

AI chatbots in public services

12.12.23



Got a health worry?
Talk to Aiden.



20:08

Hello! I'm a UK government chatbot assistant here to provide health advice regarding constipation in children. I'm sorry to hear you have a worry about your little one. Let me know more about the situation and I can help.



Got a health worry?
Talk to Aiden.

Try our new AI powered chatbot

Ask a question...



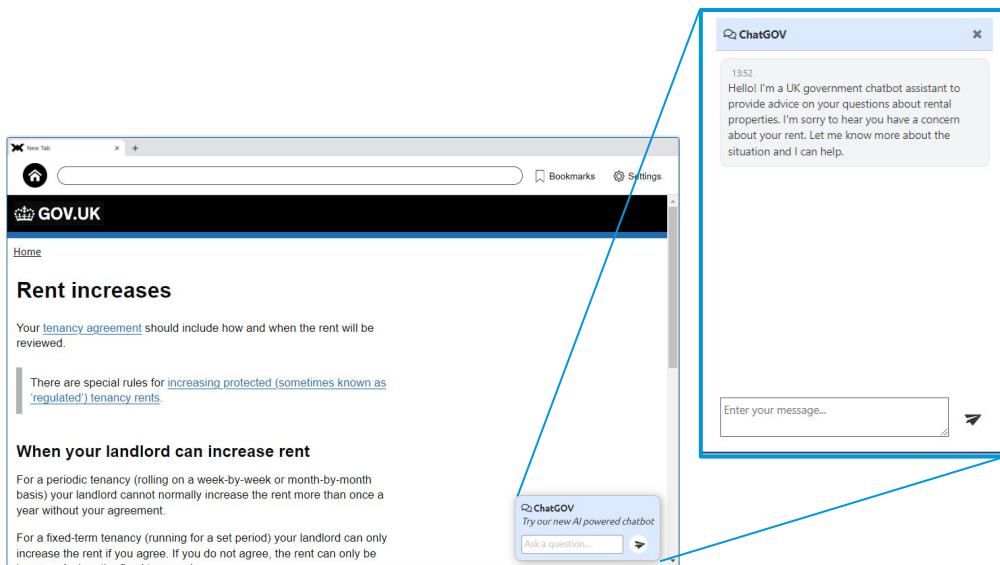
Enter your message...



We asked people to find information on government websites, with or without a chatbot

Your landlord has just told you they're putting your rent up by 35%. You're in a 1 year fixed term tenancy and you've only lived in the house for 3 months. You've come to this government website to get advice on whether that's legal.

Please use the website below and the information it contains to answer the questions that follow.



The image shows a screenshot of the GOV.UK website. The main content area is titled "Rent increases" and contains text about tenancy agreements and special rules for regulated tenancy rents. A "ChatGOV" chatbot window is open in the bottom right corner, displaying a message from the chatbot: "Hello! I'm a UK government chatbot assistant to provide advice on your questions about rental properties. I'm sorry to hear you have a concern about your rent. Let me know more about the situation and I can help." Below the message is an input field labeled "Enter your message..." and a send button. A blue box highlights the chatbot window, and a blue line connects it to a larger, detailed view of the chatbot interface on the right.

Does your landlord need your agreement to put the rent up now?

