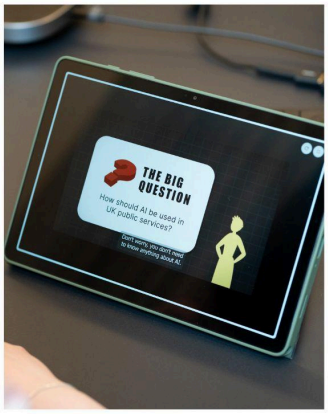
AI Social Readiness Advisory Label

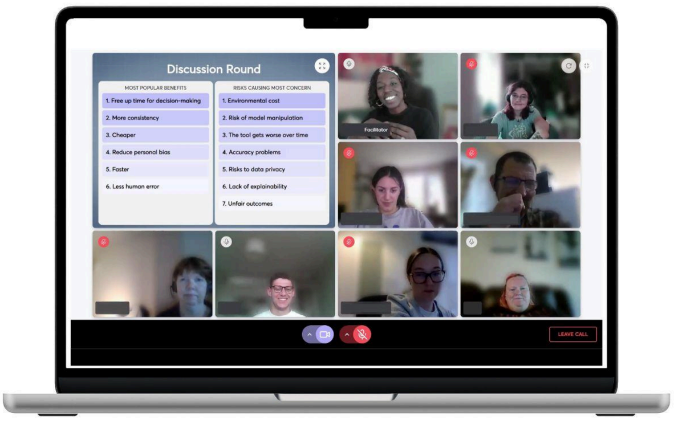
Centre for Collective Intelligence
Part of the nesta group

AI Social Readiness Advisory Label Report: For the AI tool 'Consult'

(Accessible Version)

July 2025





Contents

What is the AI Social Readiness Advisory Label?

In the UK, public sector organisations are facing increased pressure to adopt AI technology to improve the speed and quality of public services. To do this well, it is important to secure public trust by deploying AI responsibly, but there is little practical guidance on how this can be achieved.

The AI Social Readiness Advisory Label process fills an important gap in the AI assurance landscape. It measures public confidence and trust in specific AI tools being used in UK public services, and provides easy-to-understand guidance on how to address public concerns.

It has been designed to support public sector leaders as they make decisions about AI procurement, deployment and risk management. It should be used alongside other information including technical evaluations and compliance processes.

The results in this report and on the label capture the views of 144 people broadly reflective of the UK public, collected during 18 small group deliberations. We use a structured approach known as deliberative polling, which captures both quantitative and qualitative data from individuals and groups as they learn about the tool and weigh up the benefits versus the risks. Participant opinions are italicised and presented in quotation marks throughout.

See Methodology for further detail.

The AI Social Readiness Process

x 144 participants. Either online or in-person.

3. Participants are introduced to AI and its general risks and benefits to create a baseline of AI understanding. They are then introduced to a specific AI tool, and the public service it would be used in. Groups discuss risks and benefits of this AI tool, then vote on whether it should be used, and under what conditions.

Public Deliberation Sessions

4. The data from all the sessions are aggregated.

5. The results are summarised into the AI Social Readiness Advisory Label, and a detailed report.

Data Analysis & Reporting

The AI tool is assessed against a standardised set of evaluation criteria.

2. The content for the experience is developed, then reviewed by experts.

Technical Assessment & Content Development

AI Social Readiness Advisory Label - Consult

The Incubator for Artificial Intelligence (i.AI) is a team of technical experts in the Government Digital Service (part of the Department for Science, Innovation and Technology) who prototype and deliver AI tools to build a modern digital government. They have developed a tool called 'Consult' which aims to make the process of analysing responses to public consultations faster and more insightful. The tool uses AI to extract patterns and themes from the responses, giving policymakers control of which themes to use before turning them into interactive dashboards for policymakers. This means humans are in control and freed to do the work of drawing meaningful insights from those patterns.

AI Social Readiness Advisory Label for 'Consult'

Click to download the full PDF

Public views on the Consult tool

Most people were positive about the use of the Consult tool.

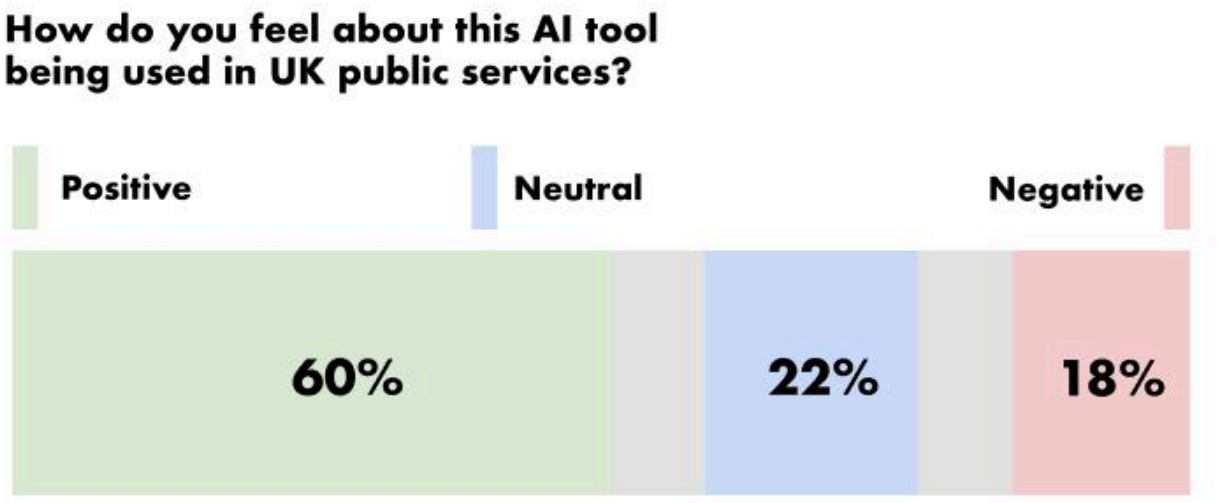


Figure 1: How do you feel about this AI tool being used in public services? n=144
(5-point Likert scale, Very negative <-> Very positive)

This is the final question in the AI Social Readiness Advisory Label process. We ask it after participants have discussed how the tool works, what they think about the benefits and risks on balance and which safeguards they would like to see in place.

People liked that the task performed by the Consult tool was well constrained, that humans are still involved in the process and that the AI isn't used to make final decisions, which made them more comfortable with its use.

Participant quotes:

"I think it's good to use in general because it seems like a simple task for AI to manage."

"I like that it includes the human touch, the constant checking of it."

"AI is doing the donkey work here and I like the idea that the experts go deeper and go into the insights."

But many people expressed reservations about the ability of AI to accurately capture and categorise nuance, for example, if the responses contained emotional language, regional slang or sarcasm.

Participant quotes:

"What slang can it understand? What grammar?"

"Languages and emotions might be very personal and AI might not be able to understand what the text really means."

A majority of people agreed that using the Consult tool will benefit the consultation process.

