Britainthinks

Insight & Strategy

CDEI | Trust in Data

Detailed Report

14.09.21

Contents

2 Objectives and methodology

Background and objectives

This research was conducted by **BritainThinks** on behalf of the **Centre for Data Ethics and Innovation** between 2nd August and 3rd September 2021.

The CDEI, as part of the National Data Strategy, want to improve public awareness of the societal benefits of responsible data use. This research aims to build on existing evidence and provide a 'pulse check' on public attitudes to data, exploring which topics and issues are most resonant, and in what contexts.

The research objectives were therefore:

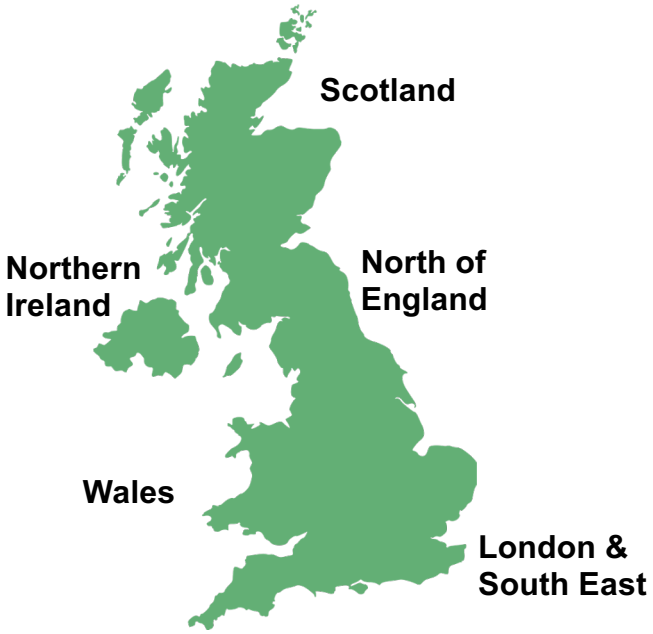
- To understand how the public spontaneously view the use of data, particularly by the public sector, for different purposes and in different settings
- To understand the characteristics of data use, or it's description, which are viewed most positively and negatively by the public
- To understand how attitudes to, and opinions on, data use are formed

Methodology

	Phase 1	Phase 2	Phase 3
When:			
What:	**What are people's spontaneous attitudes to data and data use in general?**	**What are people's attitudes to data use for different purposes, by different organisations and using different technologies?**	**What elements of different data use cases are most influential in public attitudes?**
How:	 **Online focus groups** 5 x 90 minute groups, with 5-6 participants each	 **Online task** Presenting information on 12 case studies and gathering reactions	 **Online focus groups** 5 x 90 minute groups, with 5-6 participants each
Who:	**27 members of the public completed all three phases**		

Sample

Regions represented



Participants met in groups of 6, split by age, known to be a key determinant of attitudes to technology:

Aged 18-29	Aged 30-39	Aged 40-49	Aged 50-59	Aged 60+

Each group contained participants from at least 4 different regions. Within each region, we ensured a representative mix of demographics and representation of diverse attitudes.

Demographics

- An equal balance of:
  - Gender (Male/Female)
  - Socio-economic group (ABC1/C2DE)
- Mix of urban, sub-urban and rural locations
- Ethnic minority audience to broadly reflect local demographics

Attitudes

- Spread of:
  - Trust in institutions (government, media, experts)
  - Optimism towards data use
  - Levels of comfort with using technology

2 Key findings

Key insights

1. **For most people the strongest spontaneous association with data is 'personal data'.** This means that many think of data primarily in terms of privacy and security, rather than an ability to facilitate large scale change.

2. **People often have negative views on data use because the bad examples are more memorable.** While data plays an important and positive role in their lives, this is invisible, and they only 'notice' data when something goes wrong, in the context of their own personal interactions with data or in the news.

3. **When prompted, people do recognise the potential for data use to facilitate large scale positive change but also have spontaneous concerns.** They ask questions about data security, privacy and whether specific uses go beyond the consent that has been given.

Key insights

4. **Despite concerns about *how* data is used in principle, when it comes to specific use cases the outcome or *why* has a much stronger influence on attitudes.** Where people saw significant societal benefits their concerns became less prominent.

3.1 What do you think when you hear the term ‘data’?

Spontaneous impressions and attitudes to data use

Data use is a low salience issue for the public and their knowledge is consequently low

- Participants accepted data use as a normal part of modern life.
- They assumed it must deliver some benefits, but were often unclear about what these are.
- Most people were aware that they are constantly producing and giving access to a huge amount of data about themselves, but have little knowledge of the extent, how data is stored or ultimately how it is used.
- In line with low knowledge**, perceptions of data use in general were fragile and were easily shifted** by additional information or a different context.

I do not know the full extent on what data means. It is information that is gathered, extracted and mined for their [companies] own purposes … I just know every time I swipe my credit card at the supermarket, they know something about me. That is the new world we live in.

(60+, Phase 1 Group)

[At work] we have a whole section on data protection, on protecting people's information and storing it. So I think I have an idea of what it is, but sometimes you don't know. We've become so naïve that we don't sometimes think what we're sharing when we're taking part online.

(18-29, Phase 1 Group)

Participants tended to interpret ‘data’ as ‘personal data’ rather than aggregated or anonymised data

The strongest association participants had with data is *their own* personal data and they struggle to think beyond the limitations of this, creating three big problems in terms of championing data use.