- No
- Yes
- I'm not sure

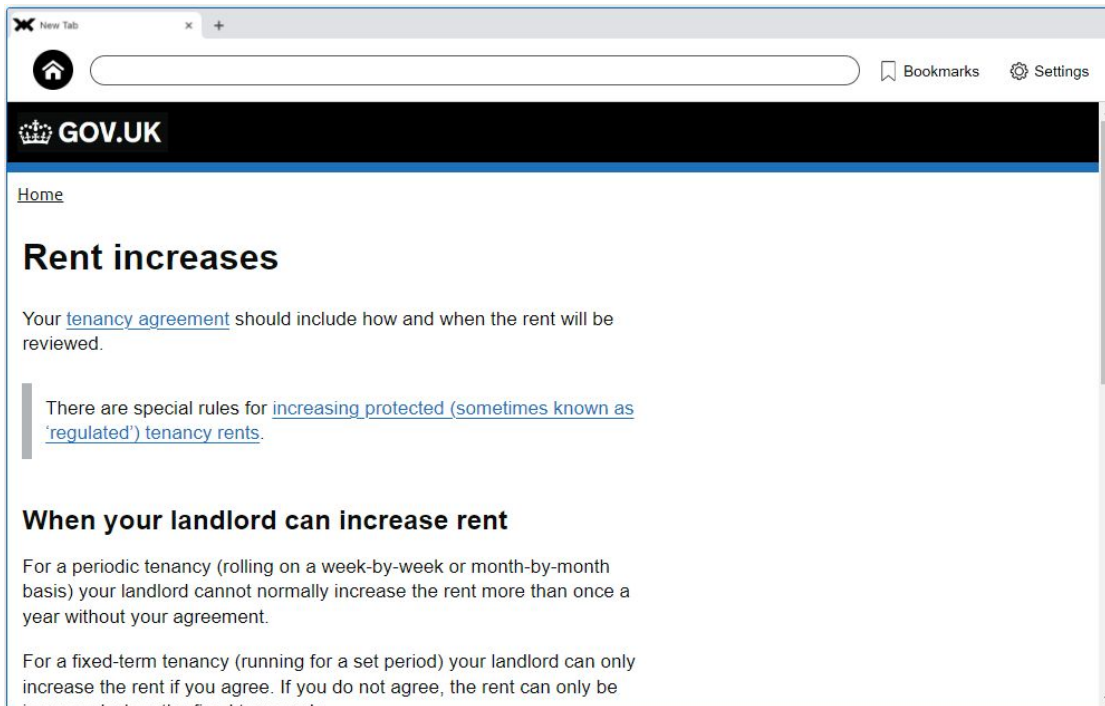
If you don't agree to the rent increase, when can the landlord legally increase the rent?

- Now, if they provide a written notice.
- In 6 months
- Only after the fixed term has ended.
- In 1 month

What should you do next?:

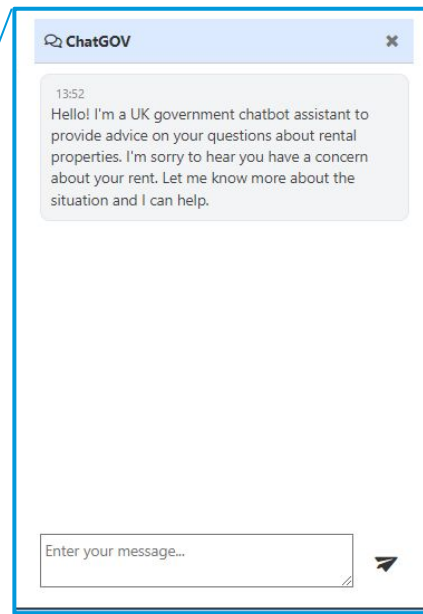
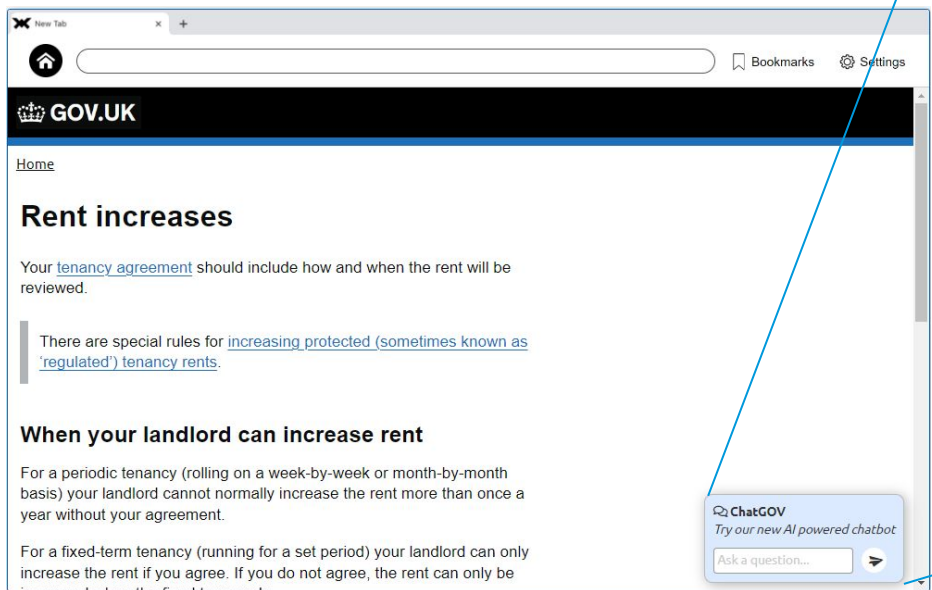
- Accept the increased rent
- Ignore the notice
- Tell the landlord you don't agree to the rent increase, and that they can't increase it until the end of the term
- Move out immediately without giving notice