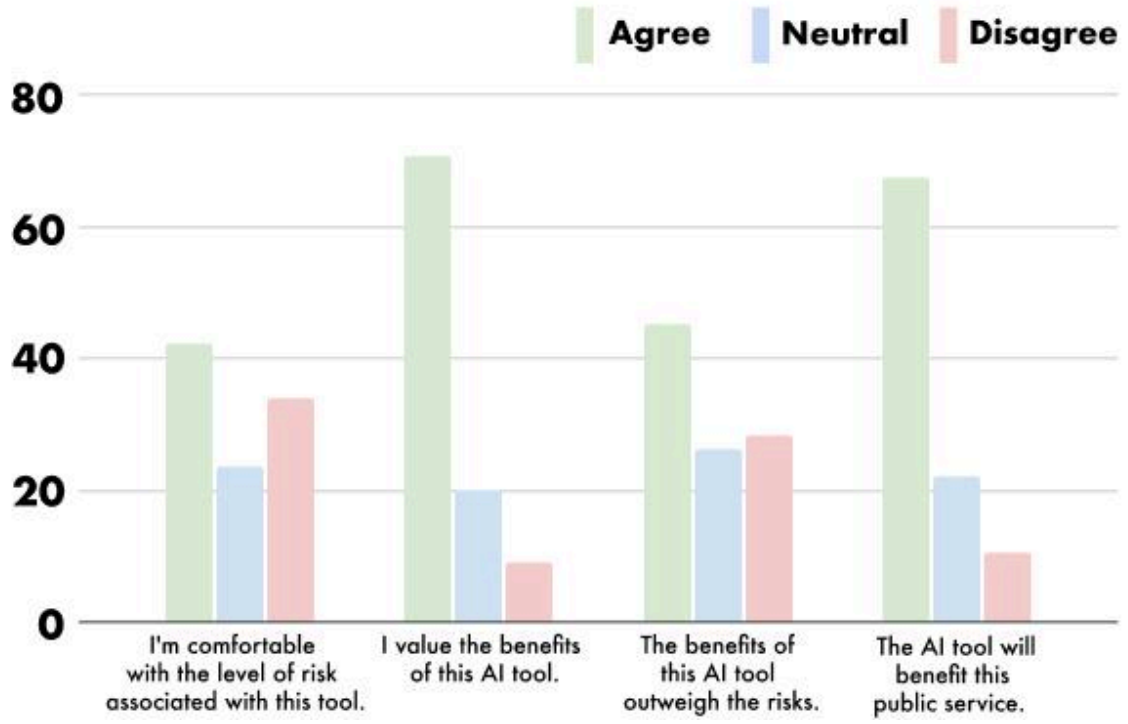


Figure 2: Considering what you've heard, how do you feel about these statements? n=144
(5-point Likert scale, Strongly disagree <-> Strongly agree)

We ask this after participants have discussed how the tool works and what they think about the benefits and risks on balance.

Overall, 67% of people agreed that the Consult tool would benefit the consultation process, while 10% disagreed.

More people (45%) thought that the benefits of the tool outweighed the risks than the opposite (29%).

During deliberations, people acknowledged that it would be impossible to entirely eliminate risks. They also discussed their satisfaction with mitigations as a factor in weighing up the trade-offs.

Participant quotes:

"I think the benefits outweigh the risks. Because a lot of these risks, they're risks that we deal with every day anyway."

"They seem to have done a lot to lessen the risks, except the ones at the top of our list [environmental, model manipulation] - but on balance I say use it."

"The current checks and balances aren't up to scratch. If they're addressed, I'd feel differently."

People think the consultation process has a lot of room for improvement and that AI can help.

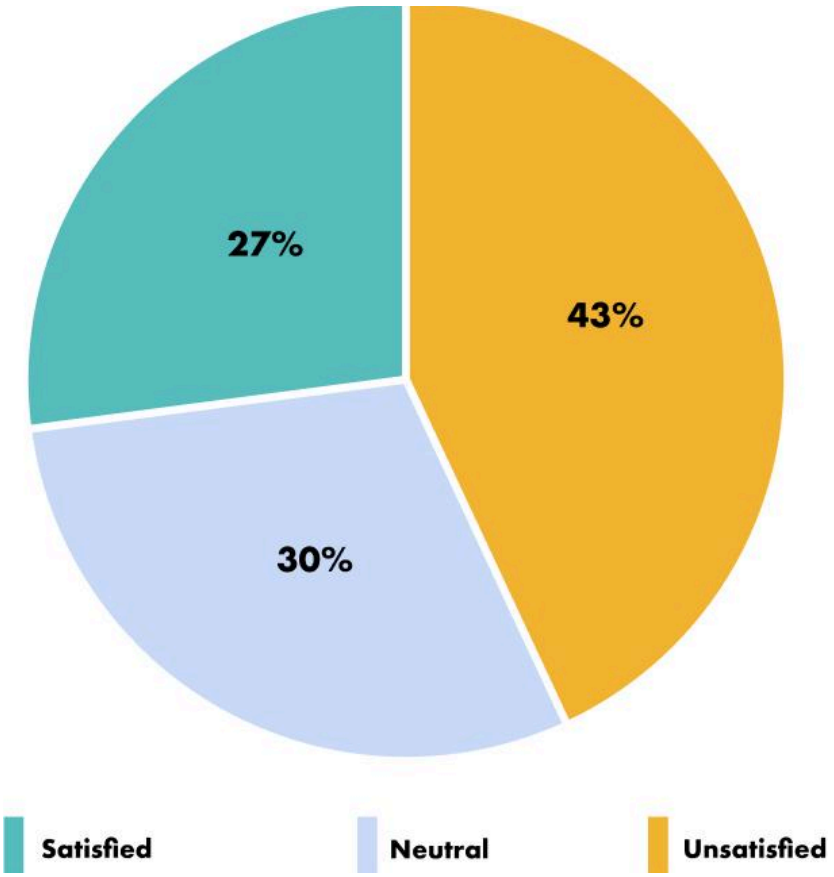


Figure 3: How satisfied are you with the current process? n=144
(5-point Likert scale, Very negative <-> Very positive)

There was low awareness of public consultations and how they worked among the participants. People valued the principle of public input into decisions but considered the current process outdated, inaccessible and inefficient. Participants expressed concerns about the limited reach of consultations and scepticism that public views would genuinely be considered.

Participant quotes:

"I think it's always important to ask people."

"This feels like a very old school process, very archaic."

"I am concerned about how they are advertised, it feels like only the savvy people would know."

People saw the potential for AI to address current inefficiencies and inconsistencies in the consultation analysis process.

Participant quotes:

"AI should be able to help with this."

"I thought it was quite a crude way of doing it, sorting it into piles - maybe AI could be better at it."

Further insights from deliberations about the current process of public consultation

Main discussion prompts:

- What do you think about this [the public consultation] process?
- What did you like?
- What didn't you like?
- Why?

People had their own ideas for how the consultation process could be improved.

People had suggestions for other ways that AI could address some of their concerns about the current process.

> **Participant quotes:**
> *"If we use these types of tools we can start thinking about audio and video responses [...] That would be more inclusive and I think people would give better answers."*
>
> *"Maybe in the future they could have the tool recognise handwritten [responses]."*

They also discussed that public consultations needed more significant changes, and, rather than just automating part of the existing process, technology could help with this change.

> **Participant quotes:**
> *"...rather than automating it, why not think of different ways to understand what the public thinks?"*
>
> *"Isn't this the point of developing new technology, so we can rethink the system?"*

Public views on the benefits and the risks of the tool

All of the potential benefits of using Consult were considered important.