1

Limited understanding of aggregated data

By only considering data in terms of personally identifiable information, it **becomes difficult to spontaneously picture the positive implications** of pooling data into larger sets.

2

Fears over loss of privacy

Participants **struggled to articulate the downsides** of data collection and use, often **defaulting to concerns about privacy**.

As such, they are wary of their personal data being passed on and used for reasons they believe they did not consent to.

3

Vector for pre-existing views

The lack of concrete knowledge and well formed attitudes about data use in the abstract means it’s **easy for people to project other views onto it.**

Conversations around data therefore centred on perceptions of the future. Where sceptical participants expect it will mainly be used to **diminish their privacy and rights**, for more optimistic participants data use is **associated with progress**.

Views of data are often informed by the times when things 'go wrong' leading to negative attitudes

As successful data use tends to facilitate smoother and more seamless experiences for the public, participants often don't notice it being used unless something goes wrong. These instances tend to focus on either:

Personal experience

Participants draw on day-to-day interactions to shape their view of data. Negative examples, where people feel data they have given is used inappropriately, are more top-of-mind than positive examples. Negative examples that came up spontaneously include:

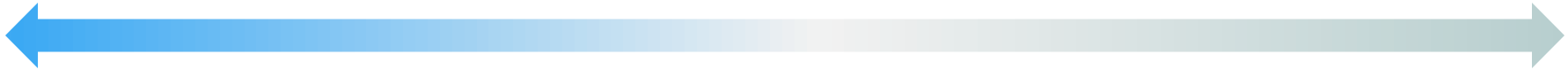
- Targeted advertising
- Data trading
- Privacy policies & cookie requests

News

Participants often said they had heard negative stories about data on the news; almost exclusively relating to data 'scandals'. Many therefore felt there are frequent **data breaches/ leaks, instances of organisations using data to manipulate politics, or politicians manipulating data to justify policy.** However very few people could name specific instances. A minority of participants gave these examples:

- NHS data sharing scheme
- Cambridge Analytica
- Ofqual's A-Level algorithm

Spontaneous associations with data were influenced by age more significantly than any other demographic



	Younger	Older
Different starting points	Digital natives who tend to take the benefits created by data for granted. Consequently, they are less likely to interrogate collection when consenting to share their data for services.	Stronger memories of a time before widespread digital technology use mean greater appreciation for the convenience they offer seen in the majority of older respondents.
Awareness & associations	Somewhat more familiar with the mechanics of data technology, and their links to specific services. For example, some spontaneously associated 'algorithms' with social media and had concerns about their use.	Lower awareness of data technology led to more neutral associations (e.g. algorithms & maths). Spontaneous associations revolved around more tangible examples, such as automated bank phone systems.
Fears & scepticism	More likely to raise concerns about the potential for systemic discrimination caused by uses of data (e.g. banks and the welfare system).	Fewer concerns about discrimination, with some seeing computer decision making as better in this respect. Concerns about privacy and scams/fraud much more pronounced, alongside fears of losing 'human touch' in services, or losing jobs to automation.

These differences were more pronounced because participants met in groups based on age. In the real world, people are less likely to encounter data in such closed groups, and other contextual factors may become more important.

Britainthinks
Insight & Strategy

3.2 What are the benefits and risks of data use?

Perceptions of the potential benefits and drawbacks of data use in general

When pushed to consider positive data use cases, people do begin to appreciate the possibilities it creates for a better future

Data can deliver direct, non-trivial personal benefits

Data, when aggregated, could be used to bring positive benefits on a societal level

These direct benefits include:

- Healthcare; specifically identifying illness earlier
- Security (e.g. banking), aiding police catching criminals
- Data linking raised by middle aged groups (e.g. free school meals).

Data-based systems can be better or fairer at some tasks then humans

Data could be used to bring an element of neutrality to decision making processes

- Some feel this could lead to fairer decision making if it is able to take personal biases out of the equation; e.g. applying for a loan or when interacting with police
- Less likely to make mistakes, removing human error to make more accurate decisions

Data use can bring about increased daily convenience

Data, information and other technology has already made life easier; and this could go further

- Many appreciated the efficiencies data has already brought (e.g. Google Maps), and are aware that other, day-to-day advancements are grounded in data usage
- This is especially true for older audiences, who value the convenience of modern life

Data can drive smarter decisions on a smaller scale

Data could also be used to save time spotting trends and highlighting problems

- Many intuitively understood that computer analysis can uncover things missed by humans
- They viewed this as particularly important for subjects closely related to science (e.g. medical trials)
- And also saw the benefit on a personal level, such as for SMEs

Data is instinctively associated with 'progress', 'efficiency', and 'future'

Top-of-mind associations are future facing, but don't always skew positively

- Participants who envisage life improving, becoming more open and streamlined in the process, understand that data will lie at the heart of this
- However some were nervous about the idea of progress for the sake of progress

But these rarely surfaced spontaneously, and tended to be brought up in the context of specific applications

Despite acknowledging the benefits of data use, participants were concerned about the risks and downsides

Data Security

Storage
Where is my data kept?

Safeguards
How is my data kept safe and secure?

Trust
Do I trust the people using data to have good practices?

Provenance
Where did it come from? Was it stolen?

Data Privacy

Who
Did I give consent to the people using it?

My identity
What does it reveal about me as a person?

