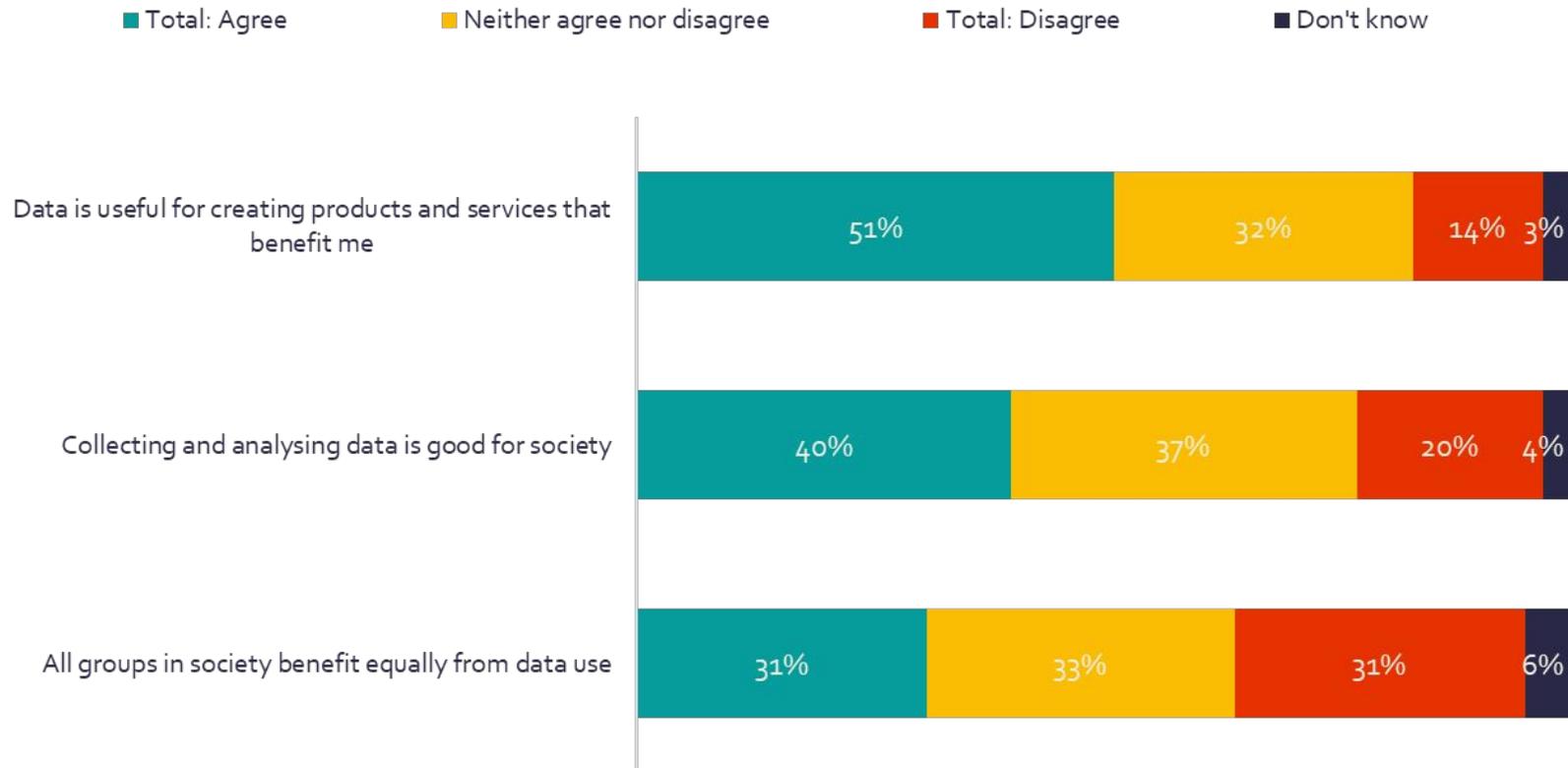
Most important issues facing the country (% selecting top 3)



COVID-19 and health were considered to be amongst the top issues facing the country - and were seen as presenting the best opportunities to use data to make improvements that benefit the public.

# Respondents recognised personal benefits in the use of data but were less certain about to what extent data benefits society and whether these benefits are equal.

Please indicate how much you agree or disagree with each of the following statements (overall, excluding CATI)



**51%** agreed that data is useful for creating products and services that have personal benefits

**Younger respondents** were more likely to agree that data is useful for creating products and services; 56% of those aged 18-24 agreed with this compared to 47% of those aged 75+ . Those from **higher socioeconomic grades** were also more likely to agree; 54% of those classified as the higher grades (ABC1) compared to 49% of those classified as lower grades (C2DE).

**40%** agreed that collecting and analysing data is good for society

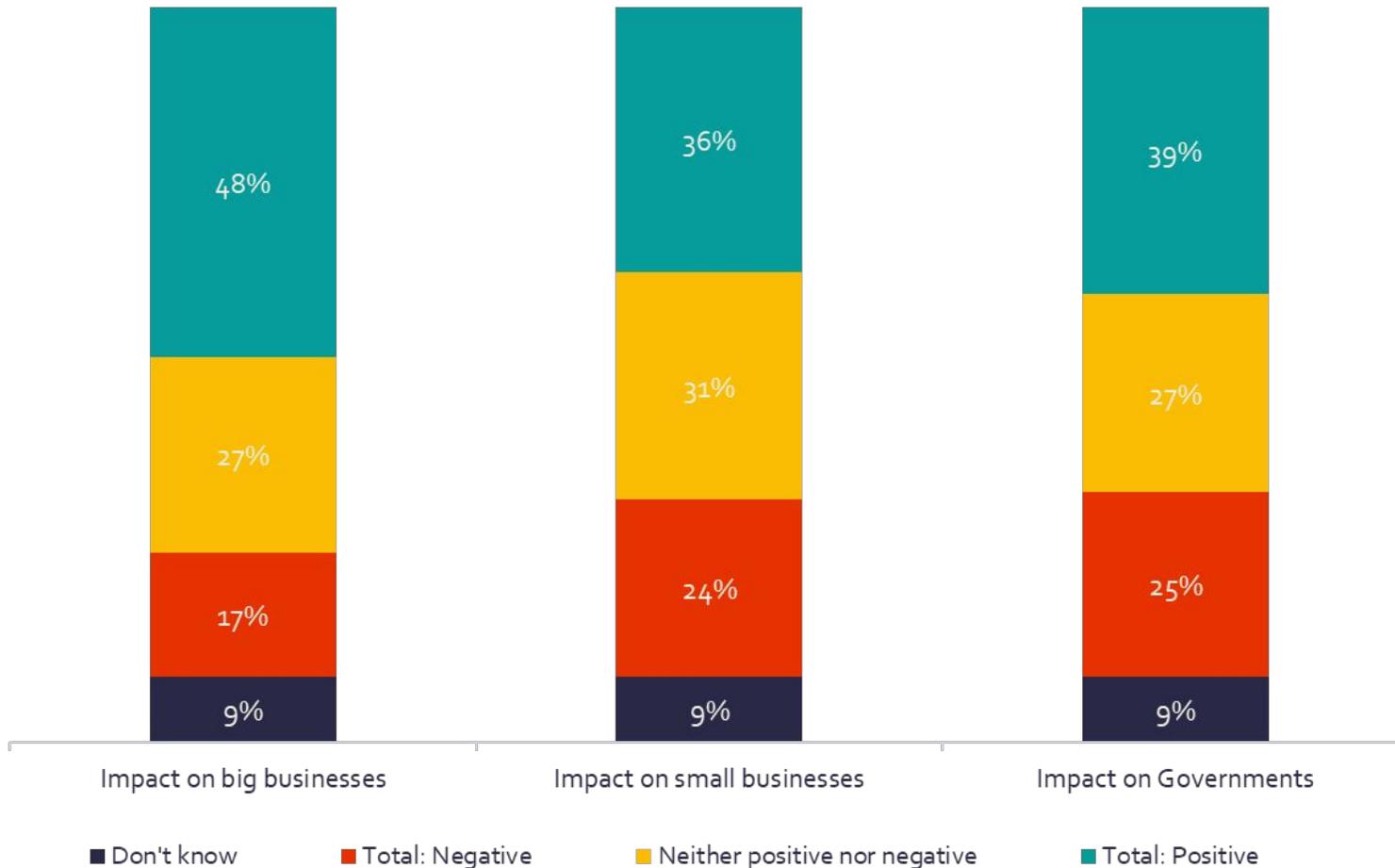
**Younger respondents** were more likely to agree that data is good for society; 49% of those aged 18-24 agreed compared to 34% of those aged 75+ . Those from **higher socioeconomic grades** were also more likely to agree that data is good for society; 42% of those from higher grades (ABC1) agreed compared to 37% for lower grades (C2DE).

**31%** agreed that all groups in society benefit equally from data use

**Younger respondents** were much more likely to agree that the benefits of data use are shared equally; 44% of those 18-24 agreed that this was true compared to 19% of those 75+ .

# Respondents had moderately positive expectations for the impact of Artificial Intelligence (AI) on large businesses.

To what extent do you think the use of Artificial Intelligence will have a positive or negative impact for the following types of people or organisations?



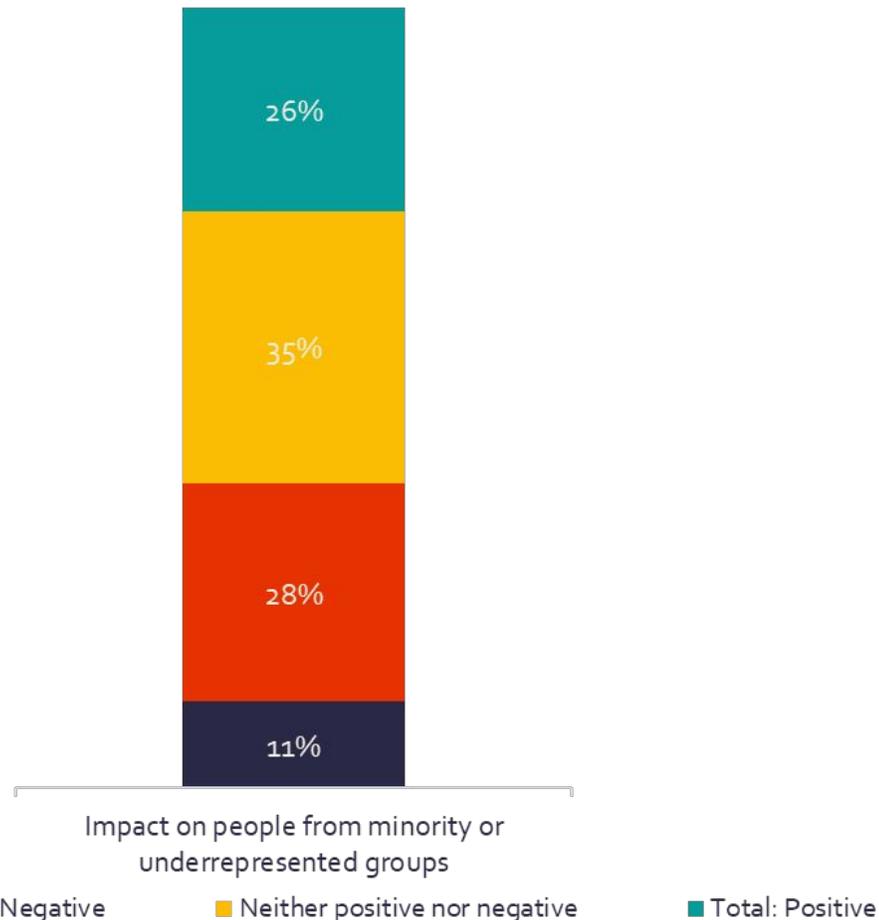
For AI, as well as data, benefits were recognised, with more respondents believing AI would have a positive impacts on big and small businesses as well as Government.

### Differing attitudes by job role

Interestingly, those whose **job involves data and technology** were significantly more likely to say they think AI will have a positive impact on big businesses (53%) and small businesses (40%) than those who don't use online tools in their roles (45% and 32% respectively).

# More respondents expected the impact of AI on minority groups to be negative than positive.

To what extent do you think the use of Artificial Intelligence will have a positive or negative impact for people from minority or underrepresented groups?



## Expectations of impact on minority groups by ethnicity of respondent

**White respondents** were more likely to expect the impact of AI on people from minority or underrepresented groups to be negative compared to respondents of other ethnicities, with 29% saying this, compared to 25% who said impacts would be positive. **Black respondents** were more likely to say the impacts would be positive, with 41% saying this compared to 18% saying negative. Similarly, 37% of **Asian respondents** said impacts would be positive compared to 22% who said negative.

This may be linked to the fact that Black and Asian respondents tended to be younger on average with **younger respondents** also more likely to expect the impact of AI on people from minority or underrepresented groups to be positive. For example, 36% of those aged 18-24 said the impacts will be positive, significantly higher than those in **older** age groups, with only 17% of those aged 75+ saying this.

## 2. Uncertainty about data use



## 2. Uncertainty about data use: key insights

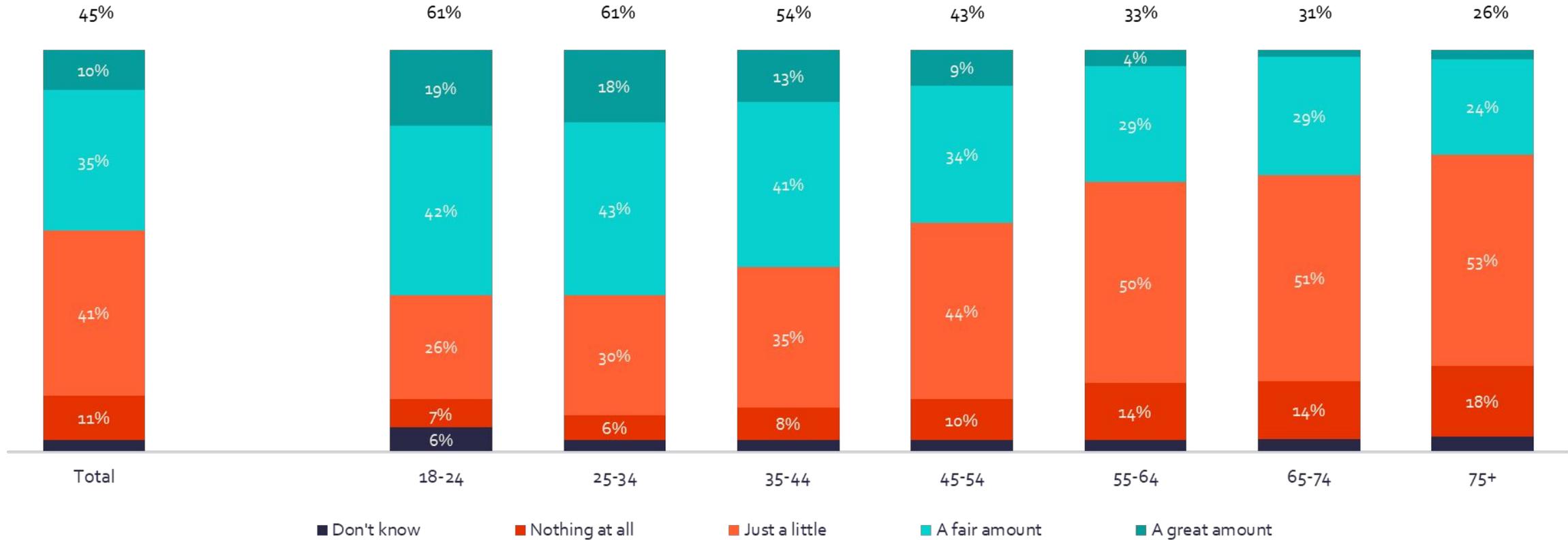
Public concerns around data use need to be addressed for the full potential of data to be realised:

- **A large proportion of the population do not feel that they know how data about them is being used;** 52% of respondents reported that they know only a little or nothing about how data about them is used and collected in their day-to-day lives.
- **There was also uncertainty amongst respondents about whether organisations are transparent about how they use data.**
- **Data security was seen as the greatest risk of data use.** Respondents expressed anxieties about the security of personal data and the trustworthiness or capabilities of organisations to protect it.
- **Amongst respondents there was a low expectation of accountability in cases of data misuse by either the Government or private sector.**

# Most respondents had limited knowledge about how data about them is used and collected in their day-to-day lives; knowledge of data use decreases with age.

How much, if anything, do you know about how data is used and collected about you in your day-to-day life?  
(by age, excluding CATI)

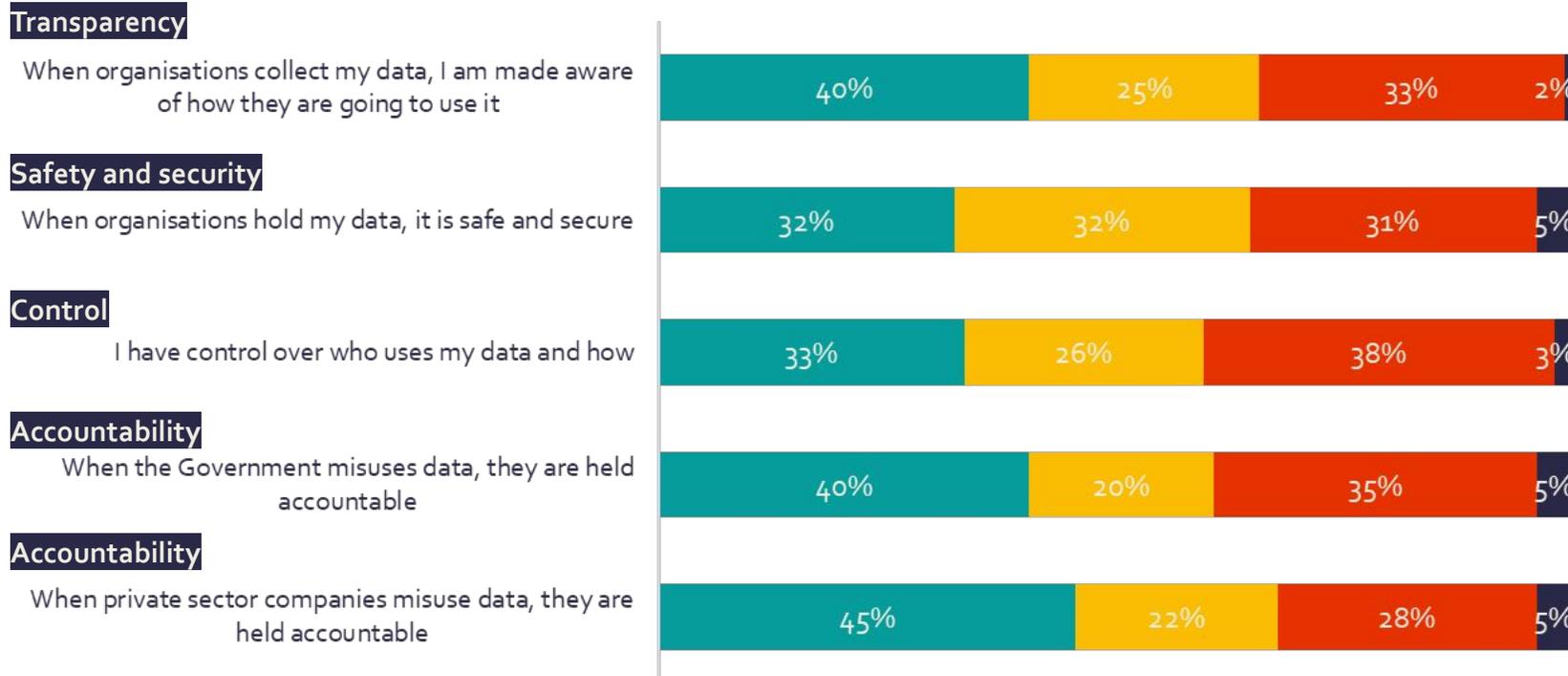
Total: At least a fair amount:



# Many respondents were unsure about how data about them is being used, whether it is being held safely and the extent to which they have control over it.

Please indicate how much you agree or disagree with each of the following statements

■ Total: Agree    ■ Neither agree nor disagree    ■ Total: Disagree    ■ Don't know



### Perceived control over data use

Younger respondents were more likely to agree that they have control over who uses their data and how, with this decreasing steadily as age increases. The difference was most stark between the **45-54 and 55-64 age bands** (35% and 24% agree respectively).

### Government accountability

Younger respondents were more likely to agree that the Government are held accountable when they misuse data. 48% of 18-24s agreed, while only 31% of those 75+ agreed. While agreement decreased with age, the biggest step change between age bands was between those aged 45-54 (42%) and aged 55-64 (33%).

Those who **trust the Government to act in their best interest** were also more likely to agree that the Government is held accountable (52%) compared to those who do not trust the Government (31%).

# Respondents were concerned about the security of personal data and trustworthiness of organisations holding data about them.

Which of the following do you think represents the greatest risk for data use in society?



Respondents reported being most concerned about **data security**, with data breaches also featuring heavily in the [negative news stories respondents reported seeing in the media](#). Data being sold without awareness or consent was also a key concern of respondents.

# 3. Data sharing preferences



## 3. Data sharing preferences: key insights

The conjoint survey experiment findings indicate strong preferences regarding when people are willing to share data:

- **Respondents were far more willing to share their data with the NHS than any other organisation.** Generally, public and third sector organisations are also preferred over private companies, particularly big technology companies and social media companies.
- **Respondents' willingness to share data with an organisation was strongly related to the levels of trust they reported in that organisation.**
- **The purpose of data use was found to influence decision-making less than the involvement of a highly trusted or untrusted actor across the range of use-cases tested.** Purposes that directly benefit society or individuals were better received, while only targeted advertising had a strong negative public reaction amongst the use-cases tested.
- **Respondents were more willing to share data when strong governance mechanisms were stated to be in place to assure them.**
- **Older respondents showed stronger preferences for sharing data with certain actors and were more concerned about data governance compared to their younger counterparts.**

# Survey design for conjoint study

A conjoint experiment was used to test the most important factors influencing respondents' willingness to share their data in different scenarios. We presented participants with a number of data sharing scenarios which differed by:

- the actor involved
- the type and identifiability of the data
- the purpose of data sharing
- the level of regulation applied

Respondents were asked to pick in which circumstance, out of a pair of randomly generated scenarios, they would be more willing to share their data. Six pairs of scenarios were shown to each participant. Analysing these results indicates which are the most important factors for making this decision.