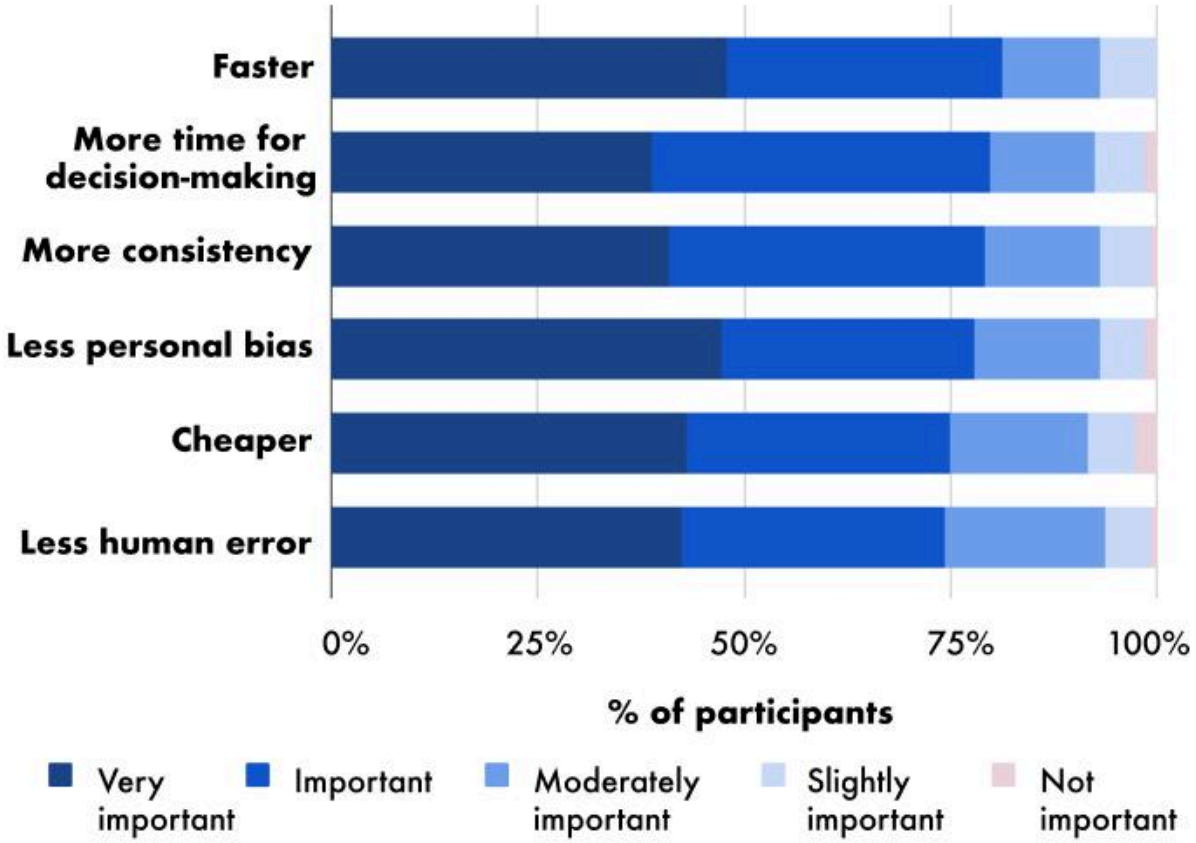


Figure 4: With everything you've learnt about this AI tool, how do you feel about these benefits? n=144
(5-point Likert scale, Very important <-> Not at all important)

All benefits were considered important by 74% of participants. Discussions revealed more nuanced views. For example, for cost-saving, people wanted to know how these weighed up against the maintenance and development costs for the tool.

Participant quotes:

"The problem is how much is this going to cost the government to try it, will it be more or less than the 80 million? [estimated annual cost of consultations]"

"I'm sure that software is expensive."

They also wanted more information on how the gains in money and time would be used by the organisations deploying Consult.

Participant quotes:

"If instead of making people redundant, it allows them to do something else."

"Will this create time for more consultation, involving the same number or more jobs?"

Participants were less concerned about the risks posed by Consult compared to risks from AI tools in general.

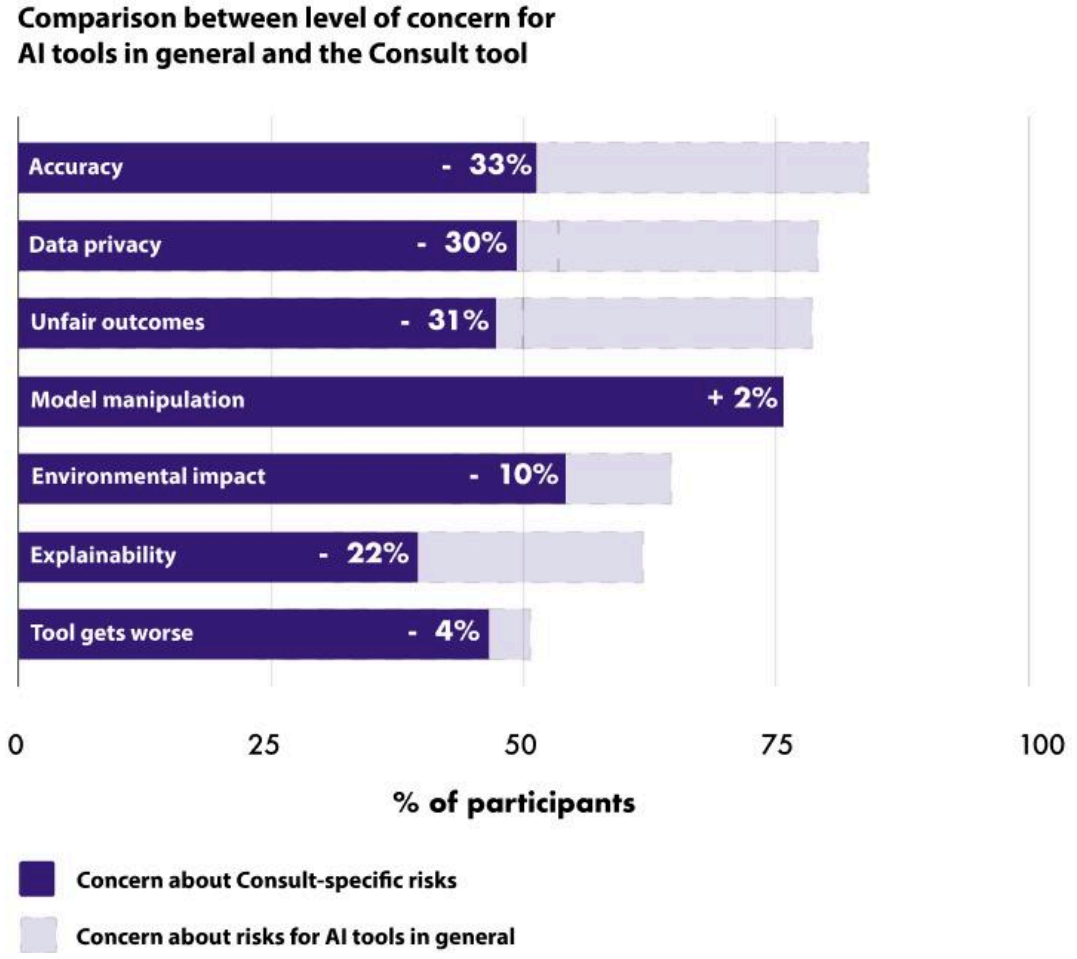


Figure 5: How concerned, if at all, are you about these risks? n=144 (5-point Likert scale, Not at all concerned <-> Very concerned. Bars show the percentage of people who were "Very concerned" or "Moderately concerned")

When we asked people how they felt about a range of risks from AI tools in general, over 50% of people were "Very concerned" about all the risks mentioned. Inaccuracy; unfair results; and data privacy threats were considered the top three risks.