Consent
Did I explicitly give consent for it to be collected?

Protections
How is my identity protected?

Data Usage

Why
Who is really gaining from my data being used?

Consideration
How carefully is my data being used?

Manipulation
Is my data being used to influence and pressure me?

Limits
Is it being used for reasons I didn't consent to?

Replaced
Will people be supplanted because of data?

Lose out
Will decisions go against me because of data?

Concerns and questions raised by participants spontaneously

3.3 What data terms are familiar, & what are the key associations?

Deep dive into spontaneous associations with data use in the public and private sector, and key data technologies

Spontaneous associations with data reflected concerns that data is currently gathered primarily to enable personalised targeting

Spontaneous associations with 'data'*

- Spontaneous associations with 'data' tend towards suspicion; they revolved around it being personal, identifying information that is thought be passed around or used in an frustrating ways (e.g. social media and call centre targeting).
- There were also more neutral associations that focussed on the more technical and generic uses of data.
- Top-of-mind associations also included some concerns founded on popular culture (e.g. "Big Brother") or media attention (particularly data breaches).

**Size of word/phrase reflects frequency of mention when participants were asked to give their top three associations with 'data'*

Reactions to public sector data use depended which public sector organisation people were thinking of at the time

Spontaneous associations with 'public sector data use'*

**Size of word/phrase reflects frequency of mention when participants were asked to give their top three associations with 'public sector use of data'*

- Spontaneous understanding focused on familiar data uses, 'government', 'medical' & 'education' records. Beyond that, reactions were based on pre-existing opinions of the organisation involved.
- With regards to 'government' and politicians, spontaneous reactions revolved around this being 'untrustworthy', allowing 'control' or being 'suspicious'.
- Public services caused some polarisation. Negative reactions revolved around fears data could be lost, or impacted by corruption. In contrast more positive associations centred around data being 'useful' and 'handy' in making public sector services more efficient and convenient.

In contrast, all private sector organisations were mistrusted and assumed to be seeking maximum profits at any cost

Spontaneous associations with 'private sector use of data'*

- Spontaneous associations focused very heavily on 'targeted marketing', largely perceived very negatively as the majority of participants felt they are harassed and bombarded on a day-to-day basis.
- Beyond this, participants either thought about data theft or data being used (or misused) for private profit; some participants even considered this profit to be made directly at the expense of the consumer.

**Size of word/phrase reflects frequency of mention when participants were asked to give their top three associations with 'private sector use of data'*

Algorithms were viewed relatively neutrally, and some participants had an awareness of what they are and how they function

Spontaneous associations with 'algorithms'*

- The most common associations with algorithms tended to be quite neutral and descriptive. Participants mentioned that they were 'complex' sequences' that use data and are found in computers, financial markets, and healthcare.
- Some participants did link algorithms to the collection of data, suggesting this could be used as the basis for targeted advertising.
- Algorithms did not feel new to participants, many of whom had long-standing associations based on studying mathematics at school.

**Size of word/phrase reflects frequency of mention when participants were asked to give their top three associations with 'algorithms'*

Much less was known about machine learning and AI, and many participants feared their potential impact in the future

Spontaneous associations with 'AI and machine learning'*

**Size of word/phrase reflects frequency of mention when participants were asked to give their top three associations with 'AI and machine learning'*

- AI and machine learning has strong future connotations, and was frequently thought of in terms of advancement and robots (often based on popular culture references).
- At the core of this were various concerns; especially around loss of jobs, empathy and a human touch. Broadly, as concepts they feel like big unknowns, which can be very scary to some.
- There were also positive spontaneous associations, although these were less common; some participants described efficiencies that can be created through using these faster and helpful technologies.

Some participants understood the potential efficiencies of automated decision systems, but most were unsure

Spontaneous associations with 'automated decision making systems'*

- Most participants had very little knowledge about what an automated decision is and what it can do in practice.
- A minority were able to mention specific associations with services, like, banking and telephone helplines and/or surveys.
- Others did speak more positively about the ability to find efficiencies with faster, more helpful methods, however this was rarely founded on anything other than a vague awareness or understanding.

**Size of word/phrase reflects frequency of mention when participants were asked to give their top three associations with 'automated decision making systems'*

4.1 Under what circumstances do people trust data use?

Analysis of participant responses to different data use cases

Participants were presented with a series of case studies of different uses of data with four categories of information

Why is this data being used?

Who is using this data?

What data is being used?

How is this data being used?

Of the four factors 'why' and 'who' were most important to our participants

More important

Why is this data being used?

Participants were most concerned with the outcomes that data use would deliver, with a strong 'why' driving trust, even in use cases where the 'how' is a concern.

Who is using this data?

The level of trust in 'who' is using data affects trust in whether the 'why' will actually happen.

What data is being used?

How is this data being used?

'What' and 'how' were seen as primarily technical questions. If the 'why' was seen to justify the use of data these were often moot points.

Less important

When judging the 'why' participants favoured big, societal benefits and cases where data is most obviously required

Substantial benefits

Big benefits (to society or the individual) are seen as 'worth' the potential downside, overriding concerns about privacy and security.

For example….

Smartphone company
Using an algorithm to scan all photos uploaded to the cloud, to detect & block child pornography, and report this to law enforcement

Though the scanning of photographs was seen as invasive, the benefit to society was seen as so positive that for most participants it was justifiable and welcome.