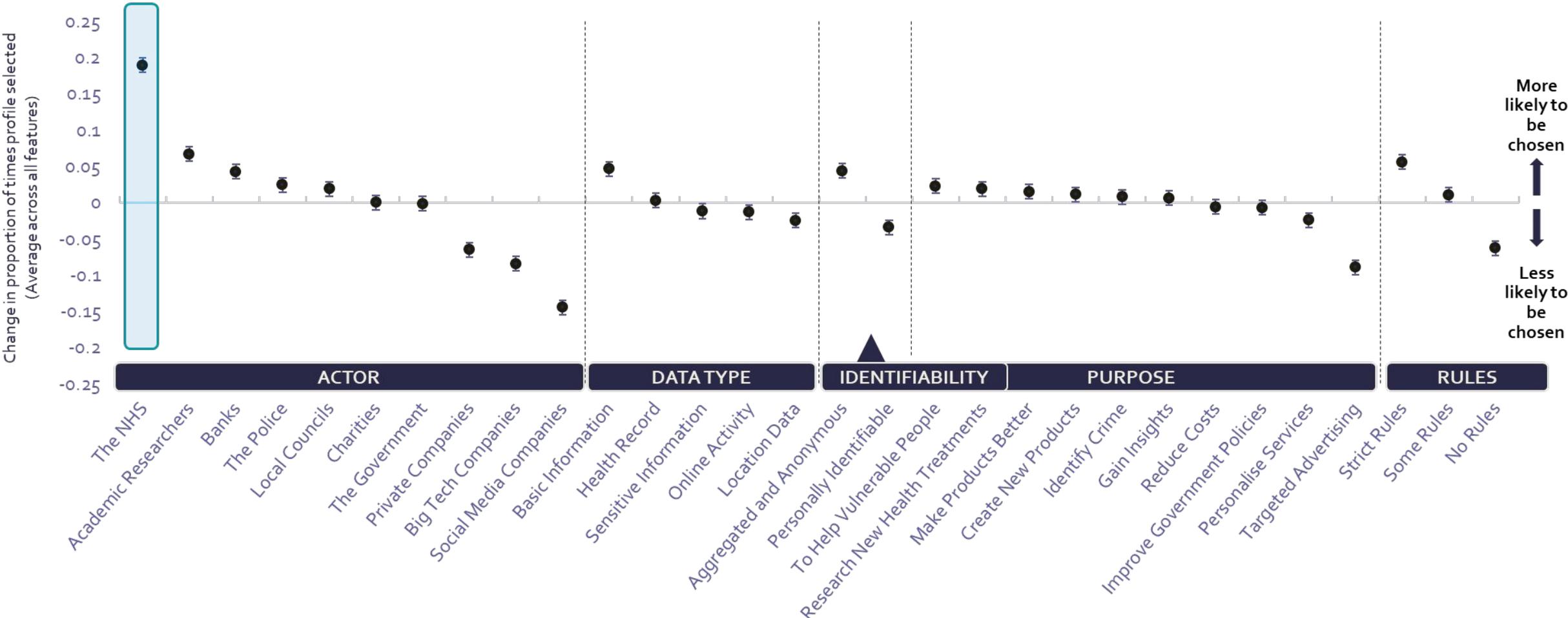
Conjoint experiments with forced choice decisions are an innovative way of researching causal relationships between different aspects of data sharing and public preferences. This is beneficial as self reported opinions might not reflect real life behaviours for complex issues like data sharing.

Example scenario shown to respondents  
[see [slide 106](#) for all hypothetical scenarios]

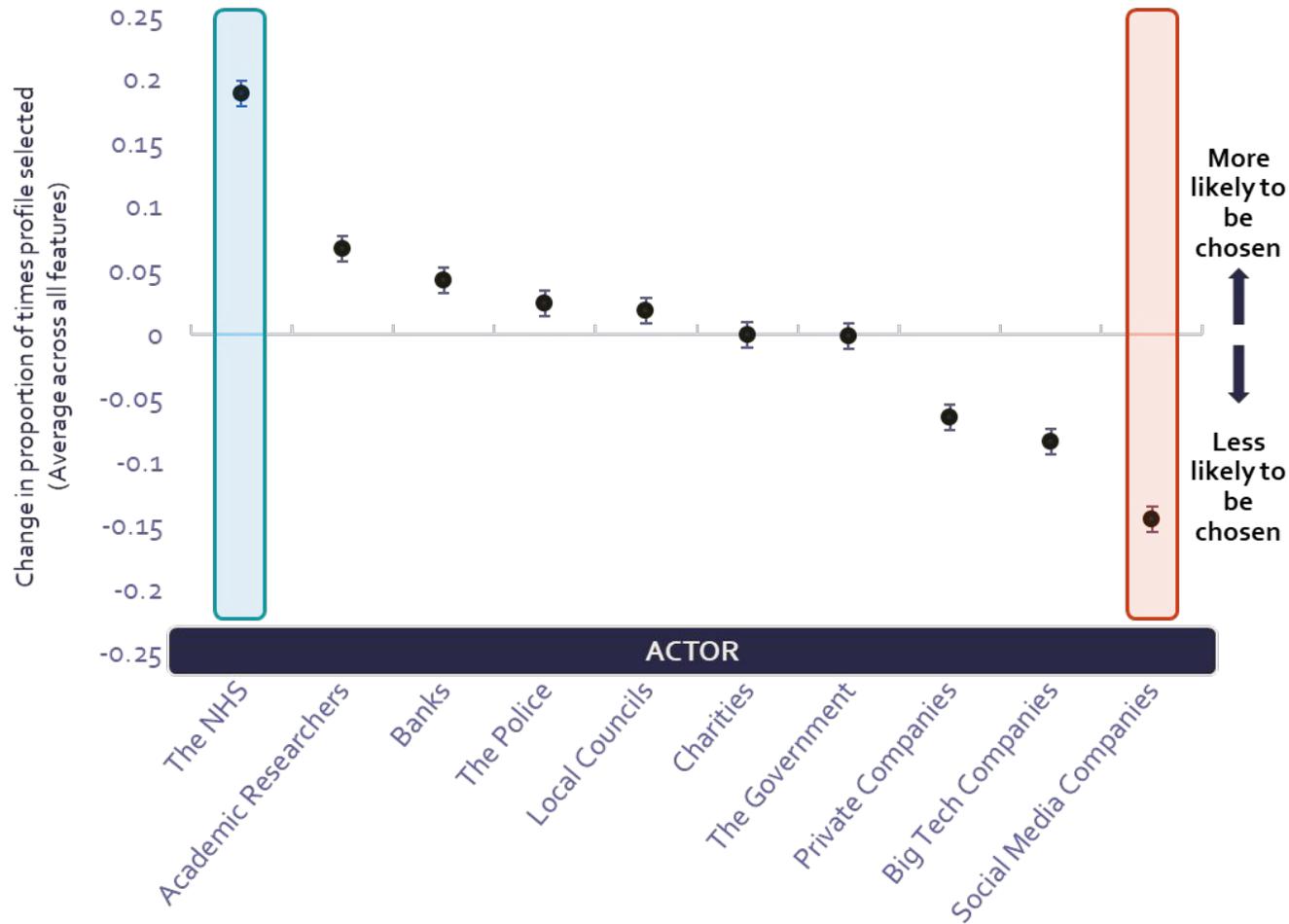
	● Option A	● Option B
	The NHS	Banks
To use...	Basic information about you e.g. gender, age	Location data from your phone about where you have been
That...	Has been averaged across the whole population so you can't be personally identified	Personally identifies you
In order to...	Improve Government policies	Create new products and services
If...	There are STRICT rules in place which require the organisation to limit and clearly explain how they will use people's data and keep this data safe and secure	There are NO rules in place which require the organisation to limit and clearly explain how they will use people's data and keep this data safe and secure

The conjoint experiment was designed by CDEI and analysed by Todd K Hartman, Professor of Quantitative Social Science at the University of Manchester. The experiment was run within the wider Savanta survey. See the Annex for further details.

# Respondents were more likely to be willing to share data when the NHS was the actor involved compared to any other feature in the conjoint experiment.



## However, a trust deficit is evident for private companies, big tech and social media companies.



Profiles containing the **NHS as the actor were chosen 69% of the time**, compared to **36% of the time** for those that contained **social media companies** as the actor. This difference of 33% points reveals the largest gap between any two features in the conjoint experiment.

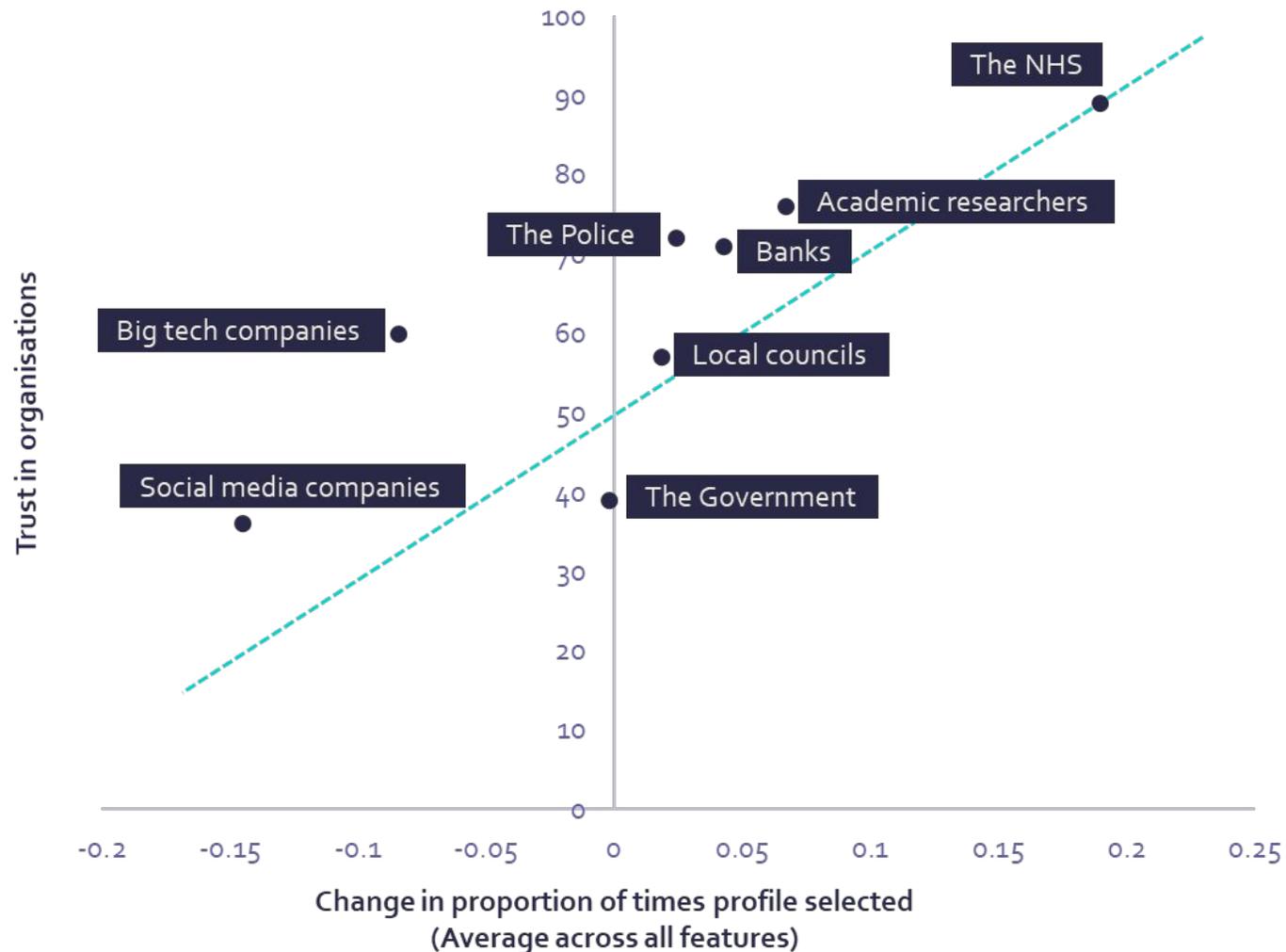
In reality, behaviours may not reflect this, as many individuals do share data with social media and big tech companies through their everyday use of technology. This highlights possible resignation by respondents to sharing data as part of using services.

Respondents were generally more likely to be willing to share their data with **public or third sector organisations** such as the NHS, academic researchers, police, and local councils, compared to **private sector organisations** including big tech and social media companies.

However, some private sector institutions such as banks are more trusted than newer sectors.

The involvements of charities or Government were not found to influence willingness to share data positively or negatively within the experiment.

# Willingness to share data with an organisation is closely related to trust in the organisation to act in the public's best interest.



Willingness to share personal data with different organisations was found to have a strong relationship with the extent to which those organisations are trusted to act in one's best interest.

The NHS and social media companies mark the high and low trust points, both in general and in trust with one's data.

Willingness to share personal data with the Government was relatively higher than other organisations people reported similar levels of trust in.

Big tech companies, on the other hand, have higher general trust but respondents report being less willing to share their data with these organisations. Within the conjoint experiment, profiles containing big tech companies were chosen in just 42% of cases, a reduction of 8% points from the average.

See the deep dive section on [Trusted Actors on Slide 62](#) for further detail.

## Respondents reported low trust in the data practices of some organisations across the public and private sectors.

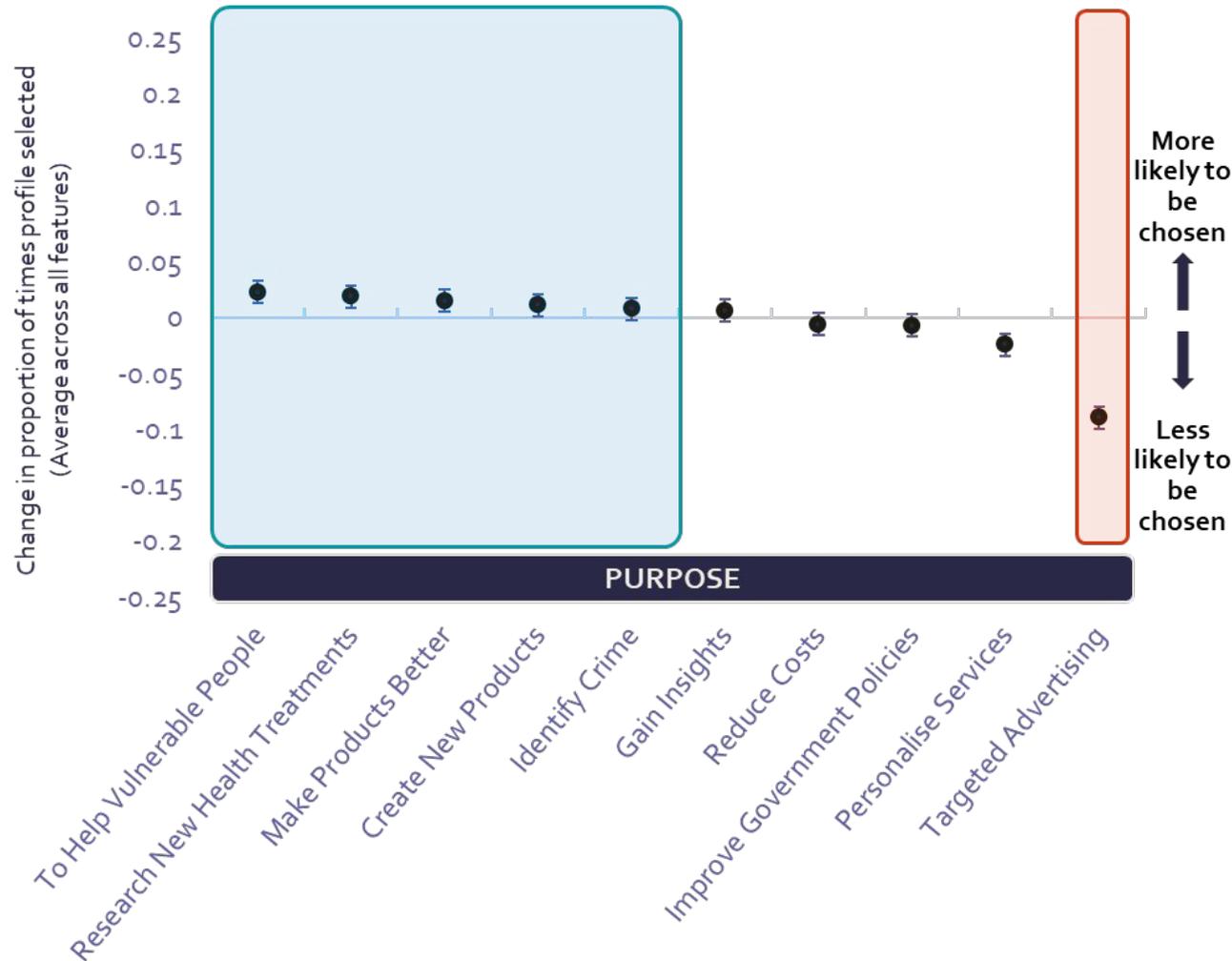
	To what extent, if at all, do you trust [organisation] to...? (% trust)						Generally trust to act in your best interest
	Effectively use the data to improve the products or services you receive	Keep data about you safe	Use your data to benefit society	Be open and transparent about what they do with data about you	Let you make decisions about how your data is used	Average trust in managing data	
The NHS	78%	75%	75%	73%	69%	74%	89%
Academic researchers at universities	65%	63%	67%	62%	60%	63%	76%
Banks	66%	69%	53%	60%	60%	62%	71%
The Police	58%	63%	57%	56%	52%	57%	72%
Utilities providers	55%	56%	48%	50%	48%	51%	61%
Local councils	54%	53%	52%	50%	46%	51%	57%
Local independent businesses	55%	50%	46%	49%	47%	49%	85%
Big technology companies	54%	49%	42%	43%	43%	46%	60%
The Government	48%	47%	46%	41%	39%	44%	39%
Social media companies	36%	33%	30%	32%	34%	33%	36%
Statement average	57%	56%	52%	52%	50%		

Trust in an organisation to manage data generally corresponded with trust in that organisation to act in individuals' best interest. One exception to this is local independent businesses which are highly trusted to act in individuals' best interest, second only to the NHS, but see much lower confidence in their data practices.

Across the statements tested, levels of trust do not differ significantly. However, on average, respondents were more likely to say they trust organisations to use data to improve products and services (57%) and less likely to say this about allowing you to make decisions about how your data is used (50%).

Social media companies received the lowest trust scores in terms of acting in individuals' best interest (36%) and in managing data (average of 33% across the metrics). However, behaviours have not necessarily been impacted by this as the majority of respondents reported that they use social media most days.

# The purpose of data use was less impactful on preferences compared to other influencing factors, with the exception of targeted advertising.



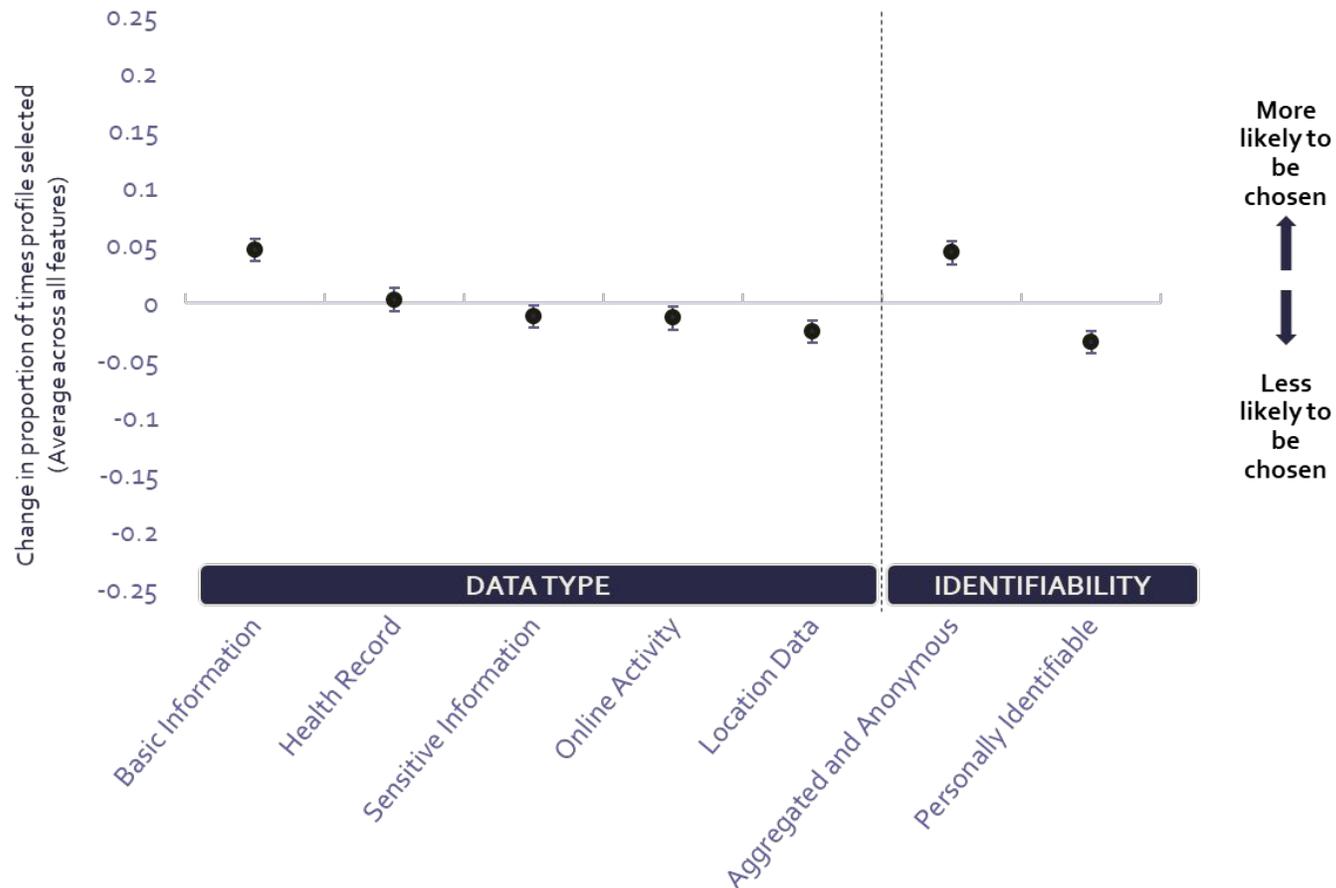
The purpose of the data use **did not especially motivate** respondents' data sharing preferences, compared to the importance of a highly trusted or untrusted actor being involved.

In reality, the purposes could be interpreted very differently depending on the who the data is shared with. For example, new products created by the NHS are much more likely to be thought of as leading to benefits to health compared to new products from social media companies which might be considered as something like an app or design feature. Therefore, the **impact of purpose in this experimental setup might have been superseded somewhat by actor.**

Respondents were more likely to share data if the purpose **directly helps other people** or has a **clear social benefit**: to identify crime, research new health treatments, help vulnerable people, or make products better.