Later in the process, participants reviewed all risks specifically in relation to the Consult tool. Overall people were much less concerned about the risks posed by Consult.
Model manipulation and environmental impact were the top two risk areas for the Consult tool. More than 75% of participants remained concerned about model manipulation after hearing about the mitigations planned by the i.AI team.

During deliberation, people discussed that these risks weren't well quantified and they weren't entirely satisfied with the mitigation strategies described.

Participant quotes:

"The risk of manipulation - [...] it sounds like something they haven't got to grips with"

"The environmental costs, sounded like it's something they plan to do - but it sounds like wishful thinking."

Further insights from deliberation about benefits and risks of the Consult tool

Main discussion prompts:

- What did you like?
- What didn't you like?
- We've heard about how Consult works, the benefits and how the team are managing some of the potential risks… So what do you think on balance?

Many groups were uncomfortable that Consult used an AI model developed by a US company, but on balance felt the benefits of the Consult tool outweighed the risks.

People felt the benefits of Consult were straightforward and useful. Despite some misgivings, most groups ultimately felt comfortable with its described use.

> **Participant quotes:**
> *"In terms of the efficiency of the process, I can see why they would do that [use the tool]"*
>
> *"With all the risks and benefits...I think with the safeguards and mitigations, I feel reassured and comfortable."*

Additional concerns reflected anxiety about the increasing domination of big tech companies, potential threats to data privacy, and whether the model might be politically biased.

> **Participant quotes:**
> *"OpenAI should have nothing to do with providing services to the government, they are doing the AI for themselves, not for public benefit."*

"Open AI is a US company and will [they] have access to UK government data? I would want to know what is happening with the data sharing."

"I want to know how it [the model] has been taught and that it's not either right or left leaning."

See the tool developer's response to recommendations.

Public views on safeguards for trustworthy deployment

85% of people thought human oversight is a necessary safeguard for using the Consult tool.

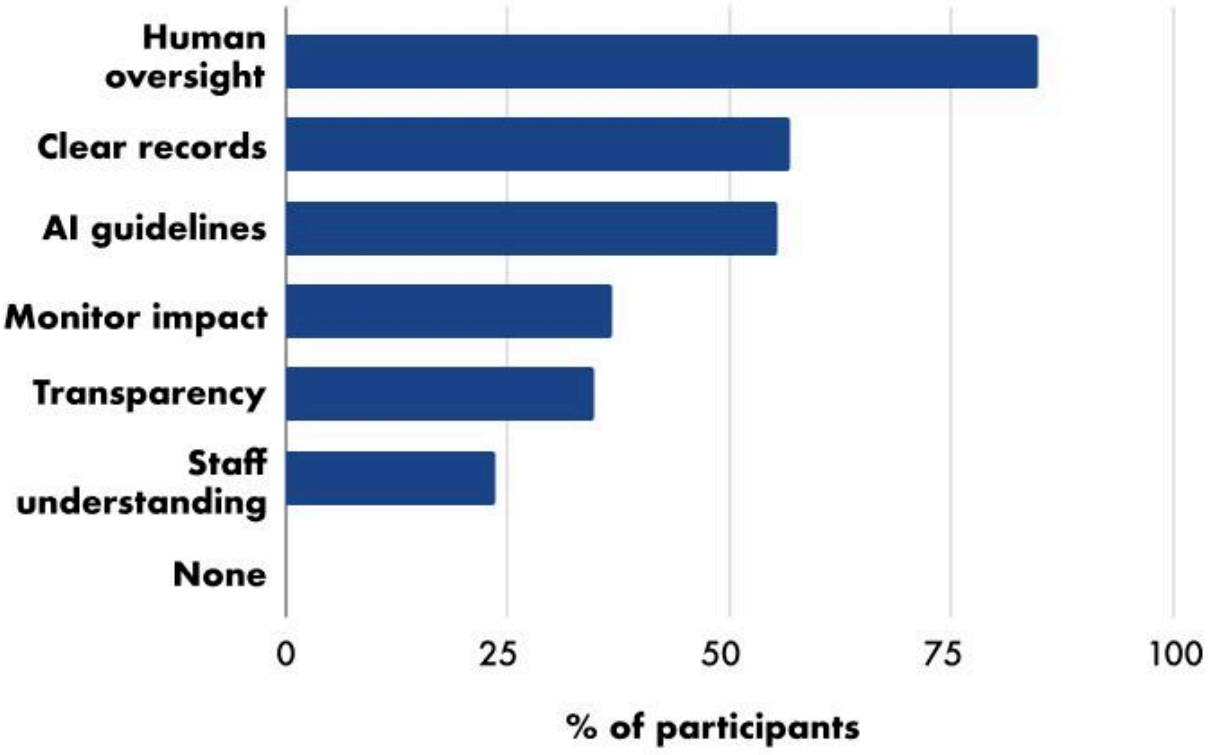


Figure 6: Which of these safety measures matter most to you? Select up to three. n=144 Participants made their selection after discussing the different safety measures with their group.

85% thought human oversight is a necessary safeguard. They recommended the proportion of responses checked by people should depend on the specific topic of the consultation.

> **Participant quote:**
> *"It depends on the weightiness of the decision, that should influence how big the sample is that is checked."*

More than half of participants also felt that keeping clear records and AI guidelines were important, as long as there was a way of enforcing these to ensure accountability.

Participant quotes:

"I wouldn't like them to be just guidelines, I want them to be mandatory. There needs to be real repercussions and accountability."

"The one I found interesting is about keeping clear records to be able to come back to the decision."

Further insights from deliberation about safety measures for deployment

Main discussion prompt:
Which of these safety measures would you most like to see in place and why?

People wanted robust safety measures in place.
All participants selected at least one safeguard as important for safe deployment of the Consult tool. People were often surprised that all the measures aren't mandatory.

> **Participant quotes:**
> *"We were asked to only pick three but I wouldn't want anyone to think the others are not important."*
>
> *"I would say all of these measures should be there as a bare minimum."*

People discussed the best ways to implement safeguards to make sure they stayed robust over time. For example, people wanted reassurance that public sector staff would maintain their skills for effective oversight.

> **Participant quote:**
> *"I think it's very important to have human oversight by someone who has a lot of experience."*

People thought transparency was important for ethical reasons but several groups also acknowledged the potential negative consequences of revealing that AI was being used.

Participant quotes:

"It might be counter-effective to tell the public [...] people might not understand what it's doing in the process and it might put people off"

"I would wanna know, regardless of if I don't understand, they could redirect you to where it can be explained."

Building public trust and understanding in AI

98% of people agreed the AI Social Readiness process was a good way of involving the public in assessing AI tools.