Importance of *their* data

Participants buy in to a case study when they can clearly see how 'their' data will contribute to the benefit.

For example….

NHS + AI company
Use breast health records and mammograms to train an AI to spot the signs of breast cancer earlier

Participants intuitively understand and buy-in to how their mammograms and health records contribute to medical research and better outcomes

Gov department & regulator
Using smart meter data and algorithms for research and to inform climate change policy

In contrast, participants do not understand how their smart meter data will help to tackle climate change

Specific and clear use of data

Generally, the clearer and more specific the collection of data, the use of data, and the benefit of a case study, the more positive participants were about it.

For example….

Car insurance company
Blackbox collecting drivers' speed, time and location to produce personalised car insurance premiums

Though most participants did not personally want this, they felt clear about the use of data and benefit and were therefore able to consider that others (or society as a whole) may benefit from it, leading to neutral instead of negative reactions

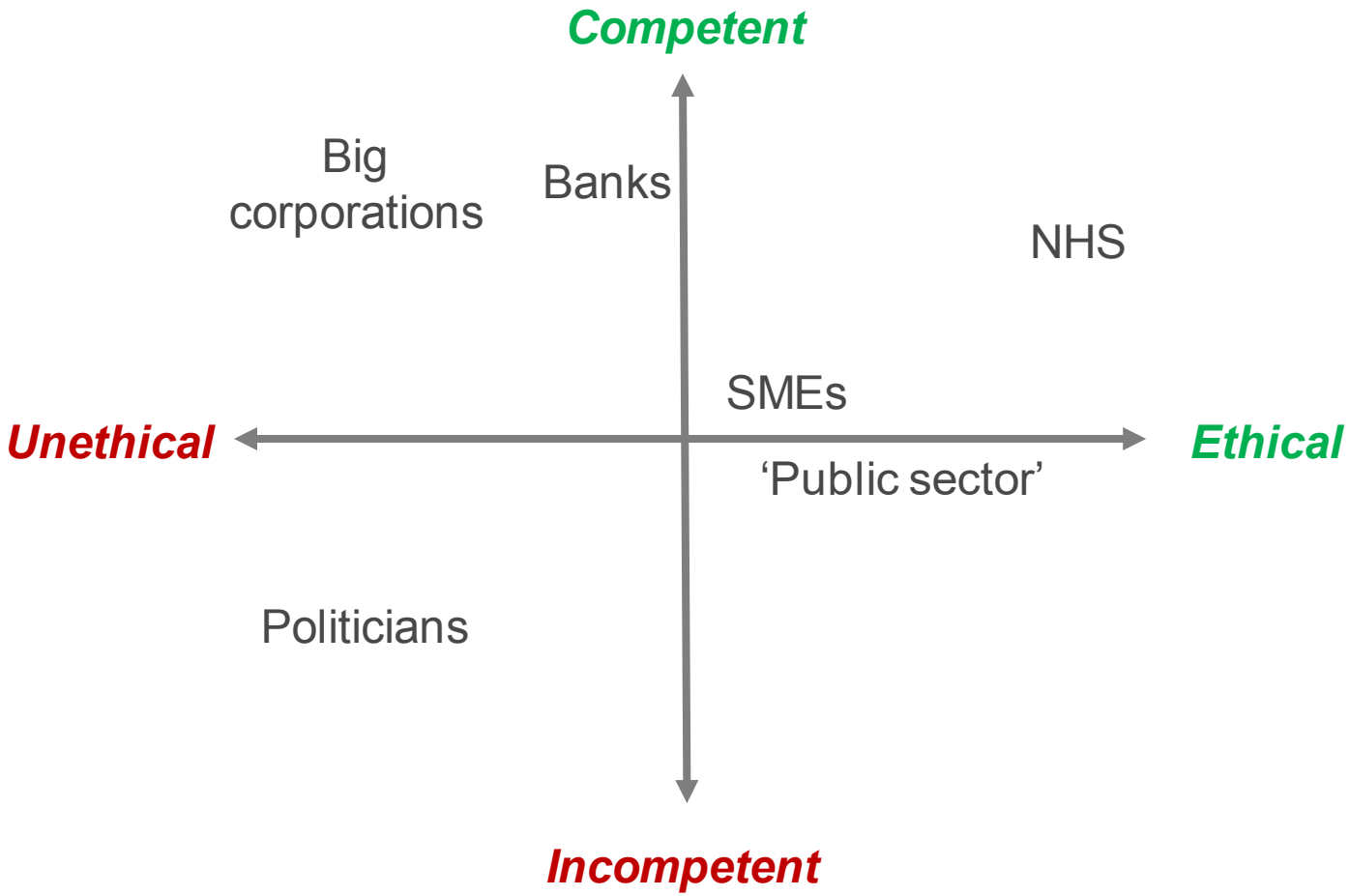
Trust in an organisation's data use was based on perceptions of their competence and their ethics

Perceptions of ethics were important for determining the credibility of the stated 'why', e.g.

- The NHS is trusted to use data for the reasons stated and not 'overextend' into other areas
- Whereas larger corporations are viewed with suspicion, with many assuming they will use data in ways they haven't explicitly declared

Perceptions of competence are important for determining confidence in data security

- The private sector is, generally, perceived to be more competent than the public sector – meaning they are trusted to 'keep data safe', even if they are using it ways the public aren't comfortable with
- Conversely, the public sector is less likely to be seen as competent, raising concerns about data security



What and *how* are viewed as technical questions. Though not directly driving attitudes towards the use of data, both can inform how participants evaluate benefits

What data is being used?