Only **targeted advertising** had a clear negative response, of the options tested.

# Considerations around personal privacy influenced respondents' willingness to share data.

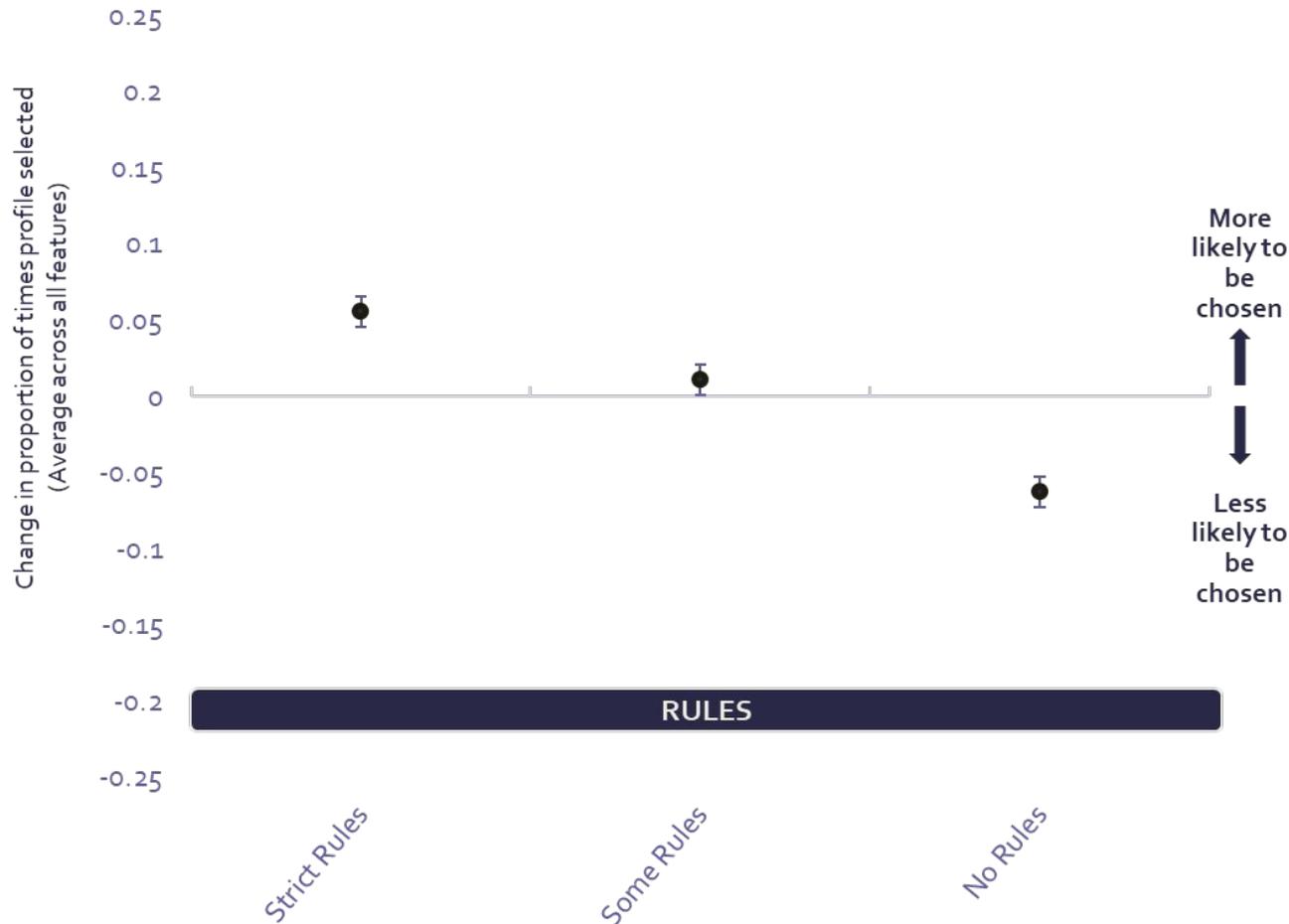


Respondents were more likely to be willing to share data if it is aggregated or anonymised, rather than personally identifiable.

Additionally, they were more likely to be **willing to share basic information**, such as gender and age, than other data types which are more sensitive and were not found to have a big impact on respondents' willingness to share data.

However, respondents did not greatly distinguish between their willingness to share other types of information including their health record, sensitive information such as income, sexual orientation, and ethnicity, online activity such as which websites they have visited or products they have bought, and location data about where they have been and when.

# Respondents were more willing to share their data if they are aware of strong governance in place to protect them.



Respondents were more likely to be willing to share data if **rules were in place** to protect them as users. This includes requiring organisations to limit data use, clearly explain how they will use people’s data, and keep data safe and secure. ‘Strict’ rules were preferred over ‘some’.

Respondents were similarly much less likely to be willing to share data if there are no rules in place.

Older respondents were less likely to want to share data with big tech and social media companies than younger respondents, and more willing to share data when it is non identifiable and strong governance is in place.



# 4. Media's influence on attitudes towards data use



## 4. Media's influence on attitudes towards data use: key insights

The perceived opportunities and fears about data use are reflected in, and influenced by, media stories:

- Front-of-mind negative media stories for respondents were largely related to misuse of personal data by social media companies, data breaches and other data risks.
- Media analysis across the last 6 months of 2021, finds that stories about data breaches in both public and private sector organisations were frequently seen in the media over this time period.
- The story with the largest online presence was about plans for new ways to share GP data.
- Positive media stories remembered by respondents about data use were dominated by COVID-19 and how data has been used to track and guide the response to the pandemic.

## Tracker Survey media monitoring methodology

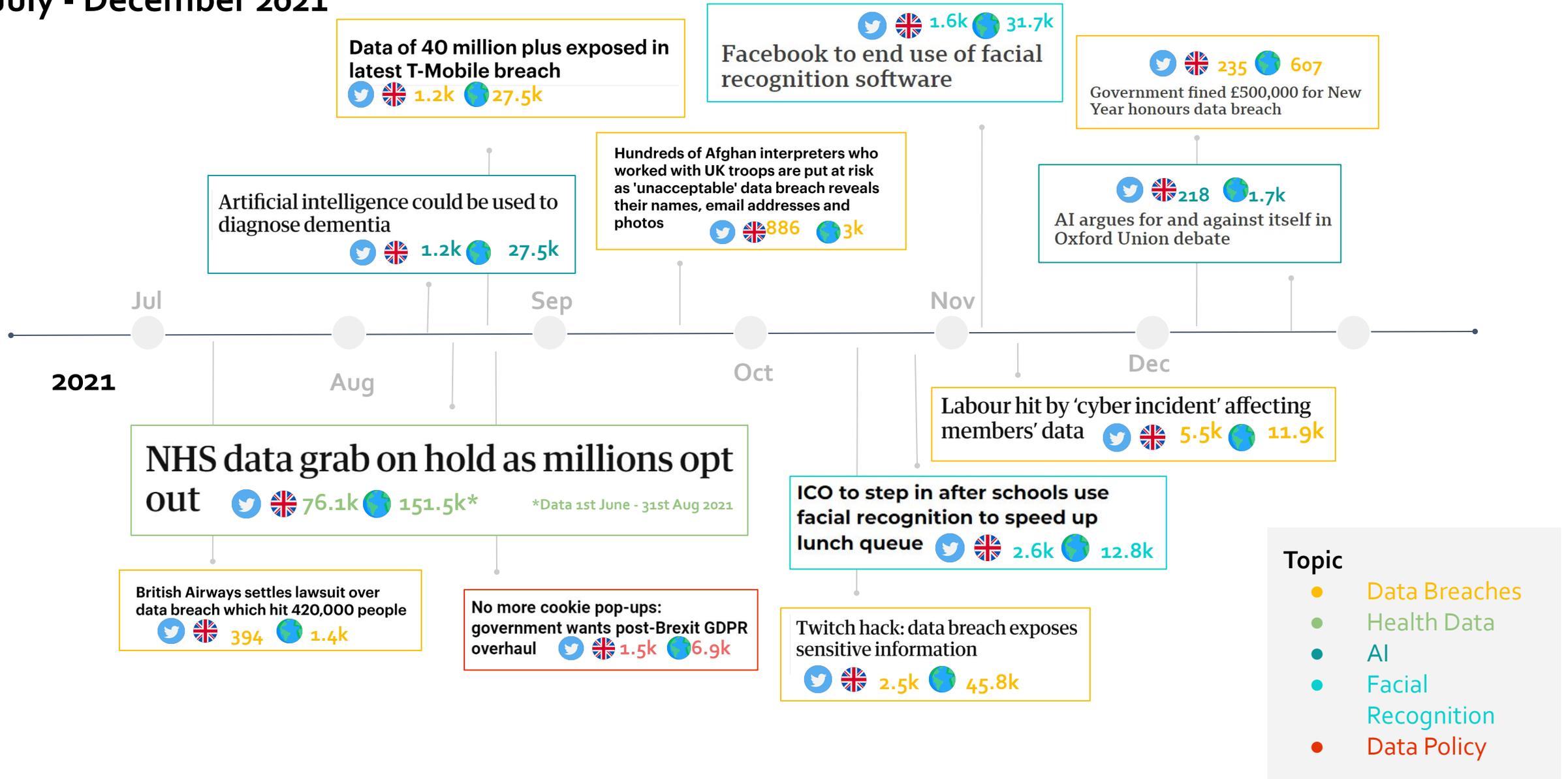
The way that data, AI and technology are presented and seen in the media is likely to both inform and reflect public attitudes about data use. As part of the Tracker Survey, CDEI undertook media monitoring to collect and analyse stories about data use for the 6 months preceding fieldwork (the latter half of 2021). This provides context of the way in which data is presented in the media and key events which may have influenced attitudes.

All media stories across the key themes of **data policy**, **data breaches**, **health data**, **AI**, and **facial recognition technology** were collected between 1st July 2021 - 31st December 2021 from the online sites of major British and international news providers, using the media monitoring tool Brandwatch. These topics were identified as key themes in the Tracker Survey free text media question, asking about recent 'data stories'. COVID-19 data stories were not collected as stories about cases, deaths, and vaccinations were daily occurrences during this time period. Top stories were selected as those with multiple articles about them or that appeared on the news sites with highest readership.

From this, Twitter search terms around key media stories were constructed to measure the reach of the individual stories. Within the following timeline these are marked with , showing numbers for UK locations only , and global Tweets . These metrics span the full timeframe unless annotated otherwise. All search terms can be found in the appendix.

# Key media stories about data use

## July - December 2021



# Tracker Survey media monitoring summary

News stories about data breaches and leaks were frequent and involved both public and private sector actors.

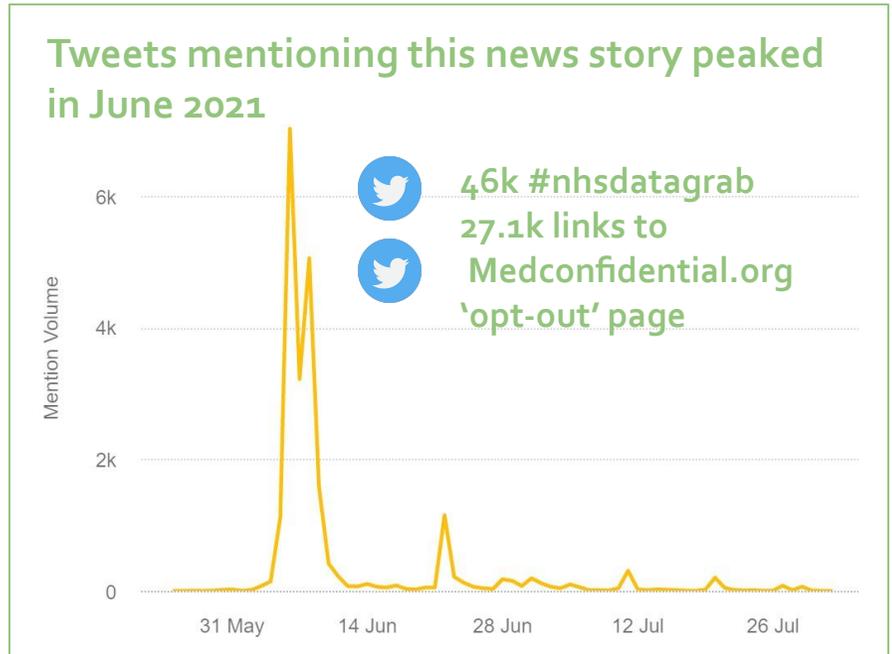
One of the most high profile news stories online was about plans for new ways to share GP data, with thousands of tweets including #nhsdatagrab and the opt-out link.

Most high profile news stories about data in the media were negative with the exceptions of data being used in healthcare, in relation to the COVID-19 response, and about the advancement of technology.

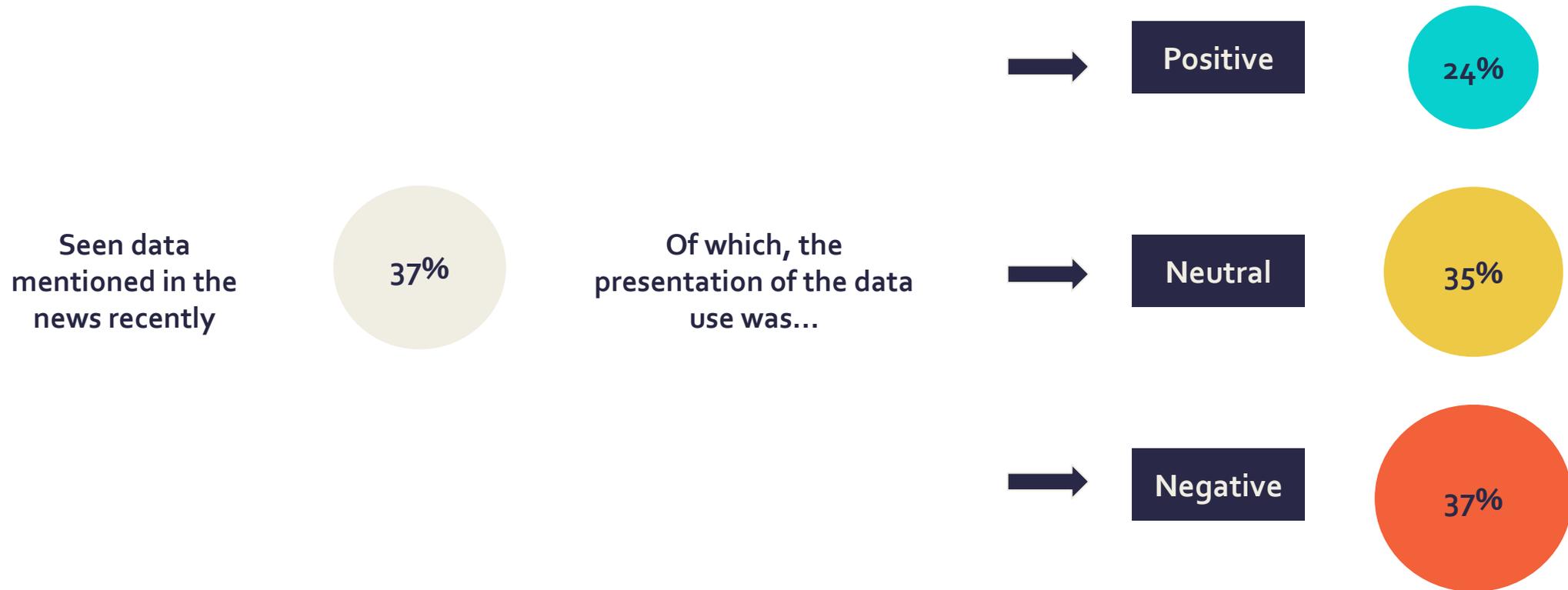
Concerns about the application of new technologies like facial recognition appeared frequently in the news.

## NHS data grab on hold as millions opt out

A plan to share GP data was set to launch in September, but an online summer campaign has prompted widespread dissent



When respondents recalled news stories about data, the stories were more frequently negative presentations of data use than positive.







# 5. Attitudes to AI

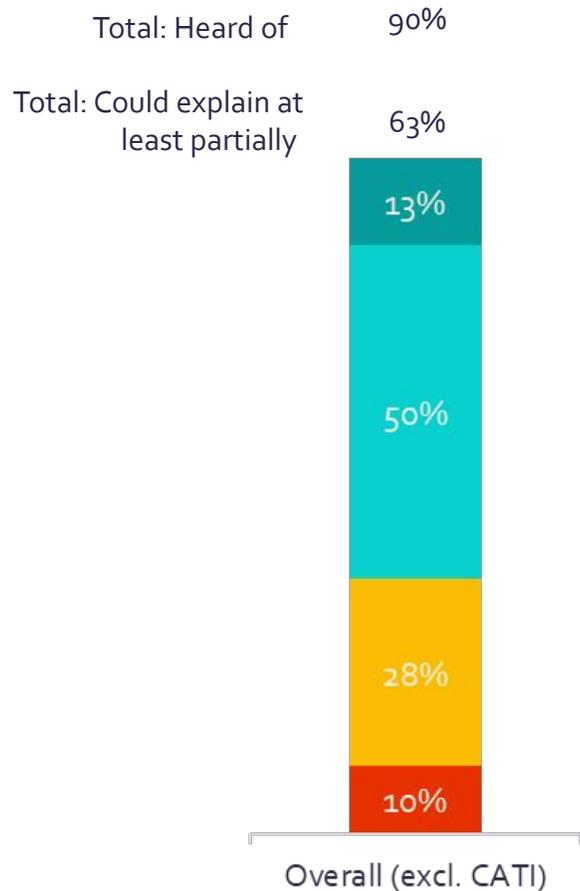


## 5. Attitudes to AI: key insights

The public does not expect the benefits of AI to be felt equally across society;

- **Respondents reported limited knowledge of Artificial Intelligence (AI);** only 13% of respondents felt they could offer a full explanation of this term. This is perhaps expected as AI covers many different technologies and applications, making it a challenging term to define and explain.
- **AI was predominant associated with being scary and futuristic amongst respondents.** Those with the lowest digital familiarity reported the lowest knowledge of AI and more frequently associated AI with feelings of worry and fear.
- **Many respondents expressed discomfort with some applications of AI,** including 32% of respondents that said they were uncomfortable with AI being used to power internet search engines which is a common practice.
- **Those with higher digital familiarity were more optimistic about applications of AI bringing benefits.**

# Knowledge of Artificial Intelligence (AI) is limited – even with an expected degree of overclaim, most respondents could not give a full explanation of AI.



The limited number of respondents who felt they could explain the term 'Artificial Intelligence' in detail is unsurprising given that AI spans a range of technologies and applications making it a challenging term to define.

### Digital familiarity

Those with **very low digital familiarity** (contacted via telephone interviewing) were **less likely** to say that they were able to explain the term 'Artificial Intelligence'; only 34% of these respondents said they would be able to give a partial or detailed explanation, compared to 68% of those with high digital familiarity.

### Gender

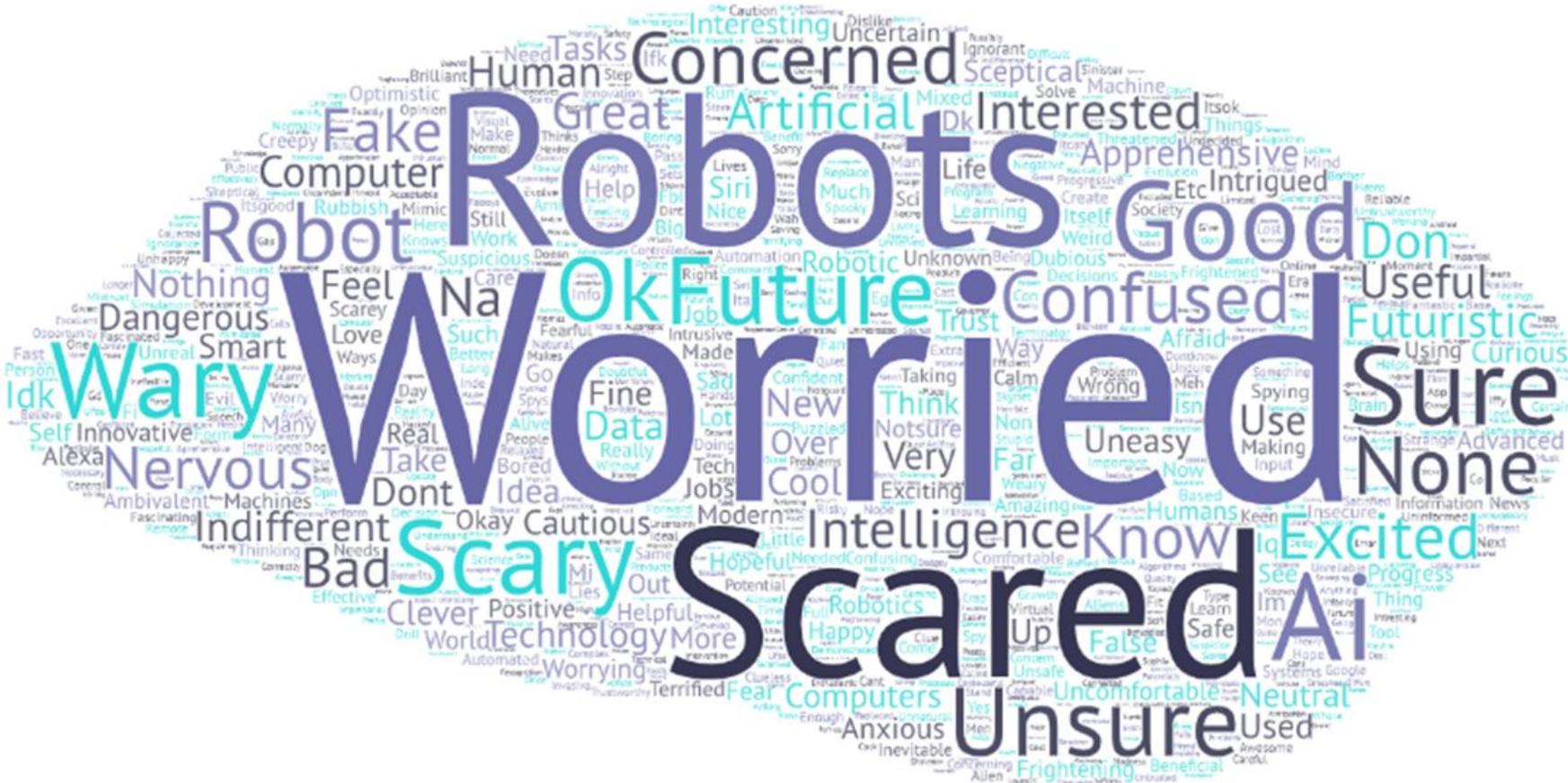
**Women** were also **less likely** to say that they could give a partial or detailed explanation of the term AI (56%) compared to men (70%). While this is a significant difference, it is also useful to note that women generally tend to rate their knowledge lower than men in surveys.