People want to be involved in decisions about how AI is used in public services. They agreed that the AI Social Readiness process is a good way of doing this because it allowed them to get balanced information on the topic and hear a diversity of views from other people. They also found it enjoyable and important.

- 4.8 out of 5 for deliberation quality. (See our Methodology for details on how deliberation quality scores are calculated.)
- 8.9 out of 10 for overall enjoyment
- 98% agreed the AI Social Readiness process is a good way of involving the public.

Participant quotes:

"I think it's important for big decisions about using a tool like this, that they can say they haven't taken this decision lightly. And can say they've spoken to x number of people, and really considered the pros and cons."

"I think the process of bringing people in is incredibly important, I'd like more of that."

"If you are getting AI involved in public services it goes without saying you should see how people feel about it, it should be a natural process."

"I thought it was brilliant - it made it fun for people to interact. It made it feel like your opinions were really valid and you were really helping to shape the future."

The AI Social Readiness process helps to build public understanding.

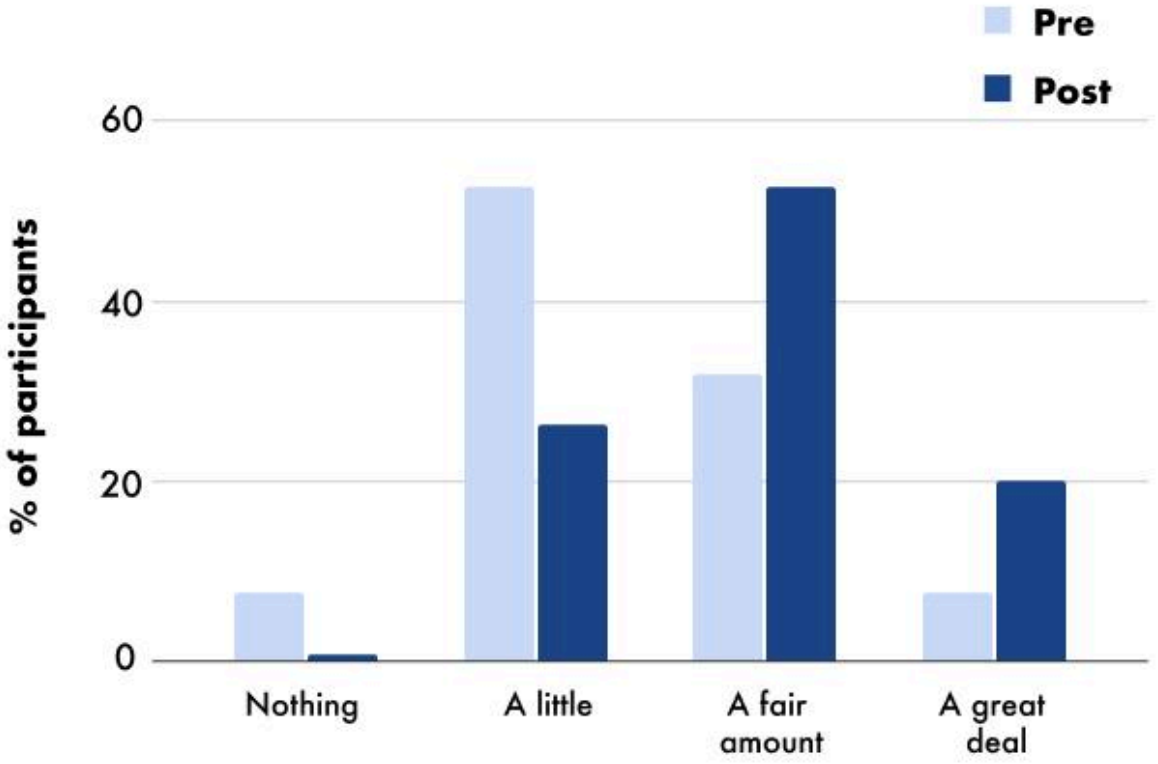


Figure 7: How much, if anything, do you know about how AI can be used for public services (e.g. education, health or benefits)? n=144

There was a 33% increase in understanding about use of AI for public services among participants.

People recognised the importance of going through the AI Social Readiness process to allow them to make an informed judgement on public sector AI.

> **Participant quote:**
> *"Hearing what other people think has challenged me to consider new things, it's different to when you're answering a questionnaire where you only have your own opinion to go on."*

"Our experience today is really key to being able to understand and judge the process."

"It's important to involve people more, because even in our group our knowledge has been improved. In terms of getting people on board with change it's important to educate people."

The AI Social Readiness process resulted in an 11% increase in trust that the public sector will use AI responsibly.

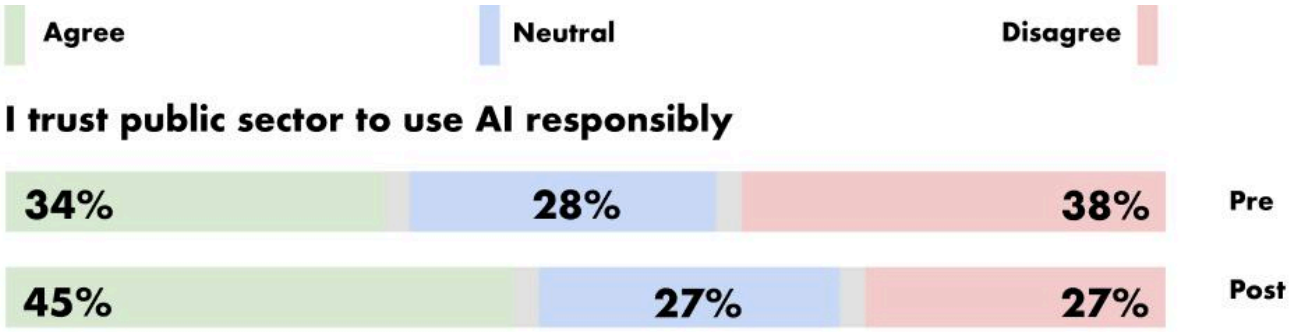


Figure 8: We ask: How do you feel about this statement? - I trust public sector organisations to use AI responsibly n=144
(5-point Likert scale, Strongly disagree <-> Strongly agree)

(Some figures may not add to 100% due to rounding.)

At the start of the AI Social Readiness process 34% of participants agreed that they trusted public sector organisations to use AI responsibly. This had increased to 45% by the end of the process.

> **Participant quote:**
> *"I think it's really positive, it gives me a bit of confidence that the Government is asking the public about what they think about AI."*

Recommendations

In this section, we provide actionable recommendations for the developers who are building the Consult AI tool; and advice for organisations planning to use it.

These are derived from the preferences captured through quantitative polling, as well as the group deliberations. All recommendations stem directly from views expressed by the public.

Recommendations for i.AI as developers of the Consult AI tool

Develop strong mitigation strategies for model manipulation and environmental impact risks: The public specifically also wanted information to understand the scale of environment risks and more detail on what was being done to mitigate both risks.

Explore alternative AI models: Consider alternative models to OpenAI's GPT-4o to address public concerns. If switching to a different model, it will be important to re-evaluate the impact of this change on performance metrics and risks.

Account for development and maintenance costs: If cost-saving is emphasised as a key benefit, estimates for the ongoing development and maintenance costs should also be accounted for.