Beyond a general desire for data to be used ethically and with full consent, the type of data **does not directly drive attitudes to specific use cases.** Instead participants largely draw on 'why' and 'who' to make judgements on benefits, and therefore whether they approve of a particular use of data.

Nonetheless, the 'what' can **inform how participants evaluate benefits** through two ways**:**

- The more 'personal' data feels, the more clear, substantial and direct the benefit should be (to an individual or society) in order to 'justify' this use of data.
- The type of data may cast doubt on the credibility of the stated outcome, where it seems inconsistent. For example, if a system will use out-of-date data, incomplete, or seemingly irrelevant data, this can undermine credibility.

How is this data being used?

This is felt to be quite **technical** and **complex**, and seen as the 'nuts and bolts' for experts to engage in.

There is an underlying view that that the '**how' can produce desirable or undesirable outcomes**. In particular, they are seen as powerful tools (potentially producing good outcomes) though lacking a 'human touch' (potentially producing bad outcomes).

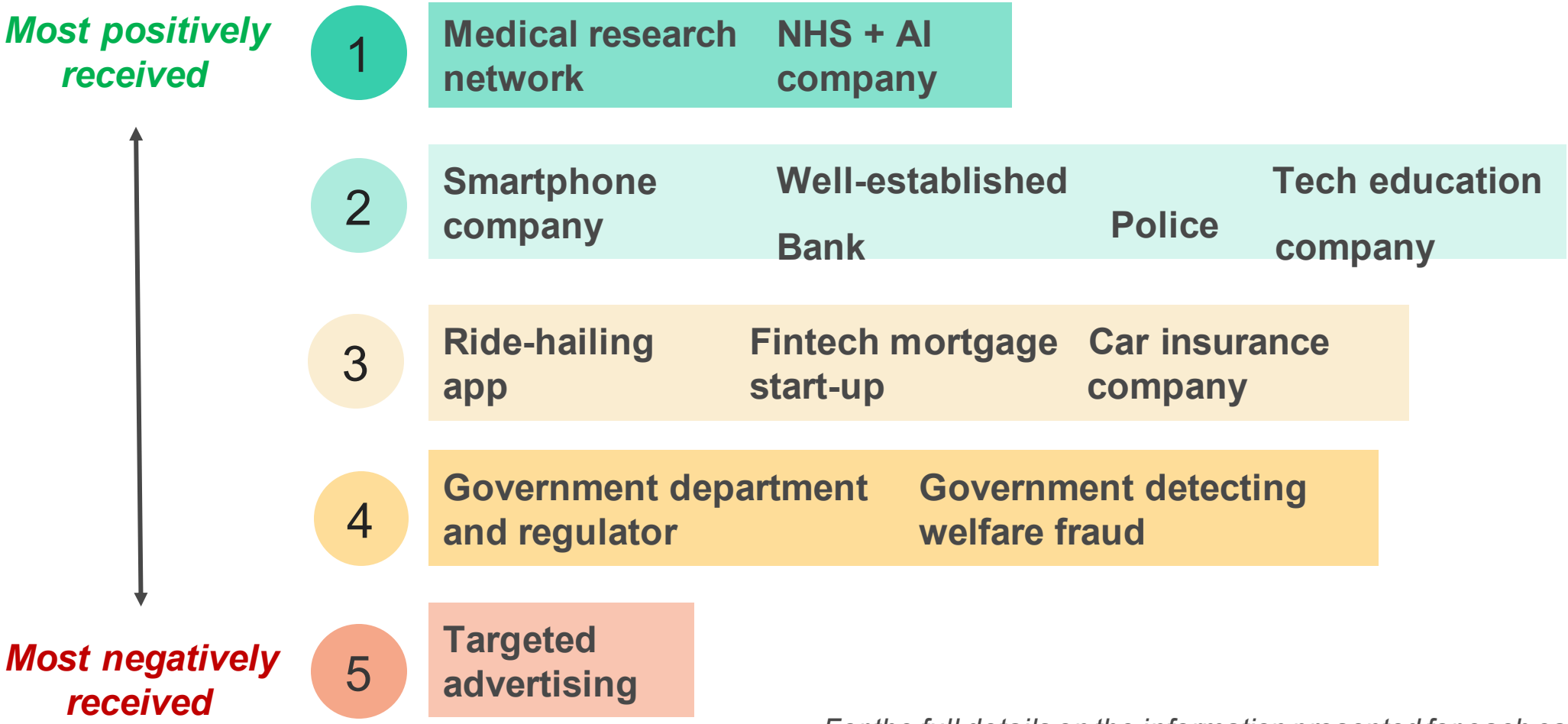
In consequence, while there is little engagement with the details of 'how', **these can still inform how participants evaluate benefits**. For example, certain processes (such as medical care) are seen as requiring more human attention to be done well, and therefore the 'how' will inform evaluations of benefits in these cases.

4.2 Which data use cases are seen most positively?

Deep dive into participant responses to 12 data use cases

We showed participants 12 case studies of data use, selected to represent a range of 'Who?', 'Why? 'How?' and 'What?'

Based on participant reactions, 5 tiers emerged…



For the full details on the information presented for each case study, see Appendix.