### Socioeconomic class

Respondents from a **lower socioeconomic background** were also **less likely** to say that they could give at least a partial explanation AI; 56% of those from lower social grades (C2DE) said this compared to 68% of those classified as higher social grades (ABC1).

■ Never heard of ■ Heard of but could not explain ■ Heard of and could partially explain ■ Heard of and could explain in detail

The words respondents thought best represented their feelings about AI are dominated by worry, fear and concern, and are shaped by a view of AI as futuristic



**Feelings about 'Artificial Intelligence'**

Those who had previously heard of the term 'Artificial Intelligence' were then asked to enter a word or phrase that best represents how they feel about it.

Respondents strongly associated AI with 'robots' and 'computers' and there was a sense that AI will be a part of the 'future' rather than the present.

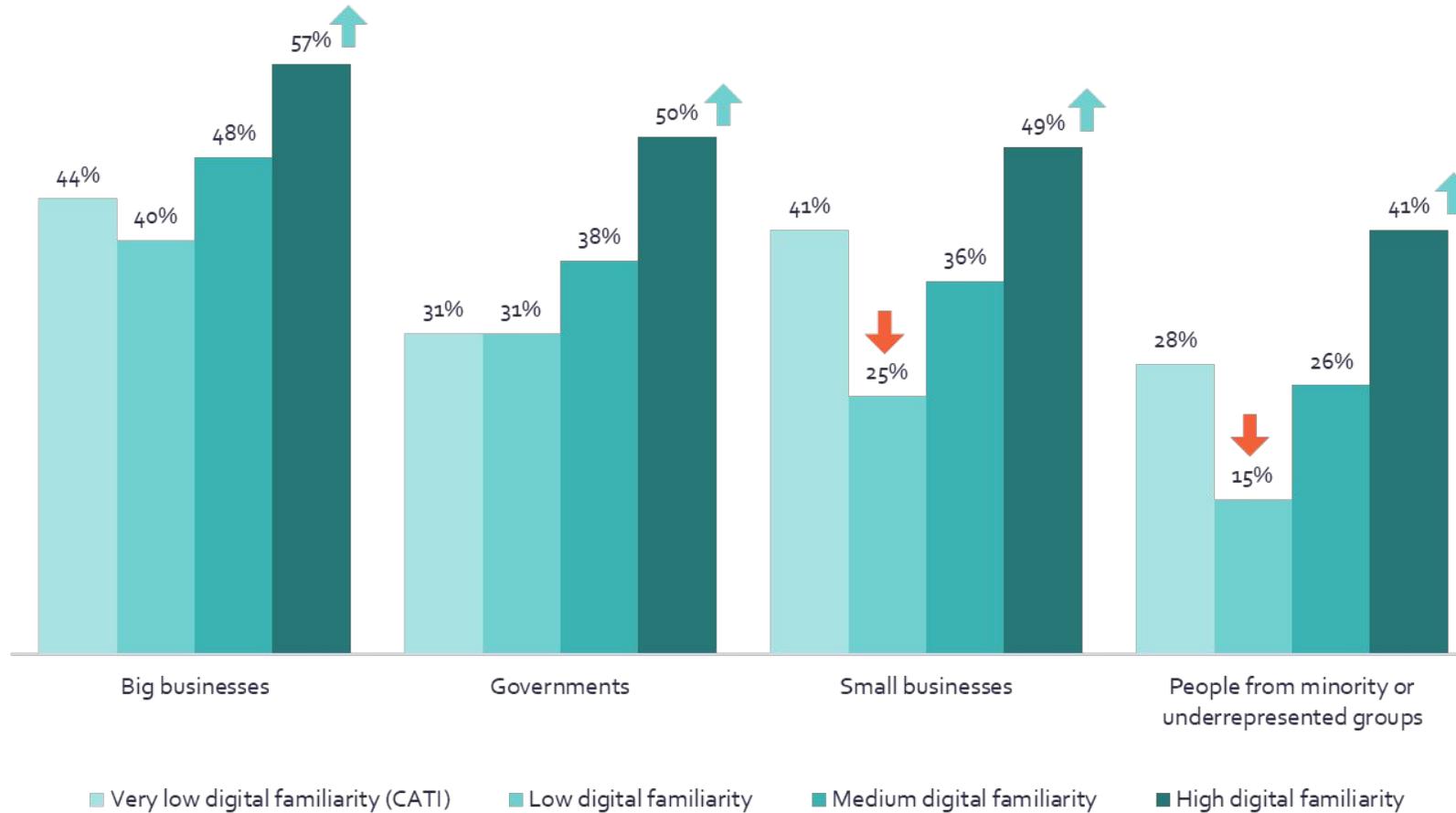
Looking to this future for many prompted feelings of apprehension. A large proportion of respondents used words such as 'worry', 'scary', 'concern' and 'nervous'.

While there was widespread uncertainty ('unsure', 'confused', 'wary') about AI amongst responses, there were some who felt more positively about AI, using words such as 'good' 'excited' and 'interested'.



# Respondents with lower digital familiarity reported less positive expectations for the impact of AI on small businesses and minority groups.

Total: Positive impact of AI on different people or organisations – by familiarity



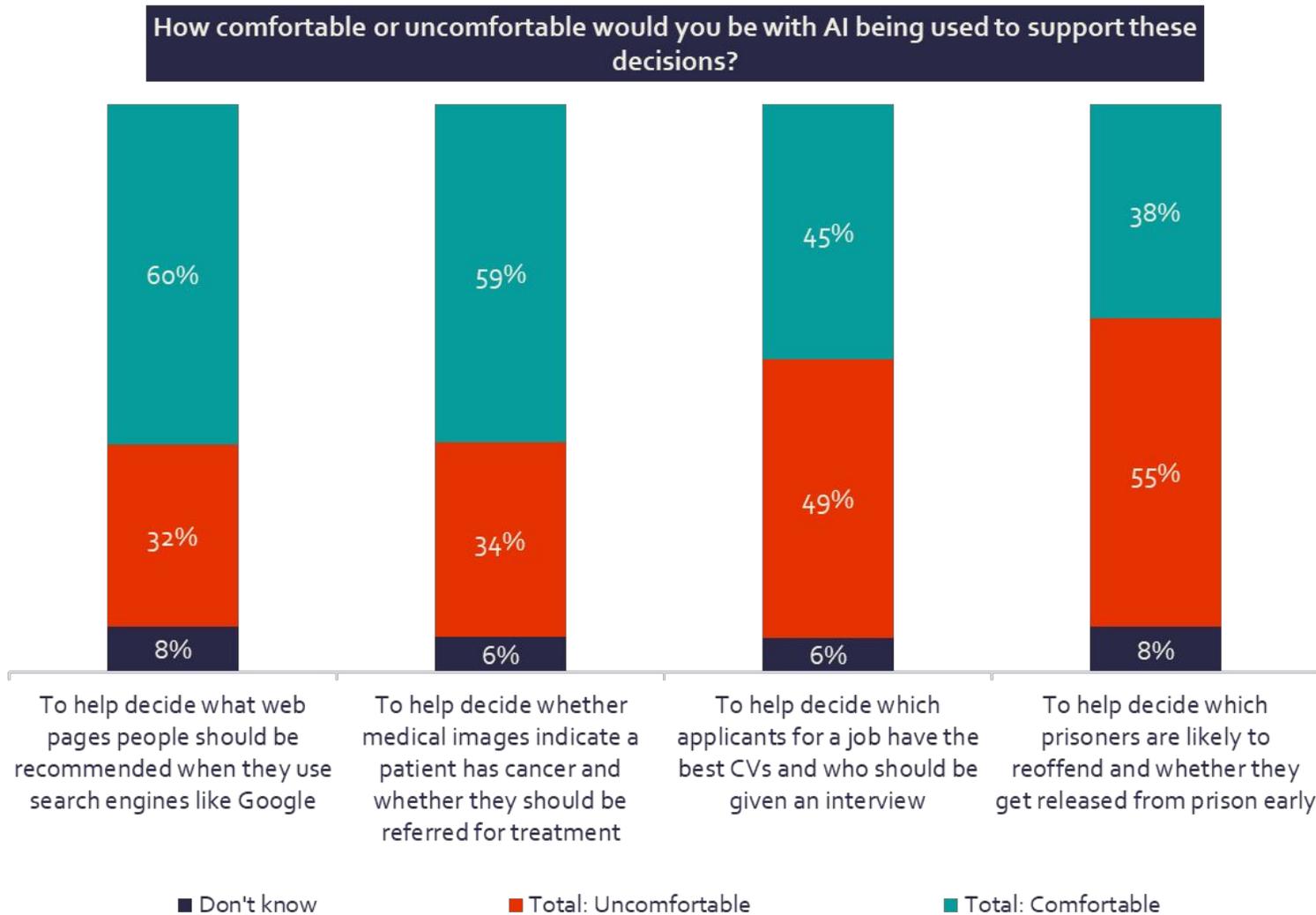
## Impact of AI by digital familiarity

41-57% of those with the highest levels of digital familiarity said that they think AI will have a positive impact on all of the types of people and organisations tested. These figures are all significantly higher than other respondents.

While those with low levels of digital familiarity were less likely to say AI will have a positive impact on all the options tested, these differences are particularly significant in relation to small businesses (25% say positive vs. 36% overall) and people from minority or underrepresented groups (15% vs. 26% overall).

Arrows indicate a statistically significant difference from all other subgroups within the same category

# Comfort with the use of AI is varied. A high number of respondents reported being uncomfortable with regular applications of AI, such as internet searching.



## Comfort with AI use cases

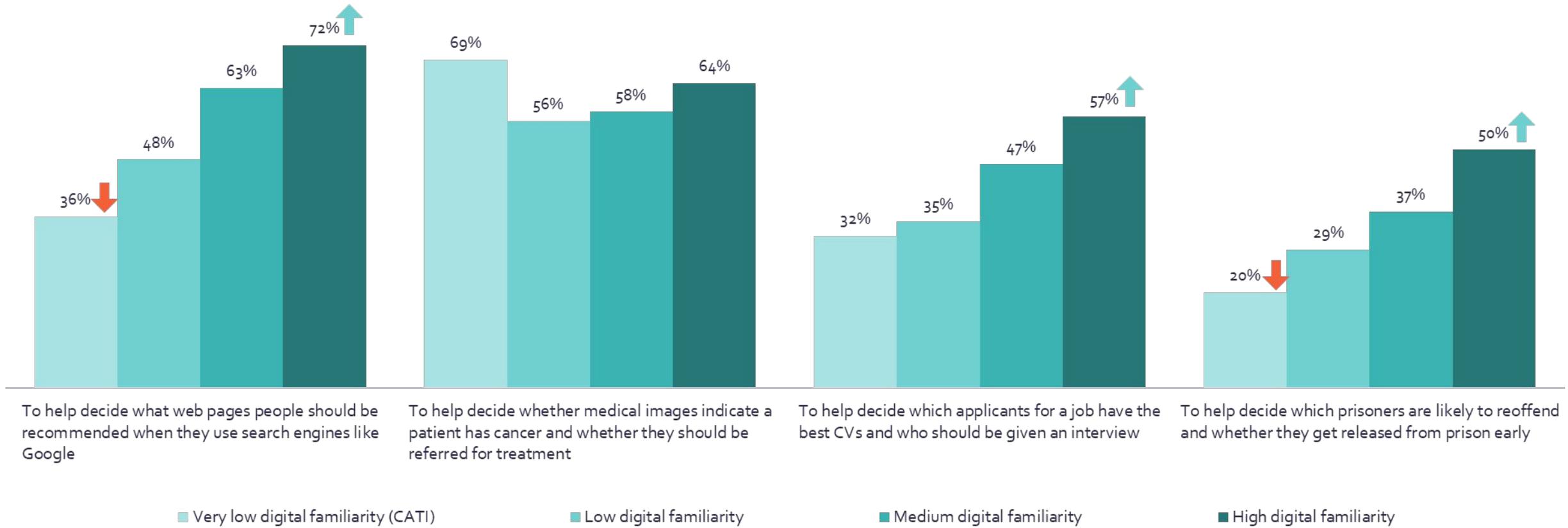
Most respondents said that they were comfortable with AI being used to help decide which web pages people should be recommended when they use search engines (60%) and whether medical images indicate a patient should be referred for cancer treatment (59%).

However, more respondents said they were uncomfortable with AI being used in other hypothetical scenarios than comfortable. This included helping decide which applicants should be offered a job (49% uncomfortable vs. 45% comfortable) and which prisoners should get released from prison early (55% vs. 38%). This greater discomfort may be because these scenarios could be seen as higher risk due to subjectivity in the decision making process and the potential to negatively impact people's futures.

Full question wording: *Below are a number of scenarios where Artificial Intelligence (AI) could be used to support decision making. In each scenario, imagine that there is a human who would be interpreting the recommendation made by AI, and that using AI would mean that the decision-making process is quicker than a human alone trying to do the task. The AI would be created to reduce the risk of an unfair decision by using lots of real-world information that is considered to be unbiased across different groups in society.*

# Respondents with lower digital familiarity were less comfortable with AI being used in most use-cases.

Total: Comfortable with AI in decision-making scenarios – by familiarity



Arrows indicate a statistically significant difference from all other subgroups within the same category

# 6. Impacts of COVID-19



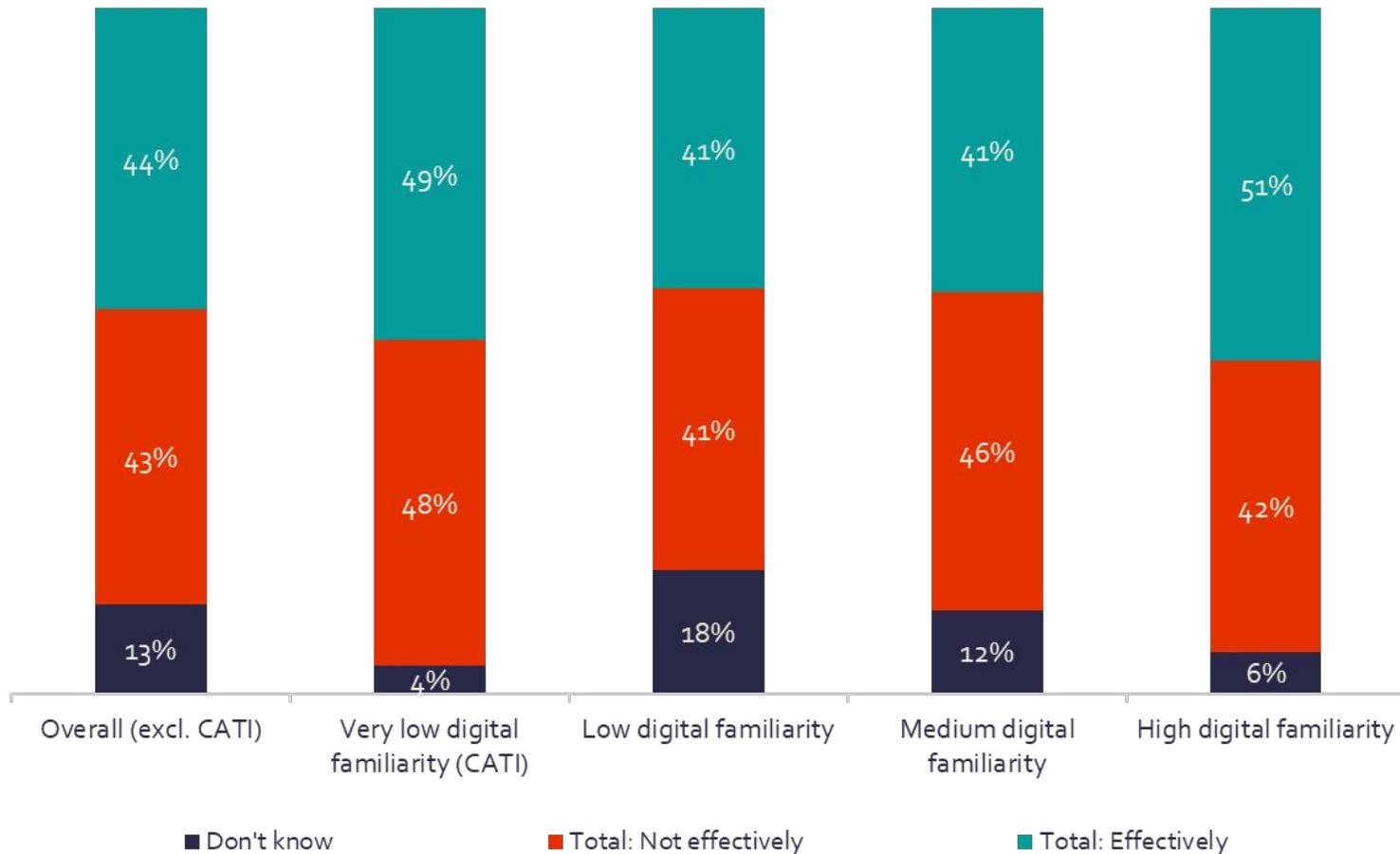
## 6. Impact of COVID-19: key insights

Views on the use of data to address challenges presented by the COVID-19 pandemic are mixed;

- Respondents were split on how effectively they felt the Government had been in using data to deal with the pandemic; 44% of respondents reported that they thought the Government had used data effectively, compared to 43% who reported that data had not been used effectively.
- Those who do not trust the Government to act in their best interests were also more likely to believe data has not been used effectively to deal with the COVID-19 pandemic.
- In general, respondents' trust in Government to use data responsibly and ethically following the pandemic remains unchanged. The highest proportion of respondents (42%) reported their views remain 'about the same'. However, more respondents (32%) reported a loss of trust in Government to use data responsibly and ethically following the pandemic than a gain (21%).

# Respondents were broadly split on how effectively the Government has used data to deal with the COVID-19 pandemic.

How effectively, if at all, do you think the Government has used data to deal with the COVID-19 pandemic?



### Trust in Government

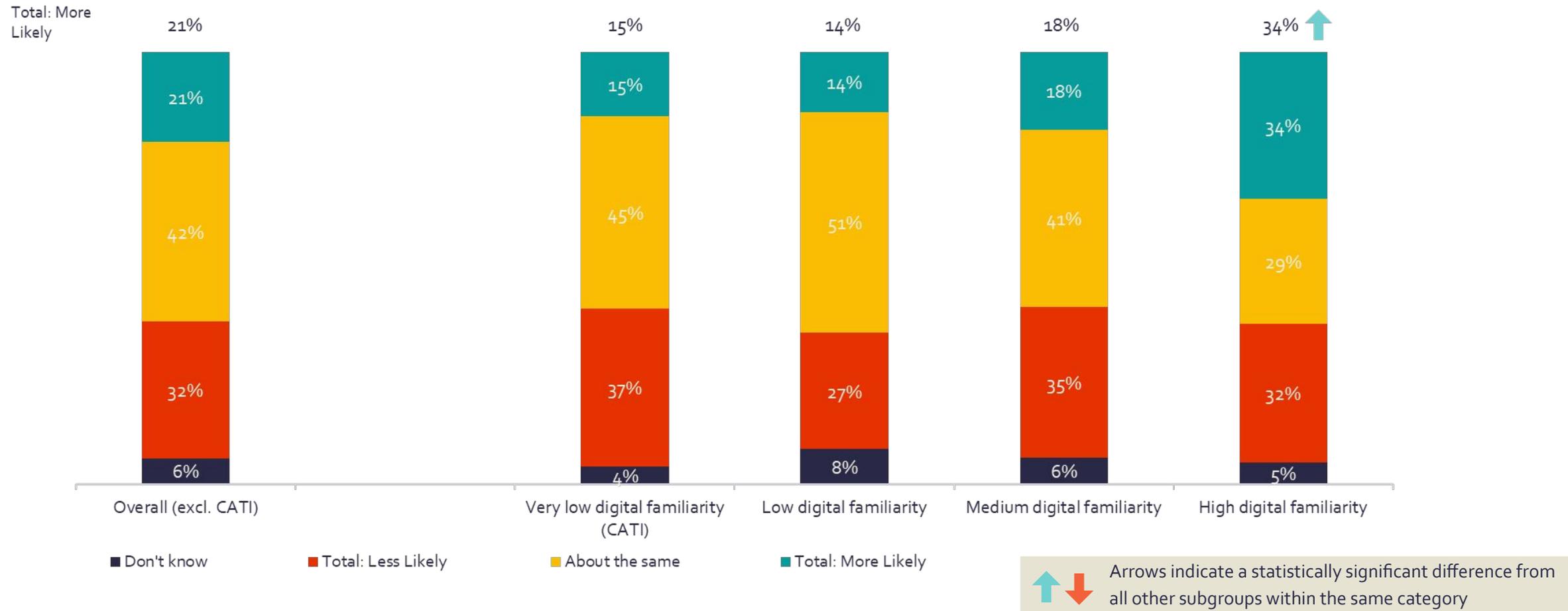
Unsurprisingly, respondents who do not trust the Government to act in their best interests were much less likely to say that the Government used data effectively to deal with COVID-19; 26% of these respondents reported that data had been used effectively in the pandemic response compared to 70% for respondents who trust the Government to act in their best interests.

### Data and technology in job role

Those whose **job involves data and technology** were much more likely to say they that the Government has used data effectively to deal with the COVID-19 pandemic (50% vs. 38% of those who don't use online tools as part of their job role).

More respondents reported they were less likely to trust the Government to use data responsibly and ethically following the pandemic than reported they were more likely. Although, trust increased amongst those with higher digital familiarity.