Consider involving the public in user-research for future development of Consult or new AI tools: People had ideas for other ways that AI might be used to support the consultation process, with a particular focus on improving the accessibility and inclusiveness of public consultations.

Explore features to support trustworthy deployment: The developers could support organisations to keep clear records about decisions

made during the analysis by adding this feature to the tool.

See the tool developer's response to recommendations.

Advice for trustworthy deployment of Consult by organisations

Always have human oversight: Organisations should take particular care with their sampling approach when dealing with topics that are sensitive, use specialist terminology or require local knowledge, ensuring it is sufficient to give confidence of accuracy. Ensure staff are able to provide effective oversight.

Record decisions: Any organisation using Consult should keep clear records of how they decided on themes to help with accountability.

Be clear with the public that AI is being used: Organisations using Consult should tell people about it and explain how it works in simple language. They should also consider additional measures to prevent deterring people from submitting responses.

Track cost savings: Organisations planning to use Consult should track costs savings and account for how these savings are being used to improve public services or benefit the public.

Measure wider impacts: Beyond efficiency and cost, it is important to monitor how Consult affects participation in the consultation process as a whole or by specific groups, like how many people respond and how diverse those responses are.

Consider wider reforms of the public consultation process. There was low awareness about consultations, and people thought the process should be updated to improve reach and diversity of responses.

Label validity

The public confidence ratings suggest overall public satisfaction that the tool, in its current form, would enhance public consultations. They should be considered alongside the other results and the recommendations in this report to ensure the tool continues to meet public expectations.

Significant changes to Consult's technical system or how users interact with it will require a new Social Readiness assessment.

This could include:

- Changes to the user workflow that remove human verification and approval of AI-identified themes.

- Extending the tool's capabilities to include recommendations that affect the interpretation and decision-making steps of the analysis.

- Changes to the tool's technical pipeline that impact performance accuracy, bias, data privacy, environmental impact or vulnerability to manipulation.

For these or similar significant changes, the i.AI team should complete a full re-evaluation of risks and mitigations, and secure renewed public acceptance

Response from the developer of the AI tool

General impressions of the process

We are delighted to have supported Nesta in developing this label. We in i.AI are committed to developing AI products that the public can trust. The process enabled us to meaningfully engage the public in the design of the tool. Alongside our wider evaluation and assurance plan this research supports us to develop tools responsibly.

Consultations are an important part of the democratic process so having an understanding of public confidence and trust about the use of AI in the process is vital. This process has confirmed that many of the decisions we have been taking – including testing alternative AI models to reduce reliance on a single provider – are on the right lines. We are also pleased to accept the other recommendations, to continue making Consult stronger.

Response to public concerns

The quality of the recommendations demonstrated the quality of the conversation which Nesta facilitated with the public. Overall we are encouraged by the fact that the public was more supportive of the use of AI in Consult than general AI products. However, we also want to address some of the specific concerns raised:

Use of OpenAI models

We agree it is important to have flexibility in model choice. All tools developed in i.AI are model agnostic, meaning that we can change the model used if it no longer meets the UK public's needs, be that on

price, performance, environmental impacts or other factors. Noting the concerns about costs, a recent consultation analysed using the tool cost less than £240 in tokens with OpenAI and prices are trending down as the models become cheaper to run.

While the tool currently uses OpenAI – and the public raised this as a concern – the team is building up an evidence base to test other models to ensure we deliver the best solution for the UK public. The contract with OpenAI makes sure that data is not stored or used for training.

Model manipulation and human oversight

Human oversight will remain an important aspect of ensuring Consult's ongoing performance, including protection from any attempts at model manipulation.

We are working with academic experts to explore the risk of model manipulation, and conduct a thorough red teaming exercise to uncover potential model manipulation risks. This will include making sure that we have robust mitigations and safeguards in place for all identified vulnerabilities.

We are also working on design choices to ensure human review is a seamless and core part of using Consult. This sits alongside statistical analysis to identify how human review can most efficiently target potential risks, ensuring they are mitigated effectively, and enables us to adjust the risk assessment based on the specific consultation.

Understanding nuance in language

Our testing of Consult across a wide range of consultations is designed to ensure it performs in all contexts, including handling dialects, slang and sarcasm. This is something we will continue to monitor in our

ongoing testing, and we are also looking at the wider academic literature to understand how Large Language Models (LLMs) in general perform with different use of language.

Planned actions

Members of the i.AI team are reviewing each recommendation and considering how to incorporate them into the design and delivery of Consult. Some specific examples include:

Assurance and measuring impact

We are conducting extensive testing of Consult to ensure it is fit for purpose before wider roll-out, and the first model evaluation and Algorithmic Transparency Recording Standard (ATRS) record have already been published as part of our commitment to working in the open.

We are also considering how best to capture wider factors, including environmental and development costs. We will aim to publish key figures for all our tools in due course, including the token costs of running them, and are hoping to add in other costs in future updates.

Explore alternative AI models

In i.AI we are continuously testing different models for our tools to optimise performance and efficiency. OpenAI's GPT model is currently used within Consult as it has outperformed other models. We will continue to evaluate different models as they are brought into the market to ensure the best possible performance.

Designing with users

We will continue to test and learn by working closely with users in government, who are analysing consultations to understand if the tool is meeting their needs. These teams will help to shape best practice for areas such as record keeping, safe deployment, guidance needed to use the tool effectively. If we move into public-facing aspects for Consult, such as changing how the public contributes to consultations, we will engage with them.

Methodology

Nesta's Centre for Collective Intelligence completed the research between May 27th and June 13th 2025. During this period we held 18 deliberative polling workshops, with 144 UK adults. 9 workshops were held online, and 9 workshops were held in person (3 each in London, Manchester and Newcastle). See Appendix for a full breakdown of demographics.

Participants were recruited through the recruitment agency Roots to be broadly reflective of the UK population, ensuring a range of demographic backgrounds across age, gender, region, ethnicity, education and socioeconomic status.

Participants receive £70 remuneration for taking part in the online experience and £90 remuneration for the in-person experience (to account for travel costs). Further details about our methodology can be found in our OSF pre-registration.

The AI Social Readiness Advisory Label process

Our public AI assurance process is immersive, educational and collaborative. During the experience, small groups undertake a mission as part of a "Public AI Task Force". Guided by expert facilitators, the groups work through a structured format to review short videos and discuss and express opinions about the use of an AI tool in public services. The videos break down complex topics into easy-to-understand, accessible content covering e.g. What is AI? How does the Consult tool work? What safeguards are available?

Through this process, ordinary people from a variety of backgrounds have informed, balanced discussions on the potential risks and benefits of a specific tool, and articulate any specific recommendations or conditions for its use.

The sessions are delivered through the Centre for Collective Intelligence's digital platform Zeitgeist, using a structured approach known as deliberative polling. This method allows for the collection of quantitative and qualitative data from individuals and groups throughout a discussion, helping to track how opinions change over time.

During in-person engagements, participants are seated around a table, each with a tablet device to review videos and record responses. During online engagements, participants connect to an online version of Zeitgeist, which hosts the same video content and polls as the in-person engagement and allows for the additional functionality of interacting with other participants via a video-calling interface.