Use of data for medical research or diagnosis was almost universally strongly welcomed

1

Medical research network	NHS + AI company
A regional NHS medical research network, linking personal and health data about blood cancer patients to better understand people with blood cancer	NHS & an artificial intelligence company, using breast cancer patient health records and mammograms to train an artificial intelligence system to spot the signs of breast cancer earlier

It could be revolutionary in a very positive way. Endorphins and other things missing in the brain when you become poorly. AI analysing these chemicals… Would be fantastic.

(50-59, Phase 2 Group)

These two case studies combined a strong 'why' with a trusted 'who'.

Strong 'why': medical research or detection was seen as providing a **clear**, **substantial** and **direct** benefit.

- ***Clear*** because participants could easily understand the value of having a cancer detected earlier, or better treatment for blood cancer.
- ***Substantial*** because participants inferred the benefit would be 'saving lives'.
- ***Direct*** as participants intuitively understood the link between data, medical research and better medical outcomes.

Trusted 'who': The NHS is seen very positively, and trusted to use this data to advance these medical causes.

For the next tier, the ‘why’ was strong though caveated by practical doubts raised by the ‘who’ and ‘how’

2

Smartphone company	Well-established Bank
Famous consumer tech company using an algorithm to scan all photos uploaded to the cloud, in order to detect child pornography, block this content and report this use to law enforcement	Bank using photos of ID of customers and an AI to speed opening an account without risking increasing identity theft

Police	Tech education company
The Police using historic data about crimes and areas, to use an algorithm to predict the likelihood of crime happening in a location.	Using AI to evaluate children's responses to tests and assesses strengths and weaknesses, to provide progress-tracking and lesson plans to teachers & personalise tasks.

These case studies had a strong ‘why’, but participants needed more reassurance

Strong ‘why’: by delivering improvements in policing, security and education, these case studies delivered a **clear**, **substantial** and **direct** benefit and were welcomed by participants. In addition, these ‘whos’ were trusted for these particular purposes (for example, banks were trusted with providing account security, and police with tackling crime).

However, practical doubts about the promised outcome led to more caveated support, rooted in:

- **How?** Some were not sure whether automated systems could deliver the promised outcome. For example, by failing to recognize students' real needs, mis-identifying child abuse images, or failing to prevent identity theft.
- **Who?** For a minority who had negative views about the ‘who’, this raised doubts about the stated outcome. For example, those with negative views of the police were cautious about systematic or racist biases arising in algorithms.

These doubts undermined how clear, substantial and direct the benefits promised by these case studies were, and therefore weakened support for this data use.

Uses of data to deliver better consumer outcomes drew mixed reactions and a lot of neutrality

3

Ride-hailing app

App using an algorithm with data about people trying to book journeys to match drivers and customers quicker and increase or decrease its prices

Fintech mortgage start-up

Fintech company using Financial information such as credit scores, and data analysis by an artificial intelligence to provide a decision on mortgages within 15 minutes

Car insurance company

Car Insurance company using a blackbox to collect information about drivers' speed, time and location and analysing it using an algorithm to produce personalised car insurance premiums

I don't like the location one on that. It's useful when your cars get stolen… Knowing if you're speeding is fine, but why do they want to know where you are?

(30-39, Phase 2 Group)

These case studies had a 'why' with appeal to only a limited audience, with mixed views on the 'who'.

'Why' lacked substantial benefit: For most participants, the consumer benefits here are **clear** and **direct**, but not **substantial** enough to justify the invasive data collection. This is because the promised benefit was minor convenience gains or cost savings. In consequence, these case studies either drew negative reactions (from those who did want to 'trade off' their data for these marginal gains) or neutral reactions (from those who would not do so personally but felt that they can 'stay away' without facing negative consequences).

A minority, with personal experience or particular demands, regard it as substantial benefit: Positivity arose for those who can see a substantial benefit to:

- Themselves personally: for example, some users of ride hailing apps rated the service highly
- Society as a whole: for example, some older participants, felt blackboxes would make roads safer overall (though they would reject it themselves, they were in favour of others, especially younger drivers, using it).

For some, the 'who' of private companies created background skepticism about the benefits to society or individuals: this was particularly prominent about the ride-hailing app, where participants brought up news stories of exploitative employment practices.

Comprehensive data collection by government agencies for 'research' and 'detection' led to negative reactions

4

Gov. department and regulator	Government detecting welfare fraud
Using Smart Meter data and algorithms for research and to inform climate change policy	A government department using data about welfare fraud demographics and financial data to train machine learning and artificial intelligence to identify and stop welfare fraud

It does not make sense. Energy is sold from different companies, what is the point of keeping this data? … it is nonsense. You know where you are using it, where you're not, you can see the hotspots in your house.

(40-49, Phase 2 Group)

If it marginalises on those who are already on benefits, it is a bad idea. When we say fraud, we say benefit system. Fraud is a huge concept. It comes down to human communication on a lower level. It will reduce human hours. For me personally is a good thing, but not for the vulnerable and those marginalised in society.

(40-49, Phase 2 Group)

These two case studies combined an unclear, indirect 'why' with low trust in 'who'.

Unpersuasive 'why': Though participants saw tackling climate change and reducing fraud as **substantial** benefits, they were sceptical about the promised benefits as they were not **clear** or **direct**.

- Not **clear**: Participants could not see a clear, tangible benefit to themselves or society. It was easier to imagine tangible harms (such as being accused of fraud).
- Not **direct**: Use of data for 'research' in non-medical areas and to 'inform policy' was vague for participants, who could not see a clear explicable link between this data and the promised outcomes. For example, participants couldn't understand how data on electricity use can identify vulnerable people.