Trust in the Government to use data responsibly and ethically following the COVID-19 pandemic



# Contact details



**Holly Clarke**  
Public Attitudes Researcher  
holly.clarke@cdei.gov.uk

**Hannah Spiro**  
Head of Public Attitudes  
hannah.spiro@cdei.gov.uk

**Ben Lyons**  
Head of External Affairs & Insight  
ben.lyons@cdei.gov.uk



**Oliver Wright**  
EVP, Public  
oliver.wright@savanta.com

**Nicola Archer**  
Director  
nicola.archer@savanta.com

**Emma Levin**  
Senior Consultant  
emma.levin@savanta.com

**George Singleton**  
Senior Executive  
george.singleton@savanta.com

# Acknowledgments

We would like to thank the following colleagues for their contributions, which included: input into the development of the PADA Tracker Survey, providing analysis, and offering their time to review and give feedback on this report:

- Todd Hartman, Professor of Quantitative Social Science at the University of Manchester
- Baobao Zhang, Assistant Professor of Political Science at Syracuse University
- Helen Kennedy and Mark Taylor, *Living With Data* researchers at the University of Sheffield
- The Ada Lovelace Institute, with special thanks to Aidan Peppin and Kira Allmann
- Olly Carr, Senior Policy Adviser, NHS Transformation Directorate

# Annex A - Deep dive sections



# Trust in actors

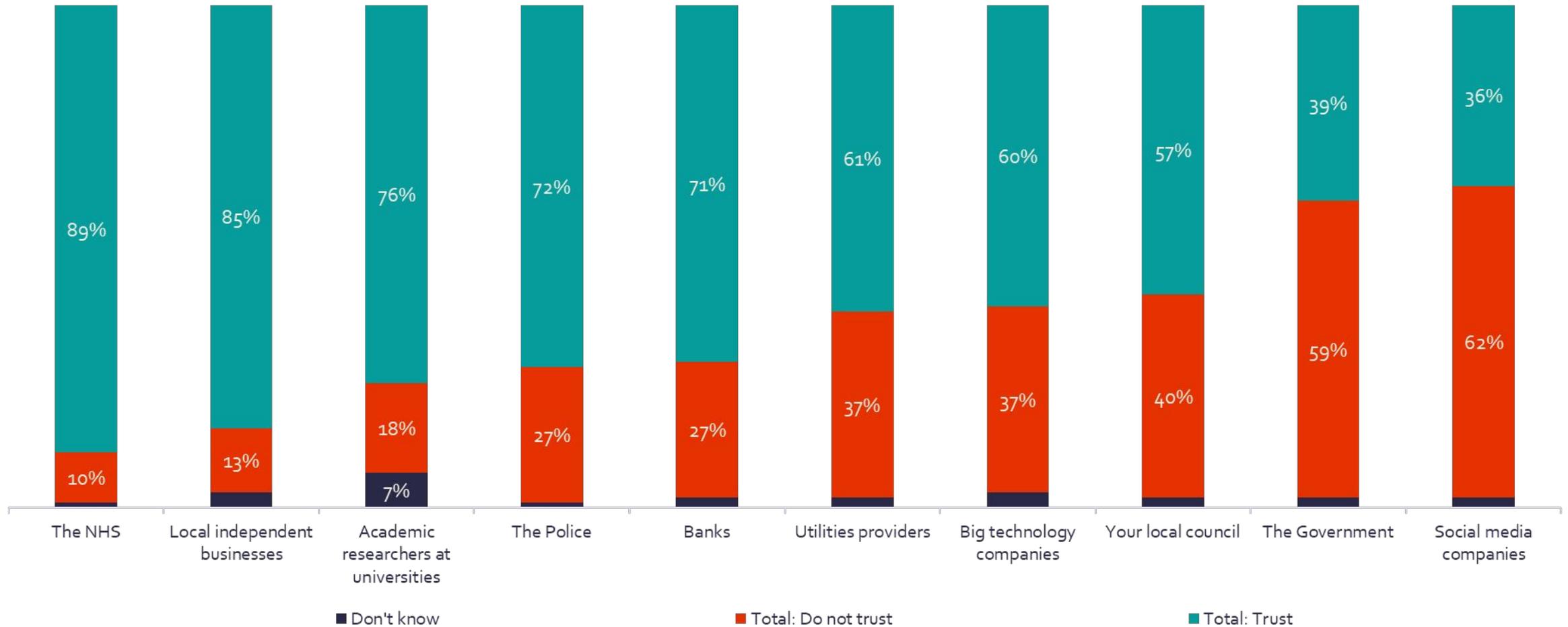


# Trust in actors: key insights

- **Respondents reported that they lack trust in a number of public and private organisations.** The NHS stands out as a highly trusted organisation.
- **Respondents with the lowest digitally familiarity expressed high trust in the police and banks, but were less likely to trust academics.**
- **Older respondents were more trusting of public sector organisation.** They were also more likely to differentiate between different types of private sector organisations compared to younger people.
- **Respondents' trust in many organisations to use data ethically was limited.** This is despite many using services that mean personal data is held by these organisations.

# Significant proportions of respondents reported general mistrust in a number of public and private organisations; under half trust the Government or social media companies.

To what extent, if at all, do you generally trust the following organisations to act in your best interest?



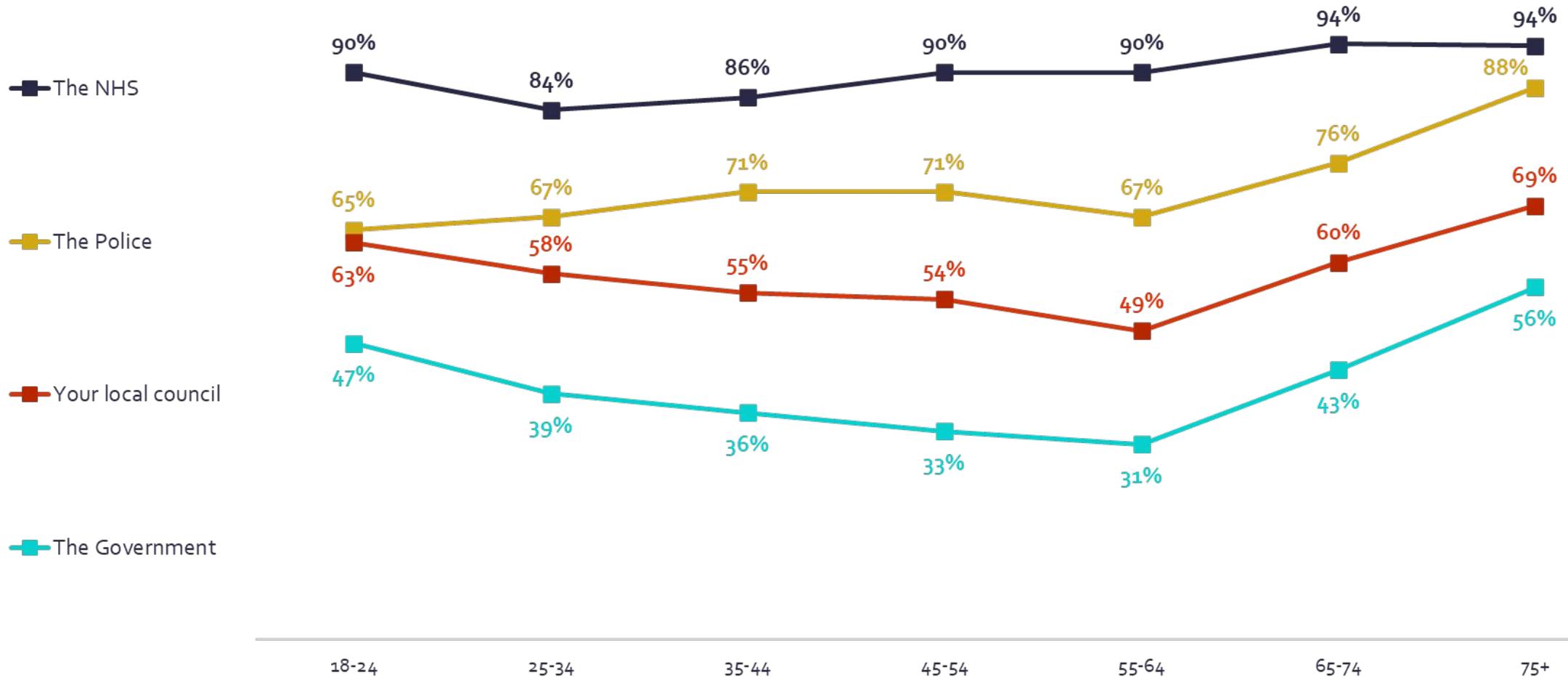
# Respondents with lowest digital familiarity expressed high trust in the police and banks, but were less likely to trust academics.



Arrows indicate a statistically significant difference from all other subgroups within the same category

# Those over the age of 65 were more likely to trust public sector organisations, with the exception of the NHS, where trust was consistently high across the age groups

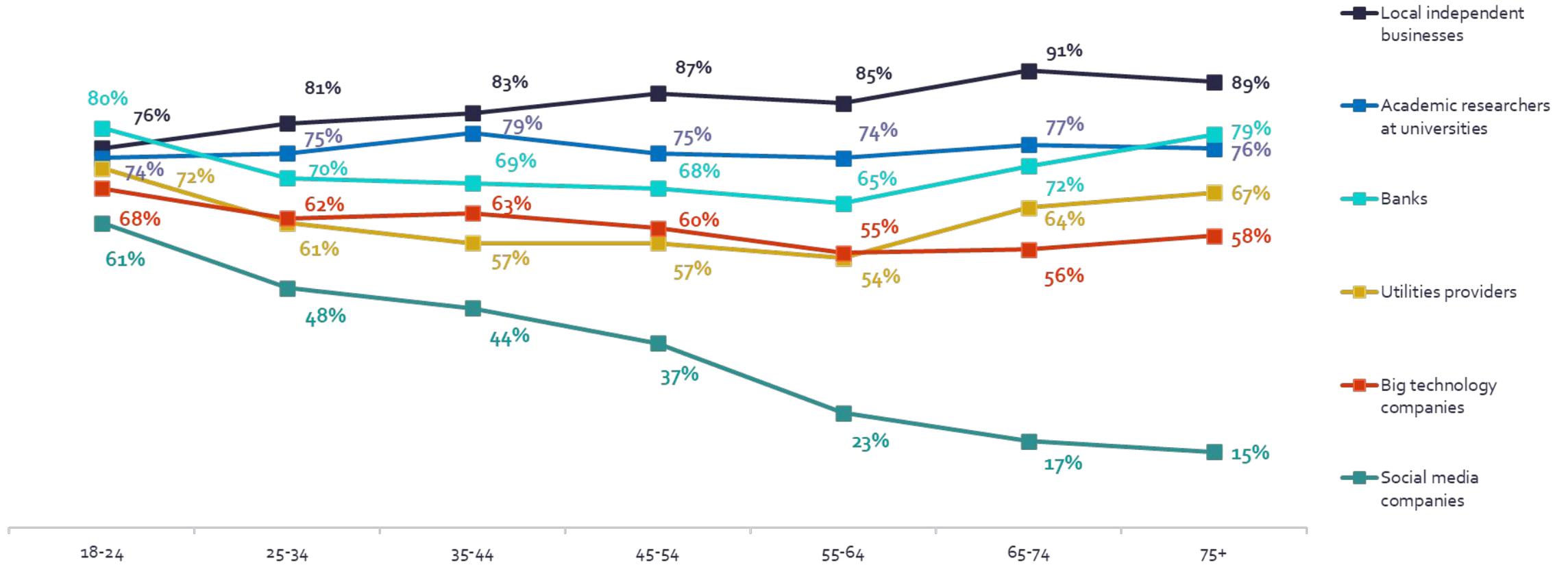
Total: Trust in public sector organisations to act 'in your best interest' – by age



Note: This chart shows categorical data rather than continuous data.

Older respondents differentiated more between different types of private sector organisations, while younger respondents had a narrower range of trust in these organisations.

Total: Trust in other organisations to act 'in your best interest' – by age



Note: This chart shows categorical data rather than continuous data.

# Digital behaviours



# Digital familiarity and the digitally excluded - a recap

A proxy measure for **digital familiarity** is used in this report. We use digital familiarity to compare responses to questions. The very low digital familiarity band is comprised of the 'digitally excluded' telephone sample. The low-high familiarity bands are based on responses to questions about respondents' frequency of online activities and confidence using devices online\*. Respondents were given a composite score out of 18 and categorised into bands based on their score. The demographic differences of these bands reflect demographic differences in digital familiarity, and are detailed in the ['Methodology'](#) section of this report.

## Number of respondents

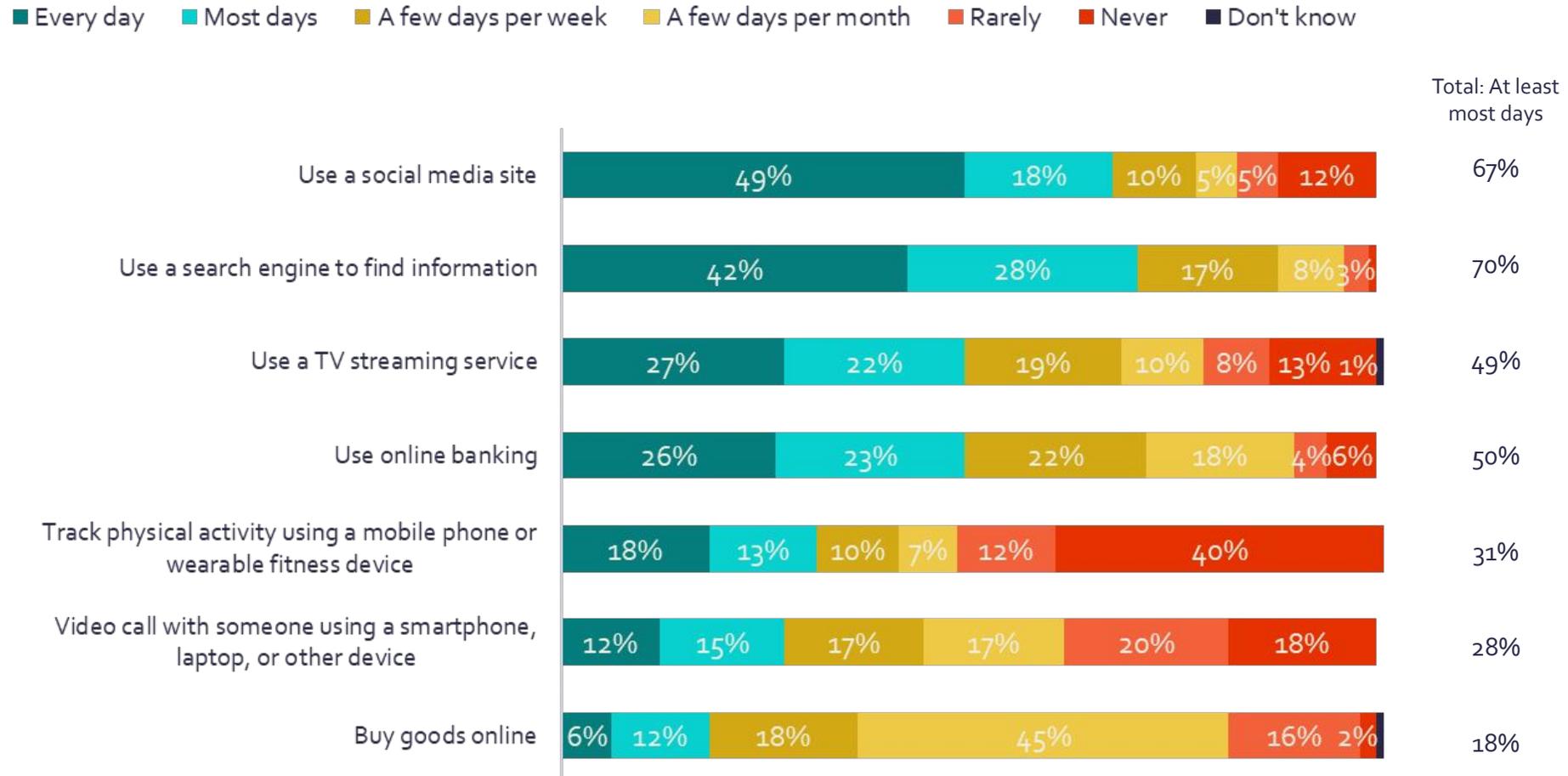
<b>Very low digital familiarity</b> the 'digitally excluded' sample (telephone booster) 'CATI'	200
<b>Low digital familiarity</b> (Score of 0-10/18)	1,476
<b>Medium digital familiarity</b> (Score of 11-14/18)	1,697
<b>High digital familiarity</b> (Score of 15-18/18)	1,084

## Digitally excluded telephone booster sample

200 respondents with very low digital familiarity were contacted and interviewed via the telephone. This telephone booster sample allows 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, to be captured. Respondents were recruited based on not using, or having low confidence in at least three of the following five following areas: using email, using online banking, preferring face to face banking, entering personal details online and getting help using the internet. Due to age differences in digital skills and the small samples size of this group, respondents are on average much older than the online survey sample and are exclusively white ethnicity. There are other intersecting reasons that individually may be digitally excluded that, therefore, may not be fully captured by this sample and merit further research.

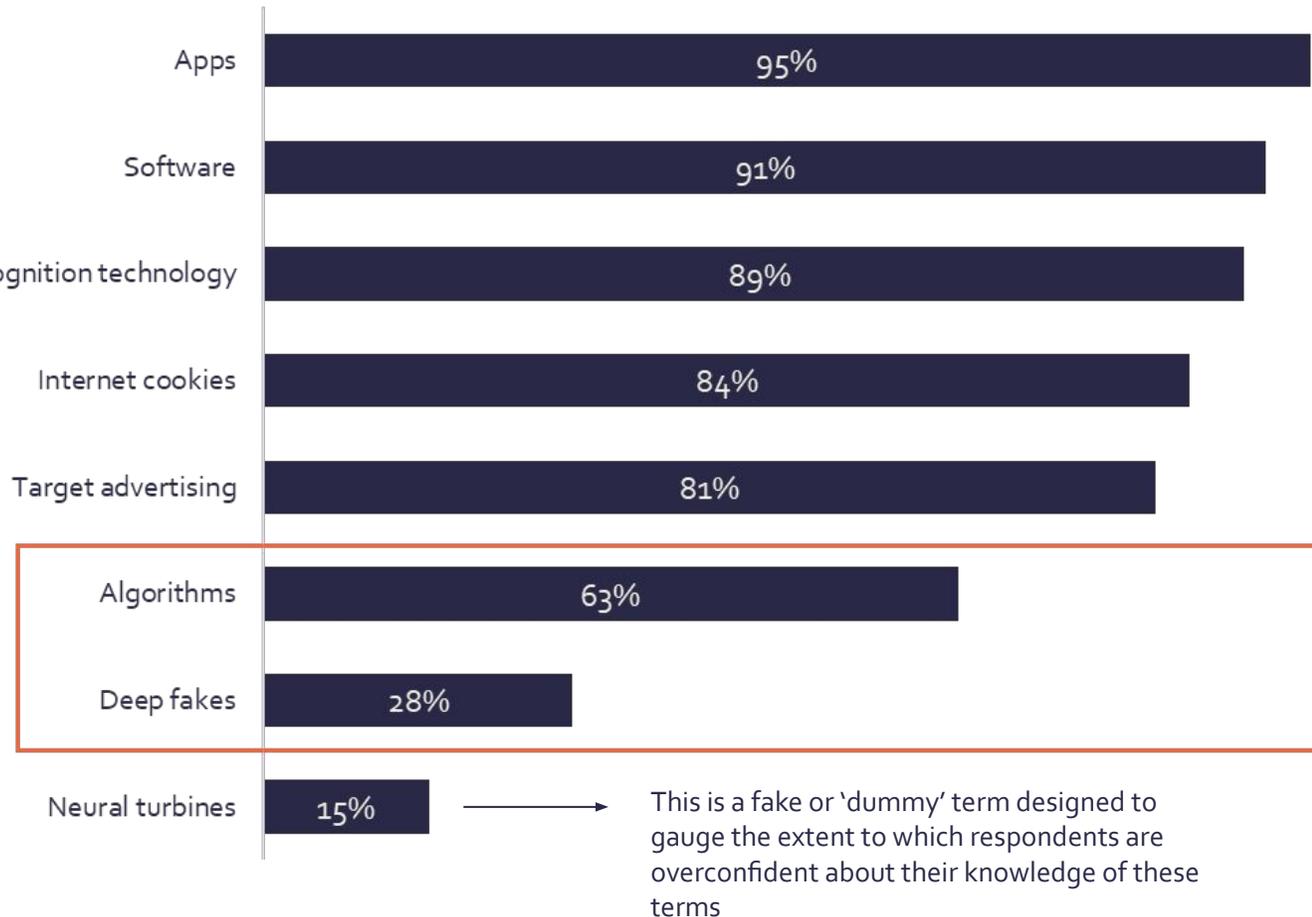
# Respondents report undertaking a range of digital activities on a regular basis, evidencing a general digital competence across much of the population.

**In a typical month, how often, if at all, do you do the following?**



# Respondents' claimed that their understanding of common technology terms was high.

If someone said the following term to you, would you understand what they meant by it?



## Device usage

Those who said they **use devices at least most days** were significantly more likely to claim understanding of most terms compared to those who say they use devices less than this. Yet, they were significantly less likely to claim knowledge about the dummy term 'neural turbines'. This dummy term was used to test the level of over claimed knowledge amongst respondents. .

## 'Algorithms'

Those who **don't trust big technology companies** are were significantly more likely to claim to understand the term 'algorithms'.

## 'Deep fakes'

Claimed understanding of the term 'deep fake' decreased sharply with age, with the age difference at its starkest at around age 55 (45-54: 30%; 55-64: 15%). Unlike the other terms, claimed understanding of 'deep fakes' has an *inverse* relationship to device usage, with **those using devices less than most days** being significantly more likely to say they understand the term than those who use them at least most days (50% vs. 27%).