Public Deliberation Session - Content Overview

The full experience lasts approximately 210 minutes, including two breaks. Each section contains a combination of video content, polling and deliberation. There are 5 distinct parts:

- **Section 0: The Briefing** - Setting the scene, demographics, baseline attitudes.

- **Section 1: The Groundwork** - What AI is, how it works, potential risks, benefits & trade-offs: AI tools "in general".

- **Section 2: Mission Deep Dive** - Overview of current public service & how the tool works.

- **Section 3: The Dilemmas** - Explore real-world risks, benefits & trade-offs of the tool & potential safeguards.

- **Section 4: Mission Debrief** - Feedback, post-experience attitudes,

overall enjoyment & deliberation quality.

Technical Assessment and Content Development

All information is provided during the session through engaging videos, which are created in advance based on a technical assessment of the AI tool performed by an independent expert. The expert evaluates the tool against a standardised set of risk criteria (see glossary for relevant tool-specific and deployment risks) using information provided by the tool developer. This might include evaluation results from pilots, or performance metrics and technical documentation from the Algorithmic Transparency Recording Standard.

Information from this assessment is used to create the tool-specific video scripts in the Mission Deep Dive and The Dilemmas sections of the AI Social Readiness process. The scripts are reviewed by the internal team, an AI expert and the tool developer for accuracy and balance. All other content, including the videos in The Briefing and The Groundwork sections (e.g. What is AI?) remain unchanged for each AI tool that goes through the process.

Data Analysis & Reporting

Quantitative analysis and data visualisation - Data are collected throughout each session, typically using 5-point Likert scales. After all sessions have been completed, the data are aggregated. We visualise the results using an adapted 3-point Likert range. For example, for a Strongly Disagree <-> Strongly Agree scale, we report Strongly disagree and Disagree as Disagree, and Strongly Agree and Agree as Agree.

Qualitative analysis - After all sessions have been completed, we review group deliberation transcripts for each deliberation in the experience. In total, there are nine deliberations ranging in length from

5-12 minutes. We identify and count themes using a combination of human and AI-supported analysis. All quotes in the report are from deliberations that happened during the sessions.

Public Confidence Ratings and Deliberation Quality

On the AI Social Readiness Label, we report Public Confidence Ratings. These are based on the public's responses to five key questions from in The Dilemmas section of the experience. Individuals answer these questions after they have learned about and discussed all of the information about the current process, how the AI tool works, its risks and benefits and further considerations about its deployment.

Considering what you've heard, how do you feel about these statements?

1. I value the benefits
2. The benefits outweigh the risks
3. I am comfortable with the risks
4. The tool benefits the public service

Ranked on a 5-point Likert scale (Strongly disagree <-> Strongly agree).

How do you feel about this AI tool being used in UK public services?
Ranked on a 5-point Likert scale (Very negative <-> Very positive).

Deliberation Quality is a calculated index based on responses to 3 statements collected at the end of the experience:

1. Everyone had the opportunity to contribute to discussions.
2. People with different opinions were able to explain their point of view.
3. Group members paid attention to others when they spoke.

Further detail on session design, Likert ratings and deliberations prompts is available in the OSF pre-registration.

Appendix - Overview of benefits, risks and safeguards discussed during the AI Social Readiness process

Benefits

During the experience, participants are asked to rate tool-specific benefits in response to the question: "**With everything you've learnt about this AI tool, how do you feel about these benefits?"** (5-point Likert scale, Not at all important </> Very important). The relevant results can be found in Figure 4. They learn about the benefits in The Dilemmas section after learning about the current public service process and how the tool works, and before the risk mitigations.

Faster and Cheaper: Delivering insights more quickly and at a significantly lower cost than manual analysis.

Frees Up Time: Saving analysts time by automating repetitive tasks. This provides additional time to focus on additional tasks.

Reduces Human Error: Unlike humans, who may miss a few responses during analysis, the Consult tool ensures every single response is captured and categorised—improving the accuracy of the results.

Consistency Across Consultations: Applying the same process every time, reducing variation between different teams and ensuring that responses are analysed consistently across different consultations.

Reducing Human Bias: Minimising unconscious bias by identifying themes in consultation responses without being influenced by analysts' expectations. This leads to more objective and inclusive insights.

Tool-specific risks

During the experience, participants are twice asked to rate seven tool-specific risks in response to the question: "**How concerned, if at all, are you about these risks?"** (5-point Likert scale, Not at all concerned <-> Very concerned). The relevant results can be found in Figure 5. They first learn about the tool-specific risks in The Groundwork section, and then again in The Dilemmas section where they return to these risks and their mitigations in the context of the Consult tool.

Unfair Outcomes: When an AI system produces biased or unequal results due to flawed or incomplete training data. This can lead to discriminatory decisions, such as overlooking qualified job applicants based on non-standard CV formats.

Accuracy Problems: Errors or incorrect outputs generated by AI models. While some mistakes are minor, others can have serious consequences in critical areas like healthcare, finance, or criminal justice.

Risks to Data Privacy: The potential for sensitive personal data processed by AI systems to be exposed, misused, or stolen if not properly protected. This includes financial, medical, or identity-related information.

Lack of Explainability: A situation where it's difficult or impossible to understand how an AI model makes its decisions. This lack of transparency can undermine trust, especially in high-stakes areas like loan approvals or public benefits.

Getting Worse Over Time: The gradual decline in an AI model's performance if it isn't updated with new data. As the real world changes, outdated models can become less accurate or relevant.

Model Manipulation: Deliberate attempts to trick or exploit an AI system by feeding it misleading or false data by malicious actors. For example, attackers might fool a fraud detection tool to let fraudulent

activity go unnoticed.

Environmental Cost: The significant energy and computing resources required to train and operate AI models, contributing to carbon emissions and environmental impact due to high electricity and cooling demands.

Deployment risks

During the experience, participants learn about five deployment risks in The Groundwork section.

Lack of Transparency & Accountability: When people aren't informed that AI is being used, or can't see how it's making decisions. This is especially concerning in public services, where people need clarity on how decisions are made and how to challenge them if necessary.

Over-Reliance on AI: Depending too heavily on AI tools without human oversight. This can lead to loss of critical thinking skills among staff and increase the risk of letting mistakes go unnoticed.

System Failures: Breakdowns or outages in AI systems that perform essential tasks. Even brief failures can disrupt services and lead to serious consequences, especially in critical sectors.

Job Changes: The impact of AI on employment, including roles being reduced, reshaped, or replaced. While AI may also create new jobs, the overall effects on the workforce remain uncertain.

Public Experience of Services: How people interact with AI-powered services. While automation can improve speed and efficiency, it

can also reduce empathy and human connection—key elements in areas like healthcare or social services.

Participants return to these risks when discussing the safeguards that can be put in place by organisations using the tool.

Safeguards

These safeguards are discussed at the end of The Dilemmas section. Participants are asked: **"Which of these safety measures matter most to you? Select up to three."** See Figure 6.

Have AI guidelines: Develop clear guidelines on how to work responsibly with any tool that is powered by AI models, and ensure all staff are aware.

Improve staff understanding of the tool: Make sure staff using the tool know its benefits and pitfalls, and how to use it appropriately.

Be Transparent: Provide information to the public about how they use the tool to analyse the consultation responses.

Keep Clear Records: Document why a theme is accepted or rejected, noting any relevant discussions or research. This would help clearly explain decisions if asked.