Unlike the previous set of case studies (to deliver better consumer outcomes) there is no option here to 'stay away' and consequently much less neutrality.

Generally, government handling of these issues (welfare fraud and smart meters) were already controversial among some participants, and this framed participants assessment of these case studies. For example, some participants rejected the welfare fraud case study, citing wider concerns about the welfare system being too punitive.

Social media targeted advertising was received with almost universal disapproval

5

Targeted advertising

A social media company, linking users' online habits and 'likes' to target ads for third-party advertisers

There's no virtuous motivation behind it. It just seems like entirely for capital gain. It's a bunch of big companies using data to make more money and I just think yeah there's no argument there that makes me wants to participate.

(18-29, Phase 2 Group)

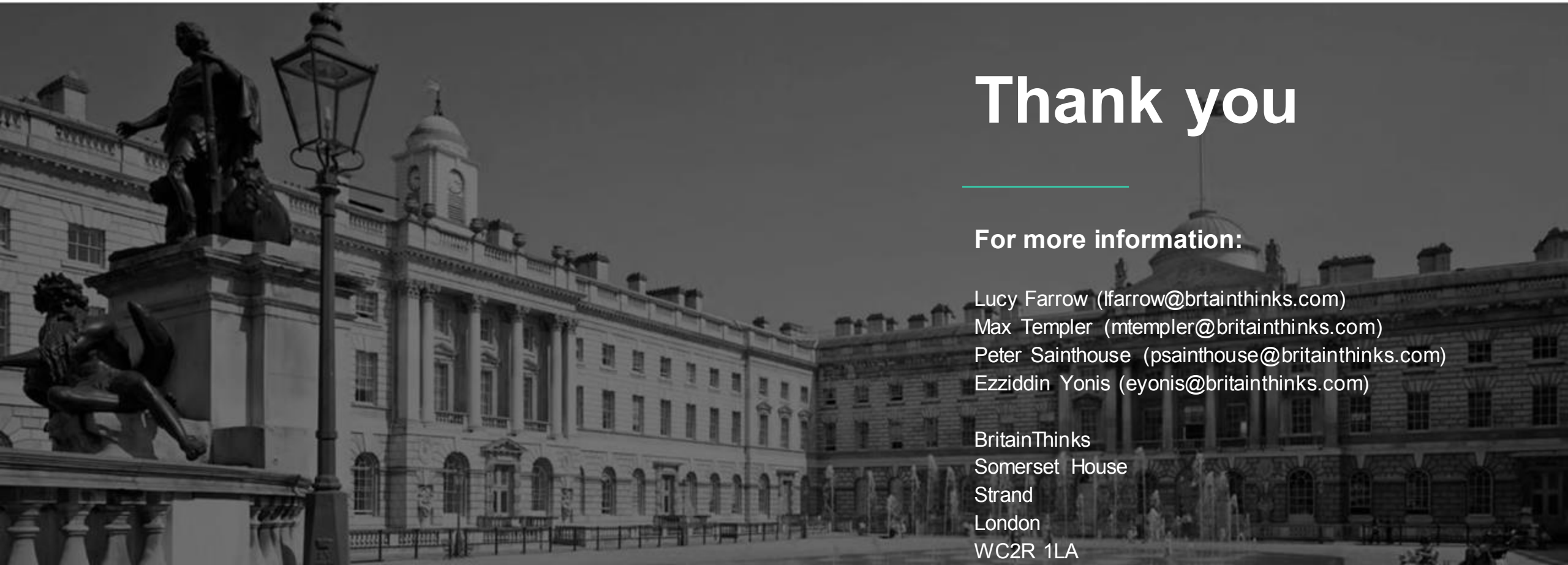
A very unpersuasive 'why' with very low trust in the 'who'

The 'why' failed to be persuasive.

- The benefits that were **clear and direct** did not feel **substantial**: the benefit of receiving more targeted adverts was seen as increasing convenience slightly, which did not 'justify' the amount of personal data collected.
- The benefits that were **substantial** did not feel **clear or direct:** Though accessing free, sophisticated online services (such as Google, or YouTube) were felt to be very substantial and tangible benefits, participants struggled to connect the collection of their data with services and content being provided for free.

Unlike the case studies to deliver better consumer outcomes, participants do not perceive an easy option to 'stay away' and consequently expressed far less neutrality (as with the government departments).

The 'who' drove scepticism about the benefits claimed for this data use. Both very low trust in their data being protected, and wider concerns about the social and political impact of social media companies, led participants to feel sceptical about the claimed benefits.



Thank you

For more information:

Lucy Farrow (lfarrow@brtainthinks.com)
Max Templer (mtempler@britainthinks.com)
Peter Sainthouse (psainthouse@britainthinks.com)
Ezziddin Yonis (eyonis@britainthinks.com)

BritainThinks
Somerset House
Strand
London
WC2R 1LA

5 Appendix

1

Medical research network (Current use of data)

Who?	**A regional NHS medical research network**
What?	**Personal and health data** about every individual with blood cancer
How?	**Linking** health records for individuals to various national data sources (e.g. the census)
Why?	Linking this data enables Doctors and Researchers to **better understand and therefore improve diagnosis and treatment of people with blood cancer**. For example, one potential use of this data is to link patients blood cancer status to their location of living, ethnicity, or family status to identify who might be most at risk. Data is anonymized before being linked.