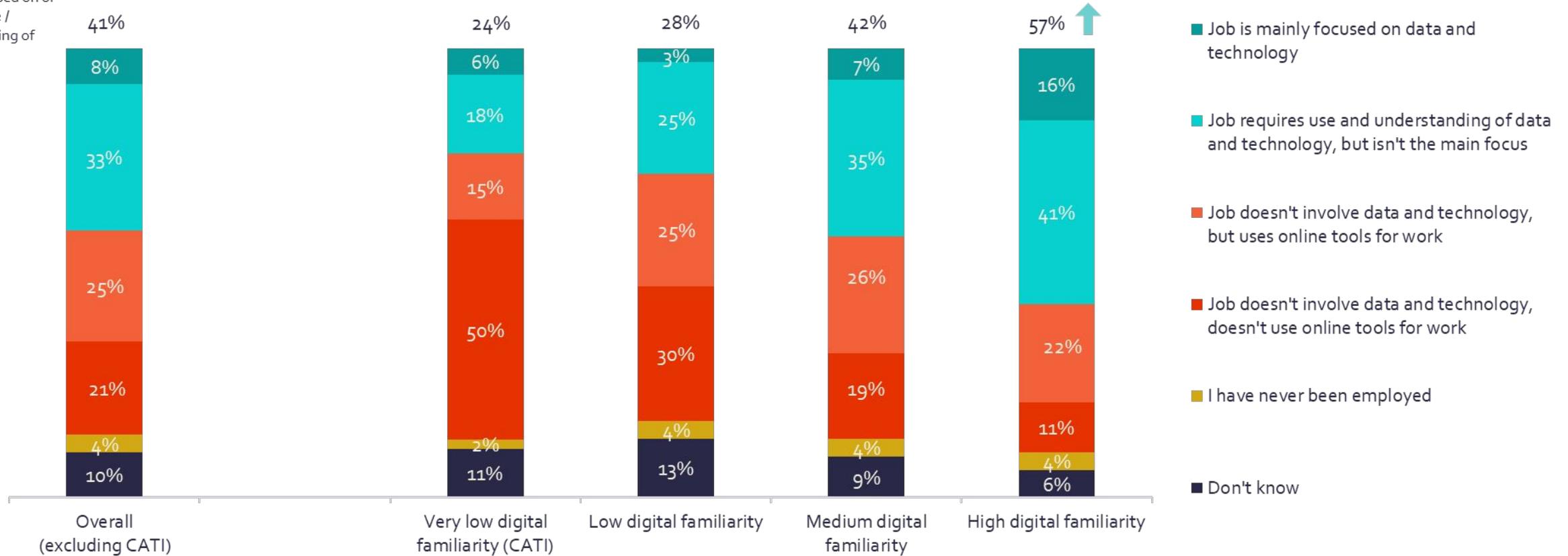
## Neural turbines and overconfidence

Claimed understanding of most terms increased significantly amongst those who are **confident doing online activities**, those who **use data and technology in their jobs**, and **men**. However, this included the dummy term – neural turbines – which should be taken into account when analysing these results. **Londoners** and **under 45s** were also significantly more likely to claim they understand 'neural turbines'.

# A large proportion of respondents reported that their job requires an understanding of data and technology, highlighting the role of data in both professional and personal lives.

Which of the following descriptions best describes your current or most recent job?

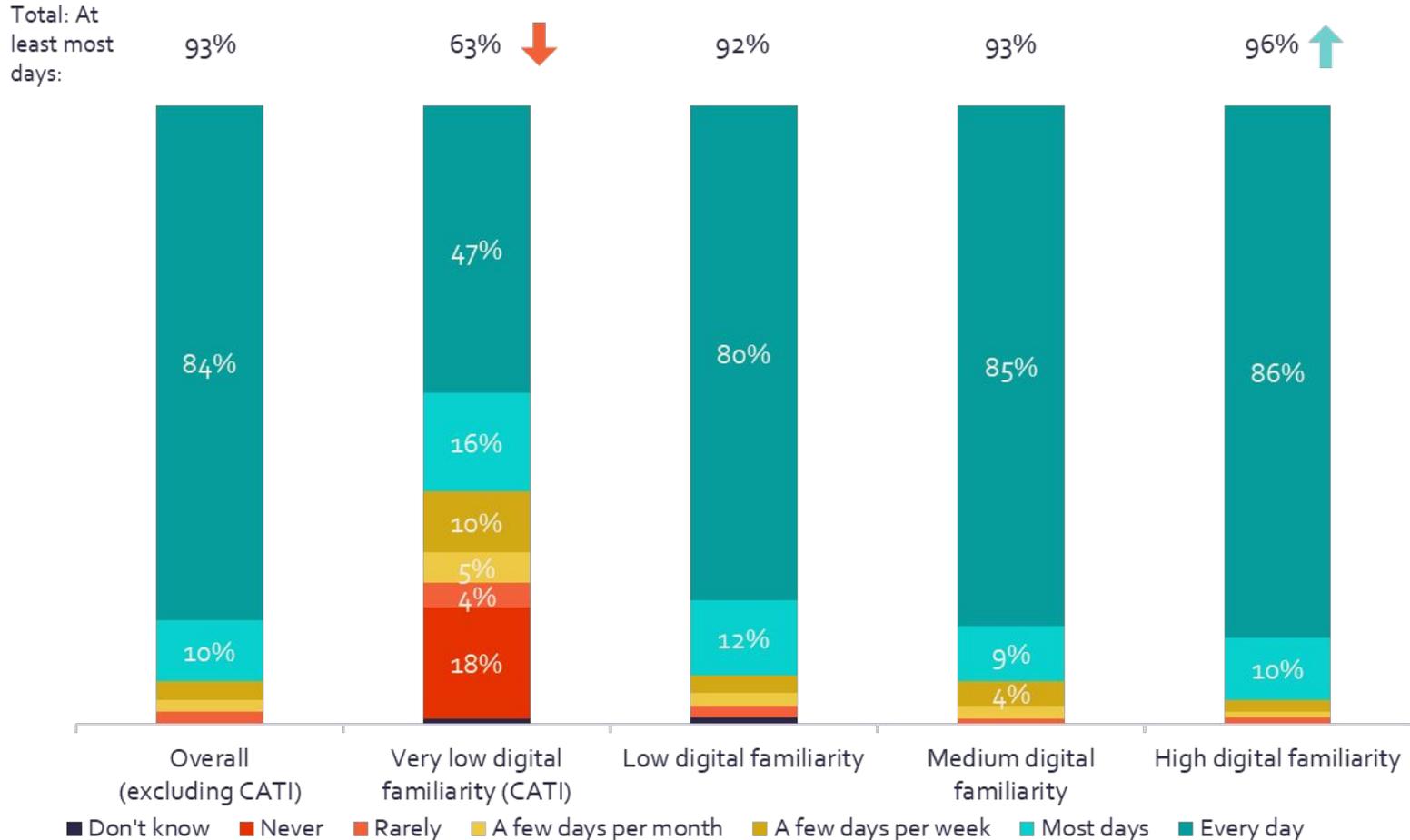
Total: Focused on or requires use / understanding of data and technology



Arrows indicate a statistically significant difference from all other subgroups within the same category

# While the majority of respondents reported frequent internet use, this is not true for all, meaning some will be unable to use services that require digital access.

In a typical month, how often, if at all, do you use the internet (for work or personal use on a computer, smartphone, or other device)?



Those with very low digital familiarity:



Use the internet every day, compared to 84% of the general population



Say they never use the internet. And 4% say they only rarely use the internet.

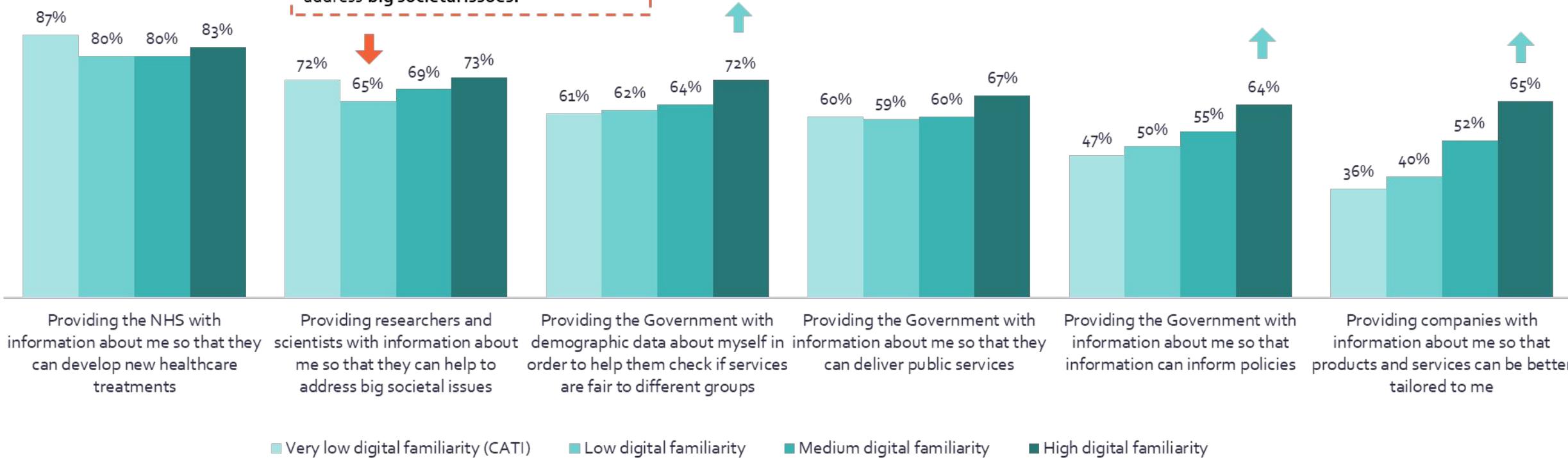
↑ ↓ Arrows indicate a statistically significant difference from all other subgroups within the same category

# Those with the highest digital familiarity were more likely to feel comfortable providing personal information about themselves.

**Total: Comfortable with providing personal data in different instances – by familiarity**

Those with **low** digital familiarity are less likely to be comfortable providing personal data to researchers/scientists to help address **big societal issues**.

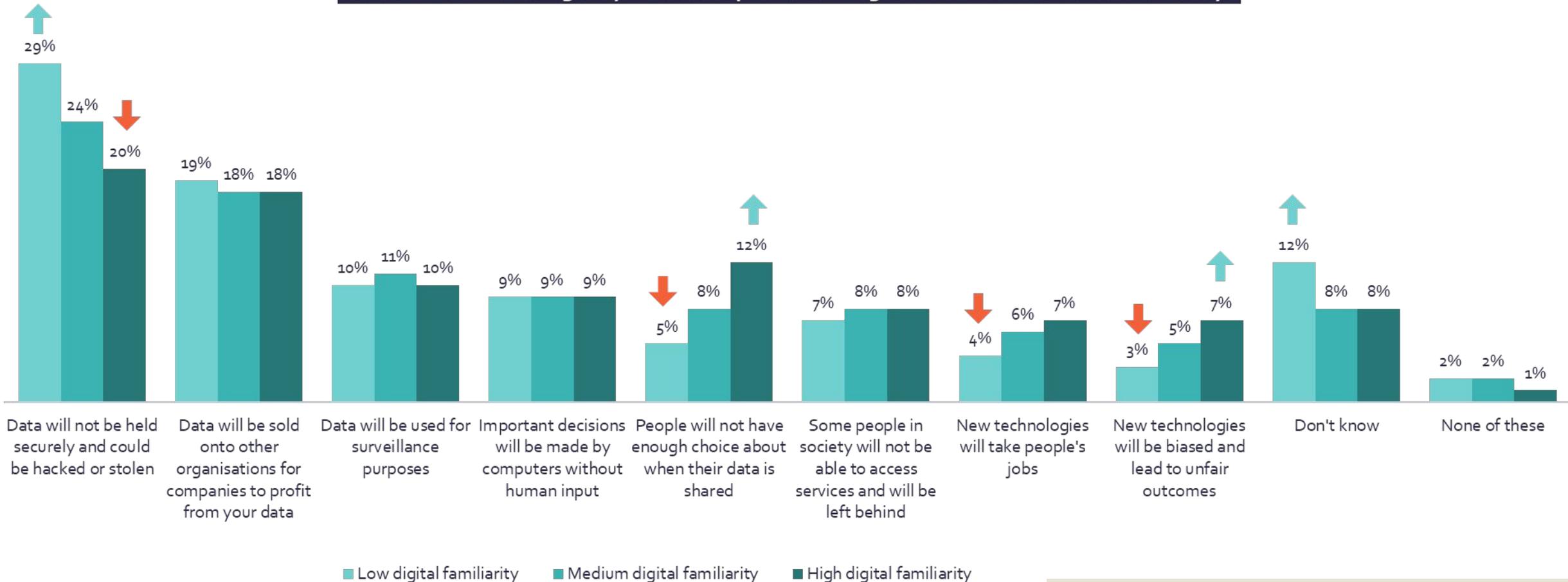
Those with **high** digital familiarity are more likely to be comfortable providing personal data to the Government to help deliver **public services or inform policies**, or to companies to help them **tailor their products and services**.



Arrows indicate a statistically significant difference from all other subgroups within the same category

# While data security was seen as the greatest risk across familiarity levels, it was especially prevalent among those with low digital familiarity.

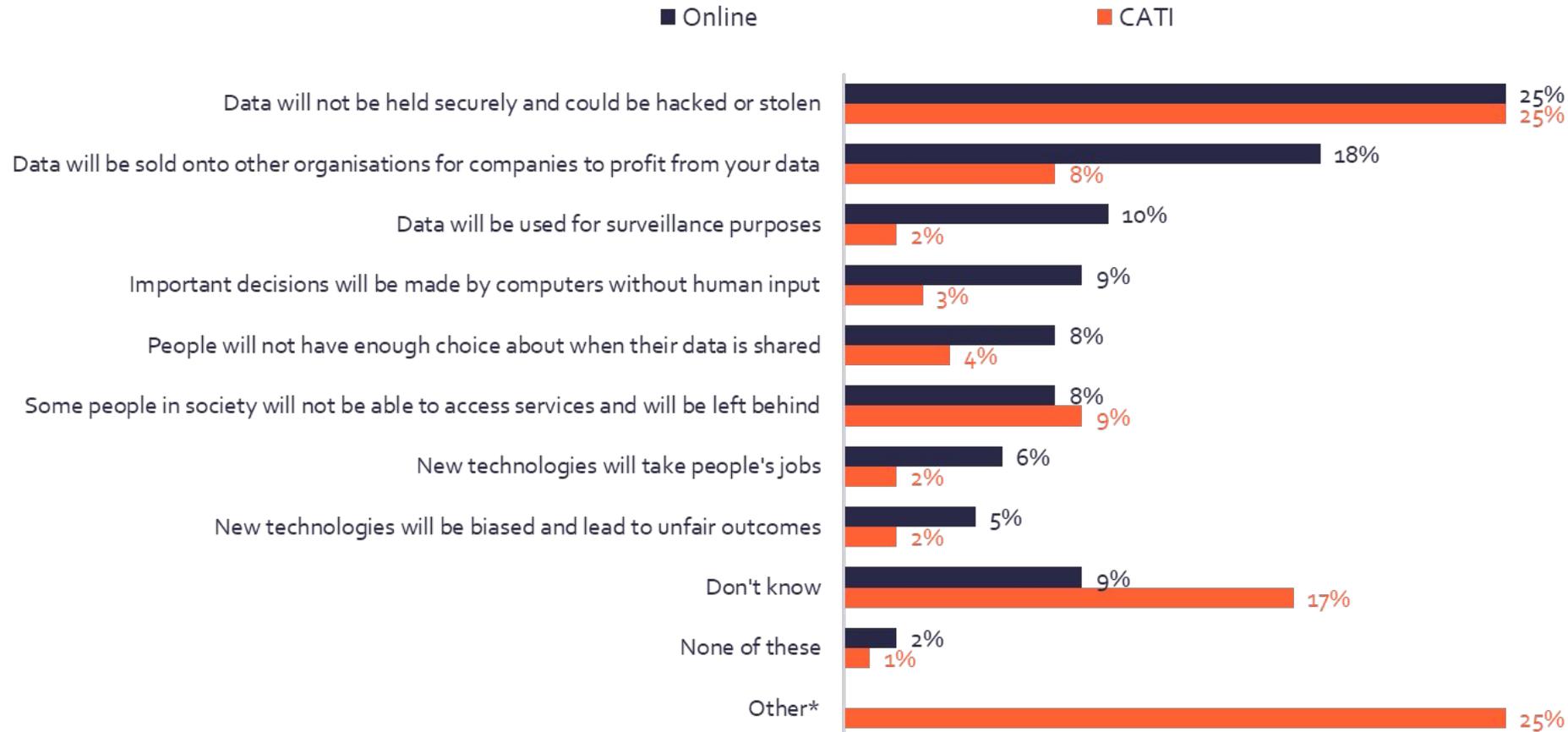
Which of the following do you think represents the greatest risk for data use in society?



Arrows indicate a statistically significant difference from all other subgroups within the same category

# The least familiar respondents were less aware of where the risks for data use in society lie and note specific concerns about data misuse and low trust.

Which of the following do you think represents the greatest risk for data use in society?

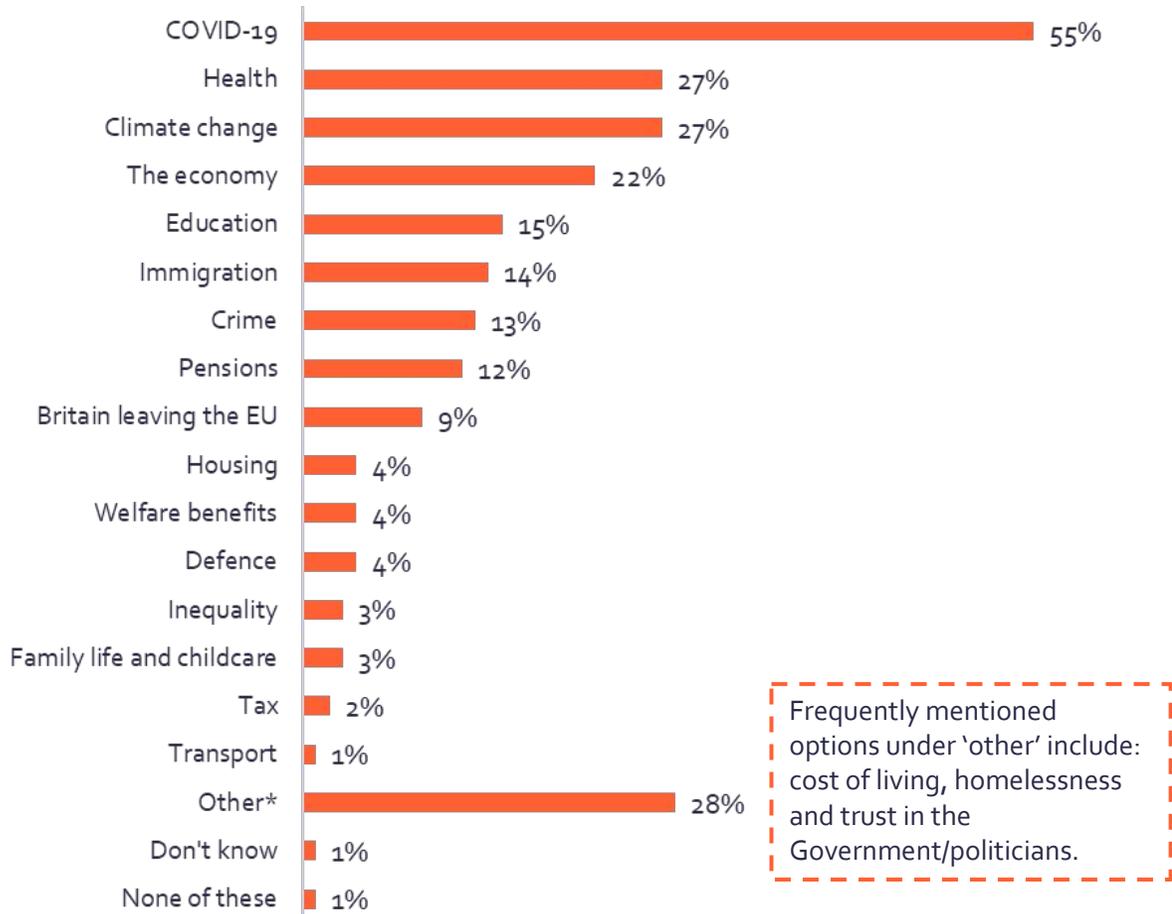


Frequently mentioned options under 'other' include: fraud and scams, privacy concerns, and lack of trust in Governments and organisations to manage data.

\*Higher 'other' scores compared with the online survey reflect the fact this question was asked openly during telephone interviews

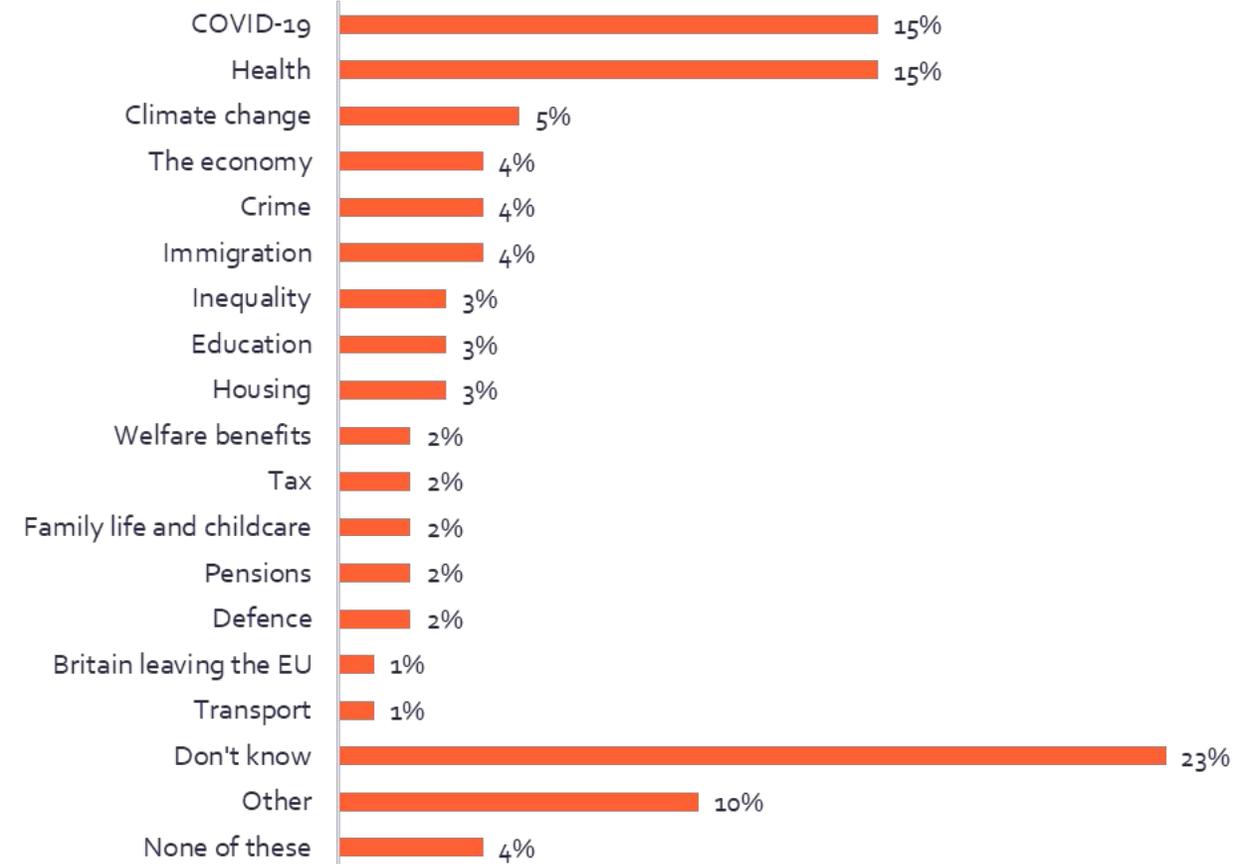
# Those with very low digital familiarity are largely consistent with the wider public, but are less sure about where data presents opportunities to make improvements.