Monitor the Impact: Track the effects of using the tool in consultation processes.

Keep Human Oversight: Use human oversight to ensure the tool is processing consultation responses appropriately. This could include having an analyst assign themes for a sample of the responses to see how they compare with the tool.

Appendix - Detailed overview of demographics

Age distribution

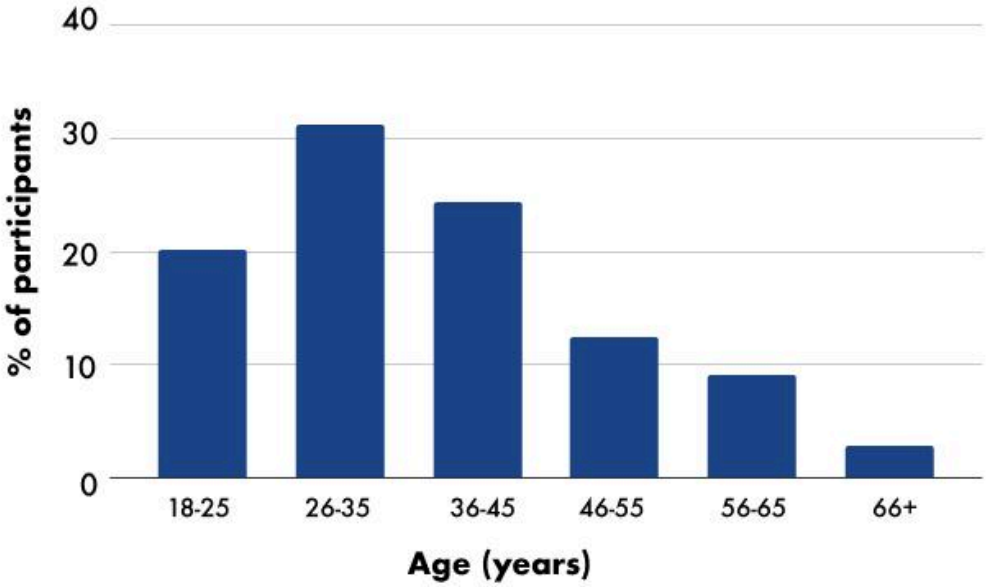


Figure 10: Participants' age distribution

Gender distribution

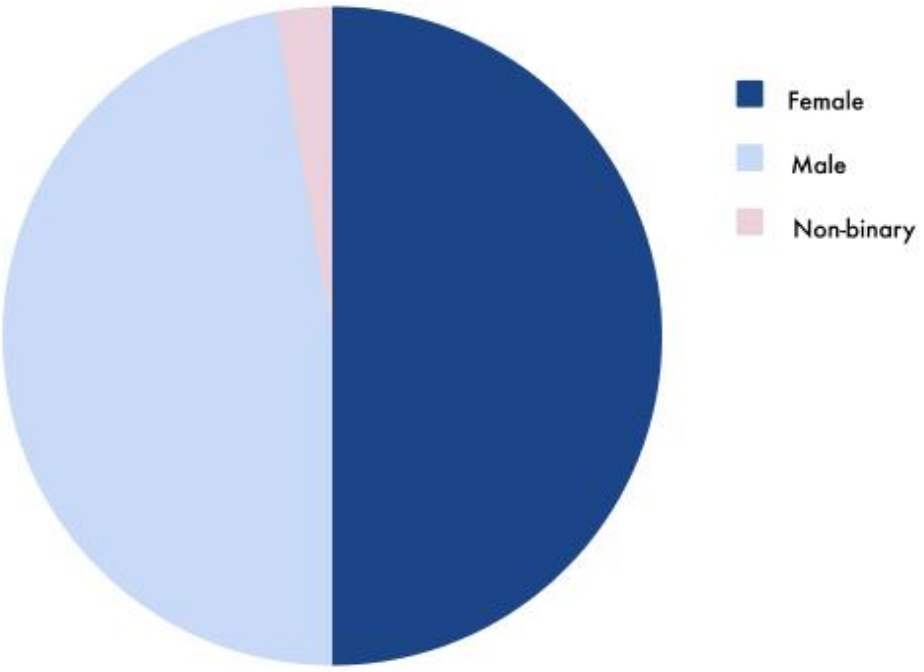


Figure 11: Participants' gender distribution

Disability distribution

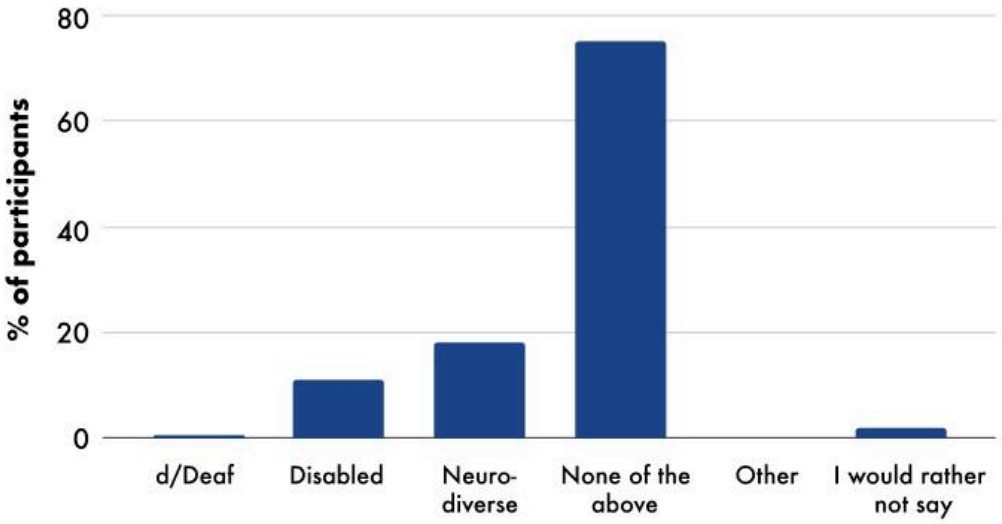


Figure 12: Participants' disability distribution

We asked: Which of the following descriptions do you identify with? (Select all that apply).
Percentages total more than 100% because people could identify with more than one description.

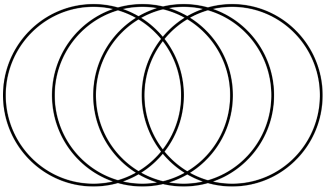
Table 1: Overview of participants by ethnicity

Ethnicity	% of participants
African	10
Any other Asian background	2
Any other Black, Black British, or Caribbean background	3
Any other ethnic group	1
Any other Mixed or multiple ethnic background	1
Any other White background	10
Bangladeshi	1
Caribbean	4
Chinese	1

English, Welsh, Scottish, Northern Irish or British	56
Indian	5
Irish	1
Pakistani	2
Roma	1
White and Black Caribbean	1

Table 2: Overview of participants by UK region

UK region	% of participants
East Midlands	3
East of England	9
Greater London	21
North East	25
North West	19
Northern Ireland	2
Scotland	3
South East	3
South West	3
Unknown Region	2
Wales	3
West Midlands	6



Centre for
Collective Intelligence
Part of the nesta group

For more information, please contact:
collective.intelligence@nesta.org.uk

Published July 2025