1

NHS + AI company (Potential future use of data)

Who?	NHS, in collaboration with a well-known machine learning and artificial intelligence company
What?	**Breast cancer patient mammograms** and **breast cancer patient health records**
How?	Training an **artificial intelligence** to identify signs of breast cancer using mammograms and data about patients who later developed breast cancer
Why?	The earlier breast cancer is caught the more able doctors are to treat it. Whilst screening exists it can be difficult for humans to spot the signs of breast cancer at early stages. **Artificial intelligence could be used to spot the signs of breast cancer earlier than human experts.**

2

Smartphone company (Current use of data)

Who?	**A technology company who produce smartphones, tablets and computers**
What?	**User photos** and **images of missing and/or exploited children**
How?	An **algorithm** scans users' photos which are stored in the cloud, using **facial recognition** to identify any images including missing and/or exploited children
Why?	This algorithm is used to **detect child pornography, block this content and report users with this content on their devices to law enforcement**.

2

Well-established Bank (Current use of data)

Who?	**Well-established Bank**
What?	**Photos of ID** and **selfie of customer**
How?	**Artificial intelligence** uses **facial recognition** to confirm they are the same person
Why?	The bank wants to **speed up the process of opening an account without risking increasing identity theft**. By using artificial intelligence and facial recognition they can **confirm that the person trying to open an account is the same as the person on their ID.**

2

Police (Potential future use of data)

Who?	**The Police**
What?	**Historic data about crimes, demographic data about areas**
How?	**Algorithms process historic data about crimes and the demographics of an area** to predict the likelihood of crime happening there
Why?	Algorithms can be used to **predict the likelihood of a certain crime happening in a location.** This information can be used to allocate resources to that location to try and prevent that crime from happening and/or be on hand to catch suspects.

2

Tech education company (Current use of data)

Who?	**An educational technology company** making technology for use by teachers in their classrooms
What?	**Children's accuracy and speed in responding to questions** on English, maths or science
How?	An online platform for use in the classroom, with tests and homework based on the national curriculum. Students complete these and **artificial Intelligence** is used to **assesses children's strengths and weaknesses**
Why?	By using artificial intelligence to analyse children's accuracy and speed in responding to questions on English, maths or science, the online platform can **provide progress-tracking and lesson plans to teachers, as well as personalizing the homeworks/tasks given to students to fit their level and needs.**

Ride-hailing app (Current use of data)

Who?	**A ride hailing service**
What?	Data about **customers' and drivers' location and** data about **how many people are trying to book journeys**
How?	**Using phone location data and algorithmic decision making**
Why?	By collecting information about where customers and drivers are and using an algorithm, **this service can match drivers and customers quicker than a traditional taxi service**. Alongside this, by collecting information about how many people are booking journeys at any time, **prices can increase or decrease** in order to manage demand and workload on drivers.

Fintech mortgage start-up (Current use of data)

Who?	A new **UK finance technology start-up**, one of the world's first digital-only mortgage lenders
What?	**Financial information about potential customers,** such as credit scores, as well as **data analysis by an artificial intelligence** on data from all customers
How?	Using **artificial intelligence** to analyse potential customers' financial information and **an automated decision-making system** to make decisions about mortgage lending
Why?	This start-up aims to be able to provide a decision **within 15 mins**, by combining a decision-making engine with an easy online application process, reducing waiting times and eliminating the cost of brokers.

Car insurance company (Current use of data)

Who?	**A car insurance company**
What?	A **blackbox** which track users' driving habits including **speed, time and locations**
How?	The **blackbox** collects information about drivers' speed, time and location and this data is analysed using an **algorithm**
Why?	The insurance company can use the data collected by the blackbox and the analysis from the algorithm to produce **personalised car insurance premiums.** Drivers who are perceived to drive more safely get cheaper insurance premiums than those who drive more dangerously.

4

Government department and regulator (Current and potential future use of data)

Who?	A government department and the energy market regulator
What?	**Household data about electricity use**, including live information about current electricity use
How?	Using data collected from **smart meters** and processed using **algorithms** to aid decision makers and to support research
Why?	To help **help people reduce their energy use** (addressing climate change directly) and to **inform policy making about energy**, including identification of vulnerable people (based on their energy use) and policies about energy consumption.

4

Government detecting welfare fraud (Potential future use of data)

Who?	**A government department**
What?	**Data about welfare fraud demographics** (e.g. age, gender, location etc.) and **financial data held by private companies (**e.g. car insurance providers, banks)
How?	Using **machine learning and artificial intelligence** trained using **data about welfare fraud**
Why?	Government currently estimates that it loses between £31-49 billion a year due to fraud. The aim of this program is to help **identify and stop welfare fraud** by using machine learning and artificial intelligence. This could be used to identify changes in the the welfare process that would make fraud harder to commit, or it could be used to identify individuals suspected of fraud for further investigation.

5

Social media company (Current use of data)

Who?	**Social media company**
What?	**Users' online habits and information** provided to Facebook
How?	This company **links users' data from its own platform** (e.g. demographic information, "Likes") and can **also link this to other online information** (e.g. websites that users log into using the email address registered with their social media account)
Why?	This social media company uses customers' information and other online information shared with them to **target ads for third-party advertisers at users of their social platform.**