Most important issues facing the country  
(% selecting top 3)



Frequently mentioned options under 'other' include: cost of living, homelessness and trust in the Government/politicians.

Issues where the use of data presents the greatest opportunity for making improvements that benefit the public



\*Higher 'other' scores compared with the online survey reflect these questions were openly during telephone interviews

# Technology user segments



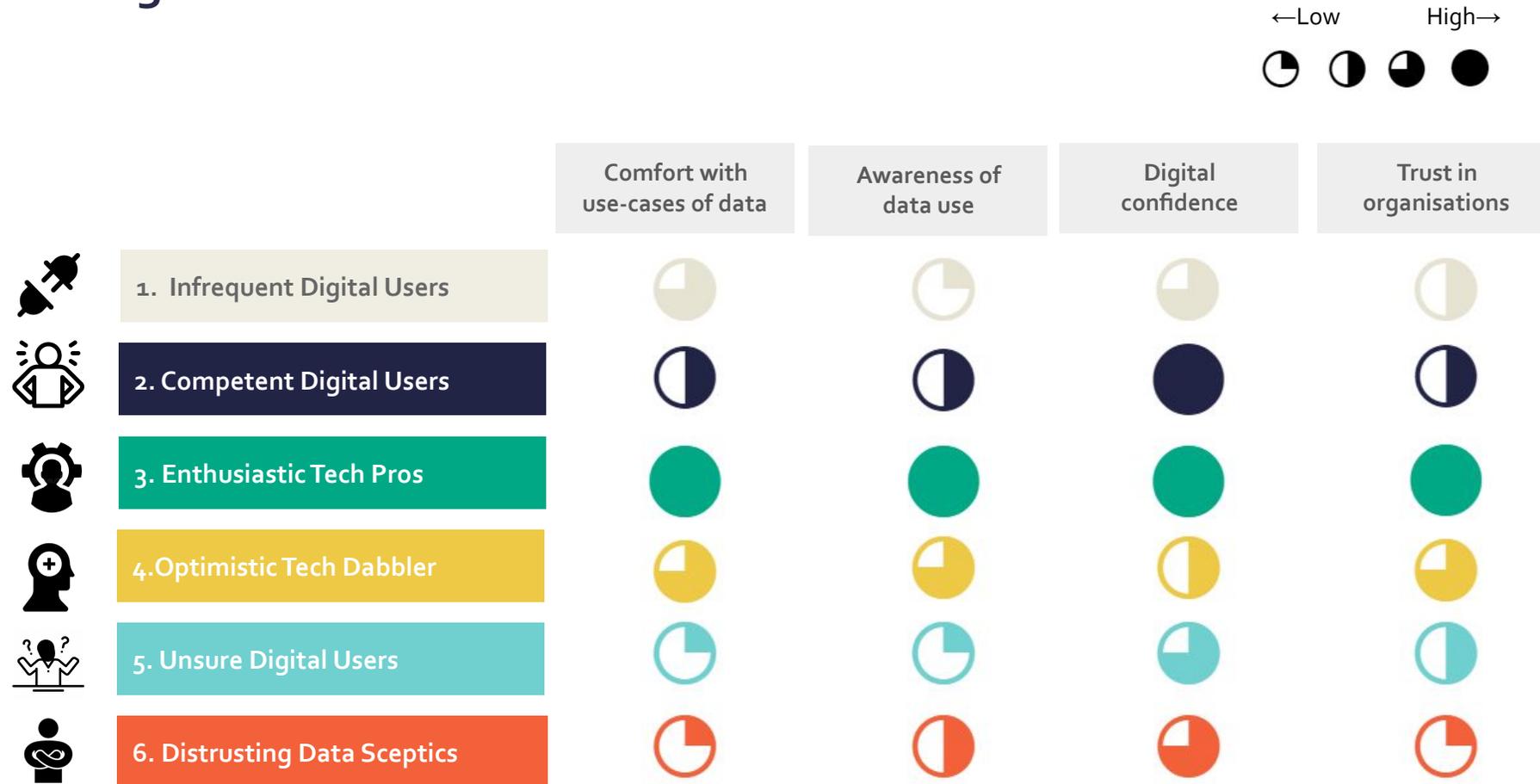
## CDEI undertook an attitudinal segmentation to identify six segments of the population that differ based on their behaviours and attitudes towards data and AI.

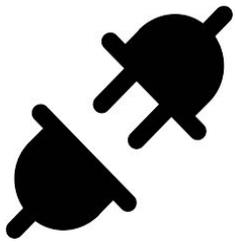
This included survey questions on\*:

- Trust in organisations
- Use of data and technology
- Awareness of data and confidence using technology
- Comfort with use-cases of data
- Perception of the risks and benefits of data use
- Attitudes towards AI

These questions specifically relate to attitudes about data. No other socio-demographic information was included to avoid segmenting the sample into non attitudinal classes, for example based on age or gender. The technique Latent Class Analysis (LCA) was used to segment respondents, this is a method of detecting and allocating respondents to groups based on the responses that they give to a set of survey items. The underlying idea is that group membership serves to determine the responses that they provide. We compared 'goodness of fit' statistics for models with different numbers of segments. Six segments was found to fit the data best [further detail available in the Annex]. The sample of digitally excluded individuals, contacted via telephone interviewing, were not included within this segmentation. This is because segmentation is based on the pattern of responses across different questions and some questions across the telephone and online samples are not directly comparable due to adaptations for telephone interviewing.

Respondents were segmented into six categories. Each segment expresses key differences in their comfort with and awareness of data use, digital confidence and the level of trust they report in organisations.





## Segment 1 – ‘Infrequent Digital Users’

23% of weighted online sample

### Key attitudes and behaviours

#### Less likely to say that they do online activities with any regularity.

Less likely to do activities such as buying goods online, using social media and using a search engine most days or more, compared with the general population.

#### Awareness of how data is used and collected is low.

Four in five (80%) say that they know just a little or nothing at all about how data about them is used and collected in day-to-day life (vs. 52% amongst general population).

#### Online security is a big concern.

More likely to say data not being held securely and being hacked or stolen as the biggest risk of data use in society (35% vs. 25% amongst general population).

#### Most likely segment to fall into the ‘low digital familiarity’ bracket.

#### Middling comfort with data use and optimism about the benefits of data use.

32% say collecting and analysing data is good for society and 83% are comfortable with Government using data to deliver public services.

### Data awareness

18%

know at least a fair amount about how data is used and collected about them in their day-to-day lives

### Trust of organisations

(Central) Government

Social media companies

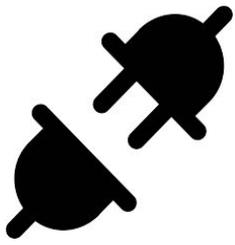
Big technology companies

To act in your best interests  
(NET: Trust)

53%

21%

63%



## Segment 1 – ‘Infrequent Digital Users’

23% of weighted online sample

### Demographic profile



No significant differences between the proportion identifying as female (49%) and the proportion identifying as male (51%).

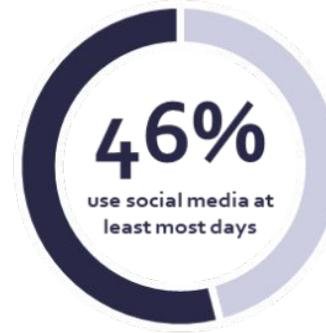


This segment is the oldest, with a **significantly higher mean age than all other segments** at 65 years old (overall mean age is 49).

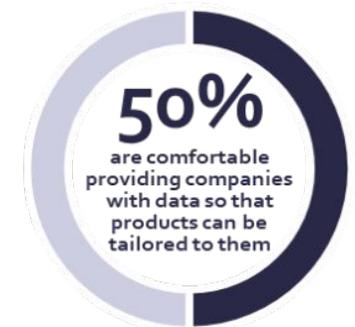
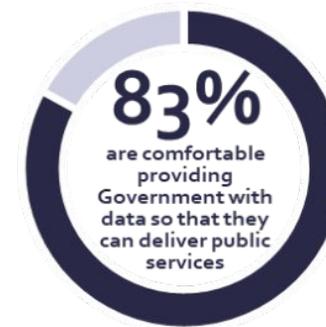


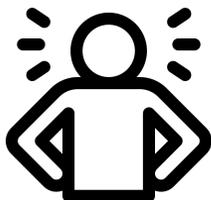
Members of this segment are **significantly more likely to be from the South-West** (11%) than almost all other segments. Conversely, they are significantly *less* likely to be from London (8%) than almost all other segments.

### Digital usage and confidence



### Comfort with data collection





## Segment 2 – ‘Competent Digital Users’

24% of weighted online sample

### Key attitudes and behaviours

**Most confident about using devices to do what they want to online.**

99% are at least somewhat confident doing this (vs. 91% amongst general population).

**More likely than other segments to work in a job role which requires them to use and understand data and technology.**

45% vs. 33% amongst general population.

**Frequent users of digital services.**

More likely to use online banking and TV streaming services frequently than any other segment except Segment 3.

**More likely to say they understand terms such as ‘algorithms’ or ‘software’.**

More likely to claim to understand terms than all other segments, except Segment 3.

### Data awareness

50%

know at least a fair amount about how data is used and collected about them in their day-to-day lives

### Trust of organisations

(Central)  
Government

Social  
media  
companies

Big  
technology  
companies

To act in your best  
interests  
(NET: Trust)

21%

31%

58%



## Segment 2 – ‘Competent Digital Users’

24% of weighted online sample

### Demographic profile

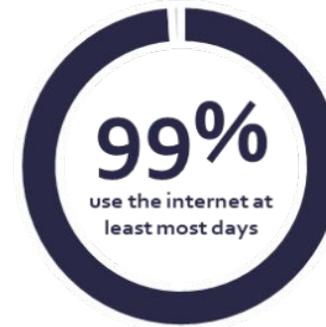
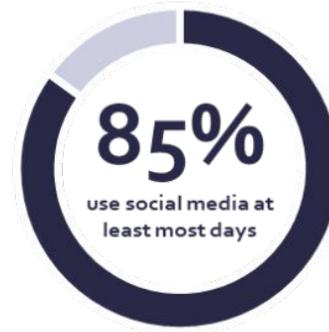


Members of this segments are significantly more likely to identify as **female** (58%) than male (42%), with this segment and Segment 5 being the only two majority female segments.

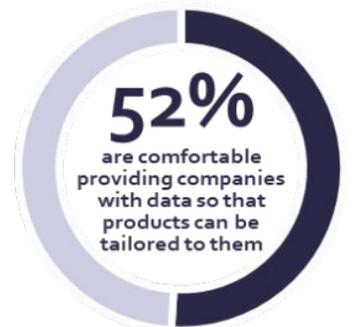
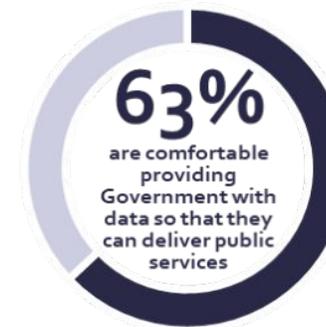


Gender split aside, the demographic makeup of this segment is fairly **close to that of the general population**. It skews **slightly younger** than average, but is not the youngest segment (mean age of 44, compared to the overall sample mean age of 49).

### Digital usage and confidence



### Comfort with data collection





## Segment 3 – ‘Enthusiastic Tech Pros’

18% of weighted online sample

### Key attitudes and behaviours

**More likely than all other segments to do several digital activities most days or more.**

Including video calling (54% vs. 28% amongst general population); tracking physical activity using wearables (51% vs. 31%); and banking online (72% vs. 50%).

**Most likely to fall into the ‘high digital familiarity’ bracket.**

56% vs. 26% amongst general population.

**Highest claimed knowledge of how data is used and collected about them in their day-to-day lives.**

73% know at least a fair amount about this (vs. 45% amongst general population).

**Most likely to be in a role which has data and technology as its main focus.**

18% vs. 8% amongst general population.

**Most comfortable providing their personal data and much more trusting that it will be used**

**ethically, irrespective of the use-case and the organisation using it.**

**Most optimistic about data and technology.**

Agree that ‘collecting and analysing data is good for society’ and ‘all groups in society benefit from data use equally’ more than any other segment. Most likely to say that the technology sector has made life better for people like them (81% vs. 57% amongst general population).

### Data awareness

73%

know at least a fair amount about how data is used and collected about them in their day-to-day lives

### Trust of organisations

(Central) Government

Social media companies

Big technology companies

To act in your best interests (NET: Trust)

63%

63%

87%



# Segment 3 – ‘Enthusiastic Tech Pros’

18% of weighted online sample

## Demographic profile

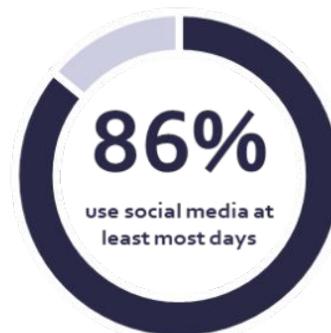


This segment is **predominantly male**, with the proportion identifying as male being significantly higher than the proportion identifying as female (53% vs. 46%).

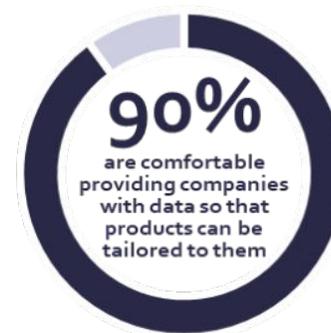
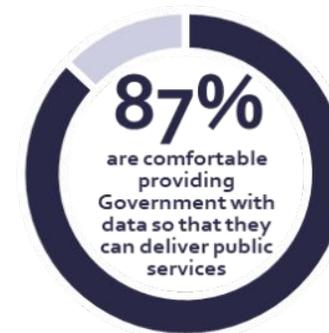


Aside from the very young Segment 4, this segment has the **lowest mean age**, at 42 (overall mean age 49).

## Digital usage and confidence



## Comfort with data collection





# Segment 4 – ‘Optimistic Tech Dabbler’

12% of weighted online sample

## Key attitudes and behaviours

### Lowest claimed internet usage per month.

Just 64% claim to use the internet at least most days (vs. 93% general population).  
Less likely to use search engines than all segments except Segment 1 (48% vs. 70% amongst general population).

### Lowest confidence using online devices.

75% vs. 91% amongst general population are confident using computers, smartphones, or other electronic devices to do the things they want to online.

### More likely to say that they don’t understand what is meant by the various terms tested.

The dummy term ‘neural turbines’ is a notable exception, where claimed understanding is significantly *higher* than all other segments.

### The perceived priorities of this segment are notably different from all others.

They are more likely to identify tax, family life, and transport as some of the most important issues facing the country and also areas where data presents opportunities to bring benefits.





## Segment 4 – ‘Optimistic Tech Dabbler’

12% of weighted online sample

### Demographic profile



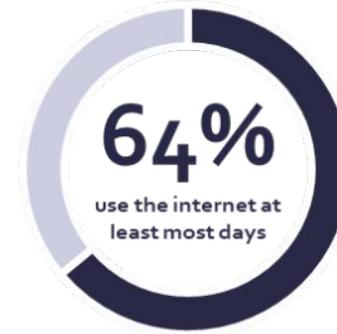
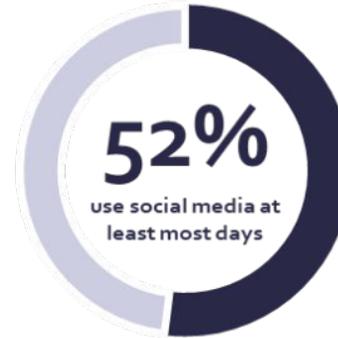
This segment has a significantly **lower mean age** than all other segments at 35 years old (overall mean age is 49).

This segment is significantly more likely to be **Asian** (14%) than all other segments and is significantly less likely to be white than all other segments (74%).

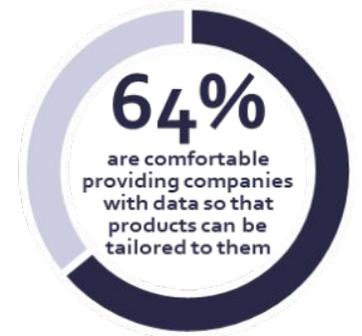
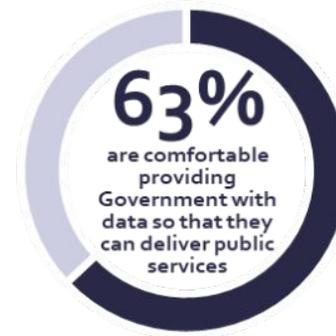


Members of this segment are significantly more likely to be from **London** than any other segment (23%).

### Digital usage and confidence



### Comfort with data collection





## Segment 5 – ‘Unsure Digital Users’

7% of weighted online sample

### Key attitudes and behaviours

**Poor knowledge about how data about them is used and collected in their day-to-day lives.**

Just 1 in 5 (20%) say that they know at least a fair amount. This is significantly lower than all segments except for Segment 1.

**Claimed understanding of tested terminology is relatively poor.**

More likely to say they do not understand what is meant by ‘algorithms’ (70% vs. 37% general population), and ‘targeted advertising’ (49% vs. 18% general population).

**The risks and opportunities presented by data use are unclear to this segment.**

Much more likely to say they ‘don’t know’ the area in which there is the greatest opportunity for data use to benefit society (45% vs. 13% amongst general population), or ‘don’t know’ what the biggest risk of data use is (47% vs. 9% amongst general population).

**Relatively low level of familiarity with Artificial Intelligence.**

Least likely segment to have heard of AI (71% vs 90% amongst general population), and are less likely to be able to explain it at least partially (34% vs 63% general population).

### Data awareness

20%

know at least a fair amount about how data is used and collected about them in their day-to-day lives

### Trust of organisations

(Central) Government

Social media companies

Big technology companies

To act in your best interests  
(NET: Trust)

24%

30%

42%



## Segment 5 – ‘Unsure Digital Users’

7% of weighted online sample

### Demographic profile



This segment is predominantly **female**, with the proportion identifying as female being significantly higher than the proportion identifying as male (58% vs. 41%).

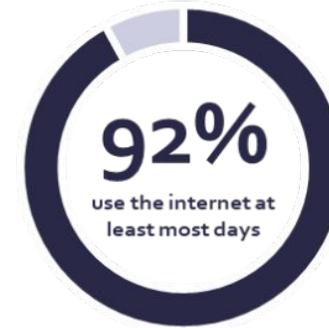
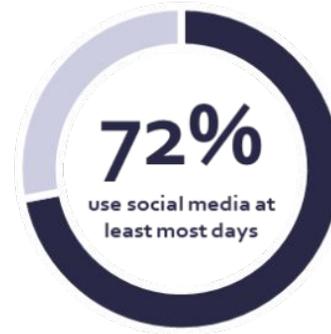


Members of this segment are significantly more likely to be in the **D or E socio-economic group** than any other segment (46%).

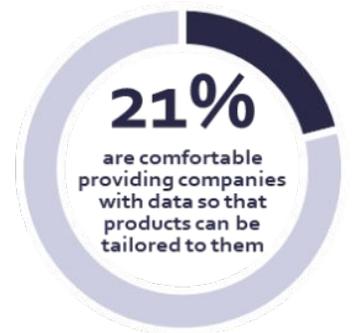
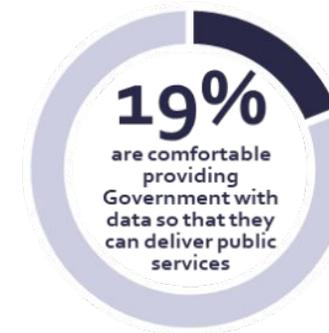


Those in this segment are significantly more likely to be from **Scotland** than almost any other segment (12%).

### Digital usage and confidence



### Comfort with data collection



Note: There was a relatively high proportion of “Don’t know” responses for these questions amongst this segment.



# Segment 6 – ‘Distrusting Data Sceptics’

14% of weighted online sample

## Key attitudes and behaviours

**Generally less trusting that organisations will act in their best interests than all other segments.**

The organisation they trust least to act in their best interests is the Government: 91% (vs. 59% general population).

**In all scenarios tested, they were more uncomfortable providing personal data and with AI-supported decision making.**

**Most likely to identify ‘data being used for surveillance’ as the single greatest risk of data use.**

**More likely than other segments to say the technology sector has made life worse for people like them. 31% vs. 11% amongst general population.**

**More pessimistic about data and technology than other segments.**

More likely to disagree that: data is useful for creating products and services that benefit them (39% vs 14% amongst general population); collecting and analysing data is good for society (54% vs 20% amongst general population); all groups in society benefit equally from data use (59% vs. 31% amongst general population).

### Data awareness

**40%**

know at least a fair amount about how data is used and collected about them in their day-to-day lives

### Trust of organisations

	(Central) Government	Social media companies	Big technology companies
--	----------------------	------------------------	--------------------------

To act in your best interests (NET: Trust)

**9%**

**11%**

**29%**



# Segment 6 – ‘Distrusting Data Sceptics’

14% of weighted online sample

## Demographic profile

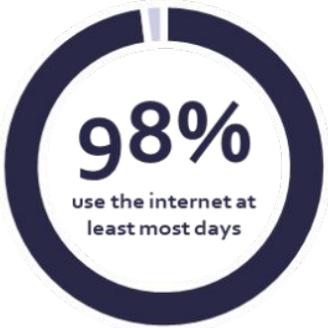


This segment is **predominantly male**, with the proportion identifying as male being significantly higher than the proportion identifying as female (54% vs. 44%).

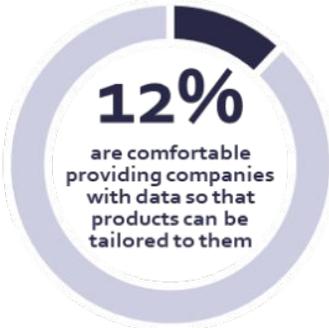
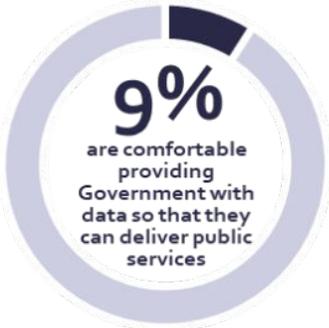


Aside from the very old Segment 1, this segment has a significantly **higher mean age** than all other segments, at 52 (overall mean age 49).

## Digital usage and confidence



## Comfort with data collection



# Annex B - Details on methodology



# Sample Profile (weighted to be nationally representative) – online survey

Total online sample size

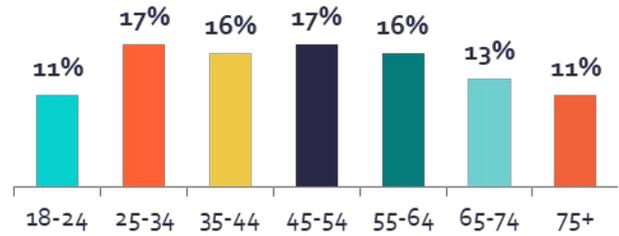
**4,257**  
UK adults

## Gender

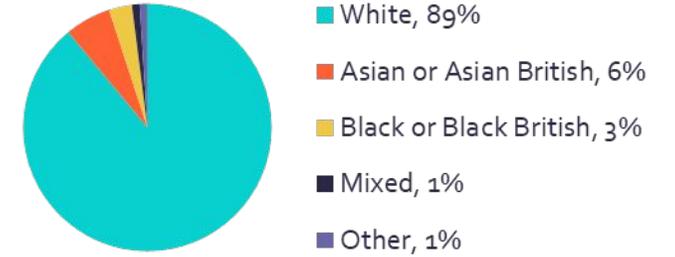


Male 49%  
Female 51%

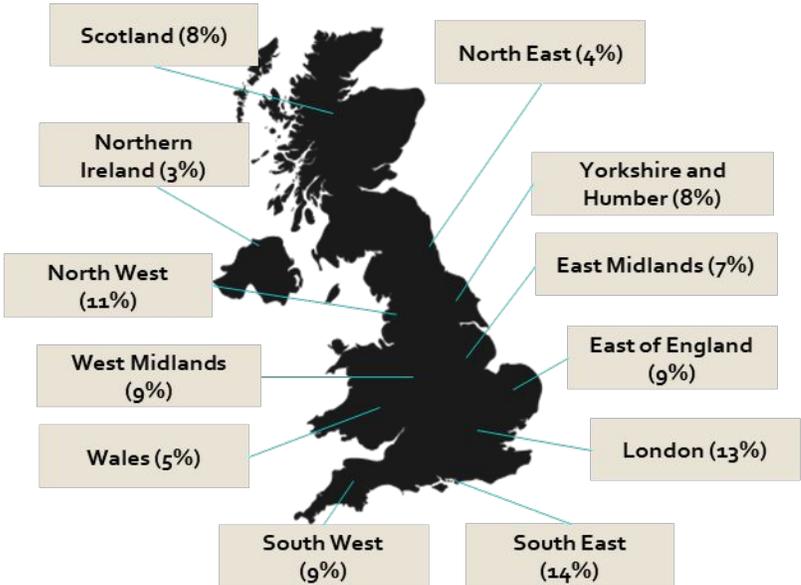
## Age



## Ethnicity



## Regions



## Socio-Economic Group

AB	27%
C1	28%
C2	20%
DE	24%

Socio-economic group is a standard definition, based on the chief income earner's economic status. See the Annex for definitions.

## Digital familiarity\*

A proxy measure for digital familiarity is used throughout, to compare responses to other questions against this. This measure is based on responses to Q4 (frequency of various online activities) and Q5 (confidence using devices online). Respondents were given a composite score out of 18 and arranged into the following bands:

Low digital familiarity (Score of 0-10/18)	35%
Medium digital familiarity (Score of 11-14/18)	40%
High digital familiarity (Score of 15-18/18)	26%

'Very low digital familiarity' is used throughout to refer to the 'digitally excluded' sample (telephone booster).

# Sample Profile – ‘High digital use/confidence’

Total ‘High digital use/confidence’ sample size

**1,084**  
UK adults

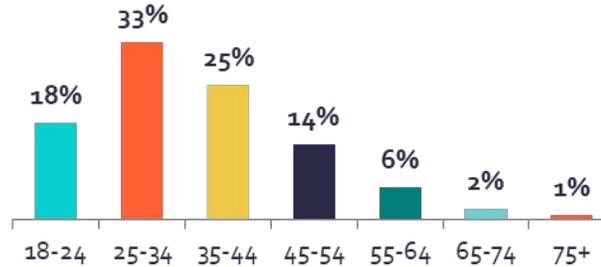
## Gender



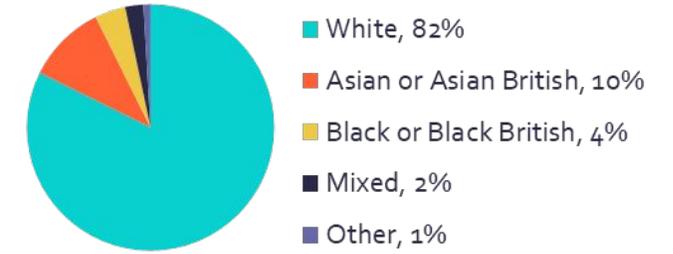
Male  
39%

Female  
61%

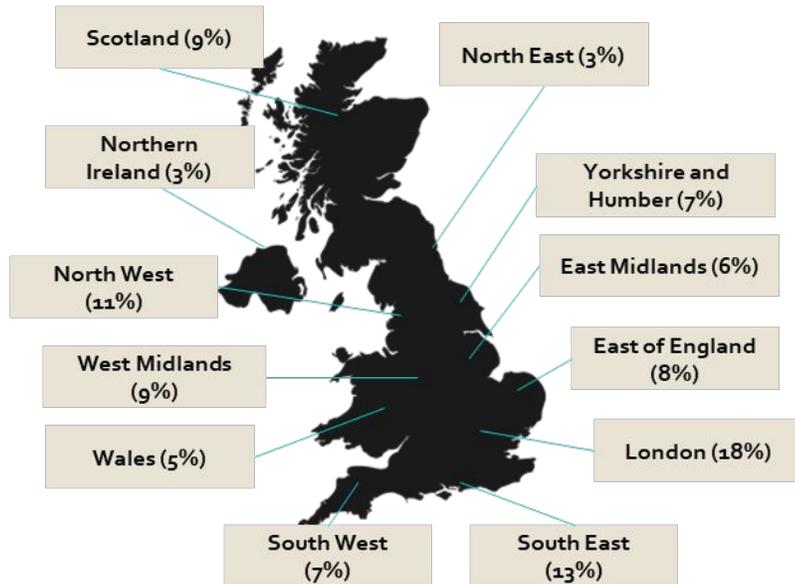
## Age



## Ethnicity



## Regions



## Socio-Economic Group

AB	33%
C1	26%
C2	20%
DE	21%

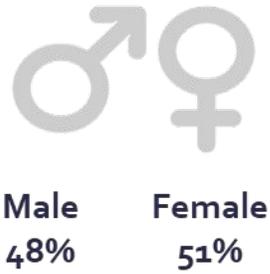
Socio-economic group is a standard definition, based on the chief income earner’s economic status. See the Annex for definitions.

# Sample Profile – ‘Medium digital use/confidence’

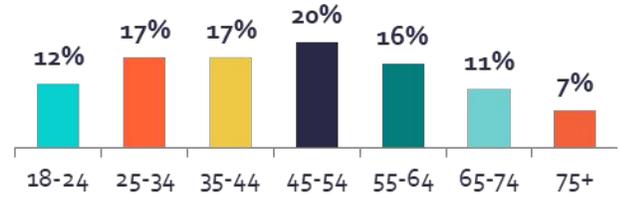
Total ‘Medium digital use/confidence’ sample size

**1,697**  
UK adults

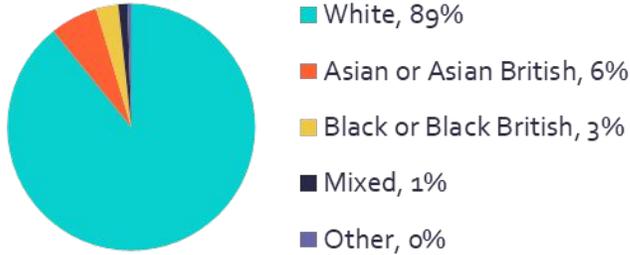
## Gender



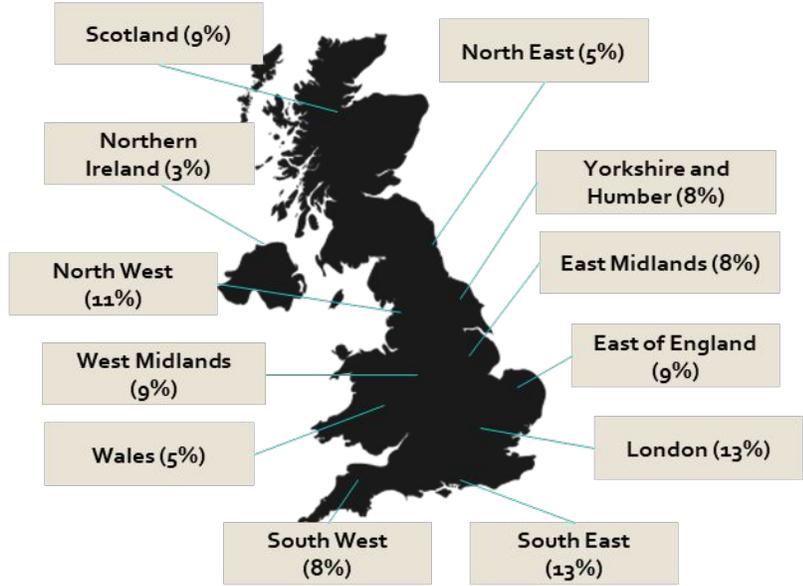
## Age



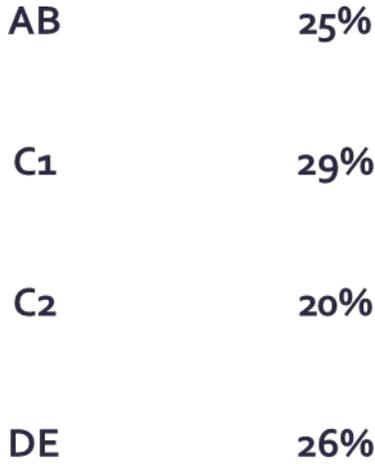
## Ethnicity



## Regions



## Socio-Economic Group

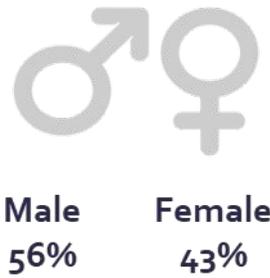


Socio-economic group is a standard definition, based on the chief income earner’s economic status. See the Annex for definitions.

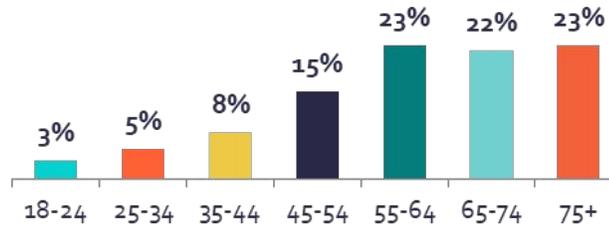
# Sample Profile – ‘Low digital use/confidence’

Total ‘Low digital use/confidence’ sample size  
1,476  
UK adults

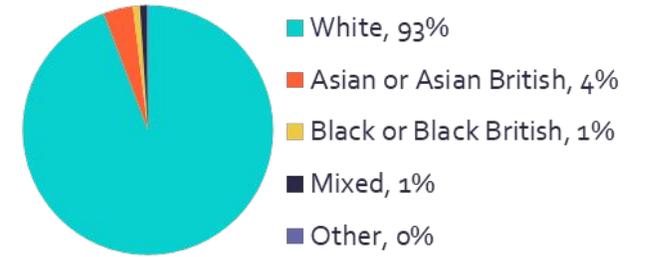
## Gender



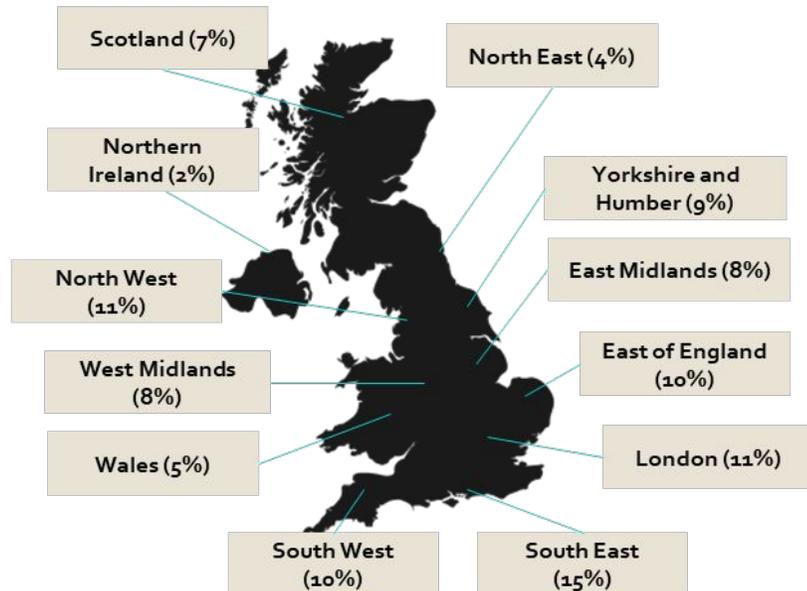
## Age



## Ethnicity



## Regions



## Socio-Economic Group

AB	26%
C1	28%
C2	20%
DE	26%

Socio-economic group is a standard definition, based on the chief income earner’s economic status. See the Annex for definitions.

# Sample Profile – ‘Digitally excluded’ telephone booster

Total ‘digitally excluded’ sample size

# 200

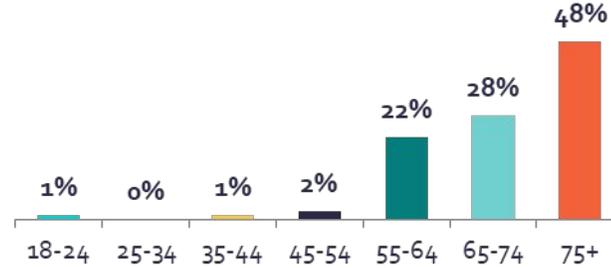
UK adults

## Gender

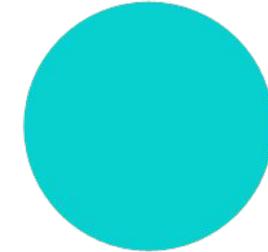


Male 36%  
Female 63%

## Age

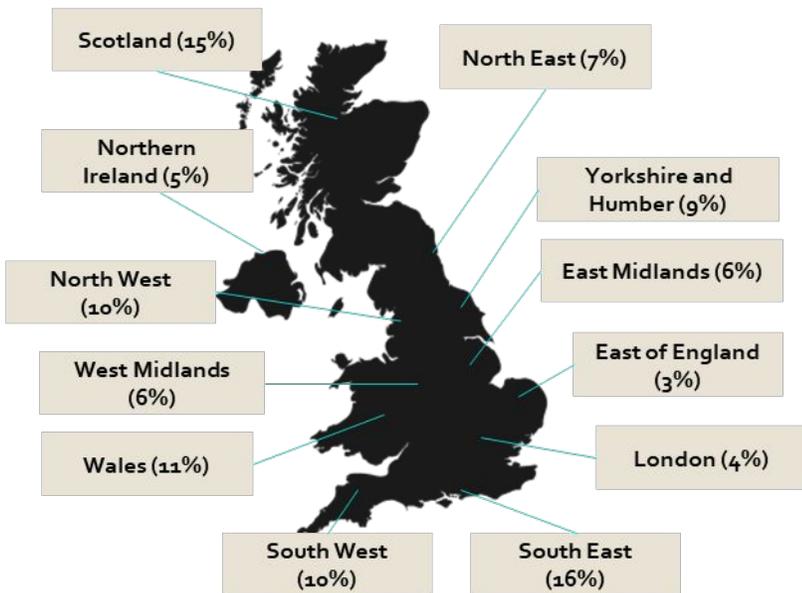


## Ethnicity



- White, 100%
- Asian or Asian British, 0%
- Black or Black British, 0%
- Mixed, 0%
- Other, 0%

## Regions



## Socio-Economic Group

AB	27%
C1	22%
C2	23%
DE	29%

Socio-economic group is a standard definition, based on the chief income earner’s economic status. See the Annex for definitions.

# Methodology – Statistical significance

- Findings marked with   arrows are statistically significant, with a 95% confidence level, when compared to *all* other members of that particular subgroup. For instance, if there is an  arrow attached to the results for a particular question for those identifying as male, this means that this figure is higher than the one for females *and* that there is a 95% this difference would persist if we repeated the fieldwork under the same conditions. Where 'significant' is used throughout this report, it is specifically statistical significance that is meant.
- Two different methodologies are used to calculate statistical significance throughout this report: **two-tailed independent z-test tests** to compare proportions/percentages, and **two-tailed independent t-tests** to compare mean scores. This involves calculating a z-score or t-score in each case, using the following formulae:

The formula for a **z-score** is:

$$\frac{p1 - p2}{\sqrt{\left(\frac{(x1 + x2)}{(n1 + n2)} \times \left(1 - \frac{(x1 + x2)}{(n1 + n2)}\right)\right) \times \left(\frac{1}{n1} + \frac{1}{n2}\right)}}$$

where:

x1 and x2 are the two raw numbers in the cells being compared  
 n1 and n2 are the two column totals (bases) of the cells being compared  
 p1 and p2 are the two ratios (p1 = x1/n1)

The formula for a **t-score** is:

$$\frac{x1 - x2}{\sqrt{\frac{(s1^2 \times (n1 - 1)) + (s2^2 \times (n2 - 1))}{n1 + n2 - 2} \times \left(\frac{1}{n1} + \frac{1}{n2}\right)}}$$

where:

x1 and x2 are the two mean scores being compared  
 n1 and n2 are the two bases  
 s1 and s2 are the two standard deviations

- Having calculated these scores, significance is then calculated based upon the significance level which is being tested, and the degree of freedom, which is equal to n1 + n2 – 2. Some common examples are shown below. If the score is greater than the figure shown in the relevant cell, the difference is significant.

Degrees of freedom:	80	120	200	10000
For 95% significance level:	1.990	1.980	1.972	1.960

# Comparing sample profile by ethnicity (1) – online survey

	Number of respondents (unweighted)	Average age	Gender	Socio-economic group
White	3,643	51 ↑	Female: 50% Male: 49% ↑	Higher (ABC1): 55% Lower (C2DE): 45% ↑
Black	132	34	Female: 58% Male: 40%	Higher (ABC1): 67% ↑ Lower (C2DE): 33%
Asian	271	34	Female: 54% Male: 45%	Higher (ABC1): 60% Lower (C2DE): 40%



Please note that the statistical significance arrows are used differently here to elsewhere in the report, in order to facilitate comparison between all three groups.

Arrows indicate statistically significant differences from other ethnic groups. The colour of the arrow indicates which group the figure is significantly higher than. For instance, an green arrow (↑) indicates that the figure is significantly higher than the figure amongst those with an Asian ethnicity. A white arrow (↑) indicates that the figure significantly differs from all of the other groups shown.

# Comparing sample profile by ethnicity (2) – online survey

	Total	White	Black	Asian
South-East	14%	14% ↑	6%	9%
London	13%	10%	58% ↑	36% ↑
North-West	11%	11%	8%	9%
West Midlands	9%	8%	9%	17% ↑
South-West	9%	9% ↑	4%	3%
East of England	9%	10% ↑	2%	3%
Scotland	8%	9% ↑	1%	3%
Yorkshire & Humberside	8%	8%	5%	12% ↑
East Midlands	7%	8%	4%	5%
Wales	5%	5% ↑	1%	1%
North-East	4%	4% ↑	1%	1%
Northern Ireland	3%	3% ↑	0%	1%

Please note that the statistical significance arrows are used differently here to elsewhere in the report, in order to facilitate comparison between all three groups.



Arrows indicate statistically significant differences from other ethnic groups. The colour of the arrow indicates which group the figure is significantly higher than. For instance, an green arrow (↑) indicates that the figure is significantly higher than the figure amongst those with an Asian ethnicity. A white arrow (↑) indicates that the figure significantly differs from all of the other groups shown.

## Annex 5: Socio-economic grades and definitions

The standard socio-economic grades used in this report are defined as follows:

AB	Higher and intermediate managerial, administrative, professional occupations
C1	Supervisory, clerical and junior managerial, administrative, professional occupations
C2	Skilled manual occupations
DE	Semi-skilled and unskilled manual occupations, unemployed and lowest grade occupations

## Annex 1: Survey design for conjoint study (2)

The conjoint design featured the generation of random pairs (ensuring that the two options in a given pair are not identical), each on a new screen. The data collected from each individual respondent included all of the variables shown for each option in each pair, and which of the options is chosen in each case.

*Respondents were given the below instructions:*

You are about to be shown some options that represent how different kinds of data might be gathered, stored, managed, and used. The options you see will be randomly generated, and your task will be to choose your preferred option among six pairs in total.

Please note that because these options are randomly generated, some of them do not currently happen. In reality you might be willing to share your data for both options or neither, we just want to hear, in theory, for which option you would be more willing to share your data.

## Annex 2: Conjoint variables and exclusions

Actor (To be piped into first row)	Data Type (To be piped into 'To use...' row)	Identifiability (To be piped into 'That...' row)	Purpose (To be piped into 'In order to...' row)	Rules (To be piped into 'If...' row)
The NHS	Sensitive information about you (e.g. income, sexual orientation, ethnicity)	Personally identifies you	Personalise services to you (exclude averaged data) 3.2	There are strict rules in place which require organisations to clearly explain how they will use people's data and keep this data safe and secure
The police	Basic information about you (e.g. gender, age)	Has been averaged across the whole population so you can't be personally identified	Reduce their costs	There are some rules in place which require organisations to explain how they will use people's data and keep this data safe and secure
Banks	Location data from your mobile phone about where you go		Improve Government policies	There are no rules in place which require organisations to explain how they will use people's data and keep this data safe and secure
Private companies	Your online activity (e.g. which websites you have visited or products you have bought)		Create new products and services	
Academic researchers at universities	Your health record		Research new health treatments (exclude banks, police) 1.3 1.2	
The Government			Target adverts at you (exclude averaged data) 3.2	
Big tech companies (e.g. Amazon or Google)			Gain insights about people's behaviour and attitudes	
Charities			Identify crime (exclude NHS) 1.1	
Local councils			Make products and services quicker and easier for users	
Social media companies (e.g. Facebook, Twitter, TikTok)			To help vulnerable people	