Control - No bot



N = 1,090

Treatment 1 - Basic bot



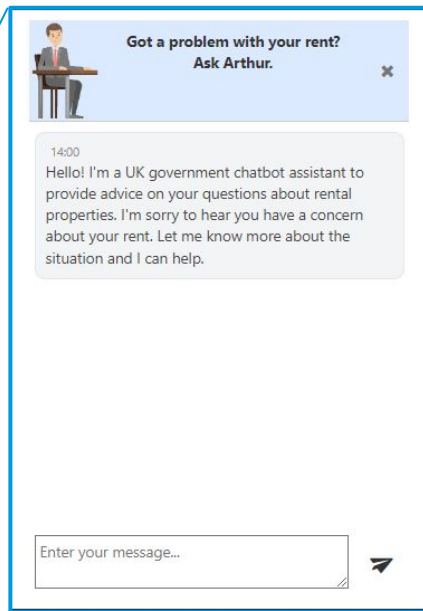
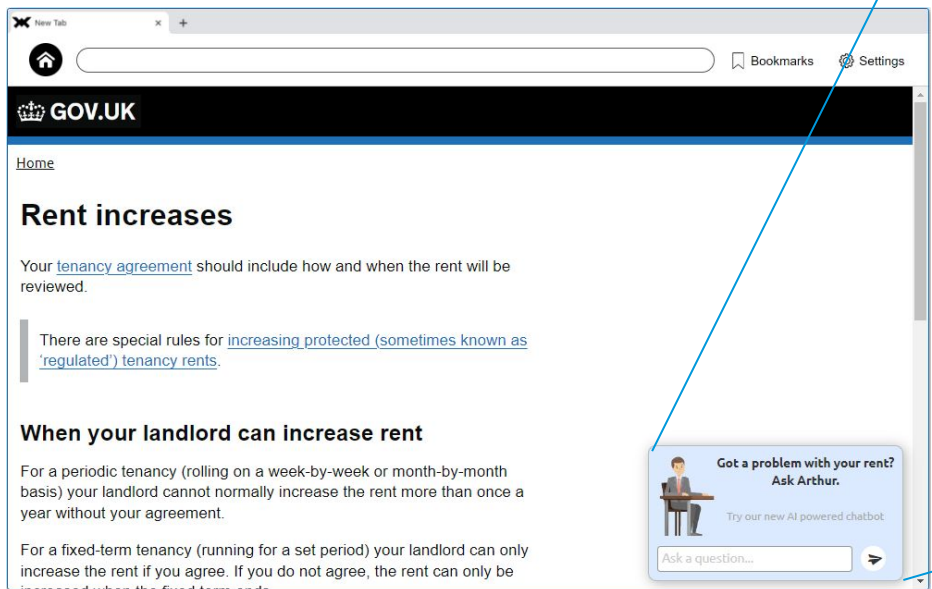
N = 1,026



15 seconds

If the user didn't engage with the chatbot after 15s on the page, it would automatically expand to prompt engagement

Treatment 2 - Cartoon bot

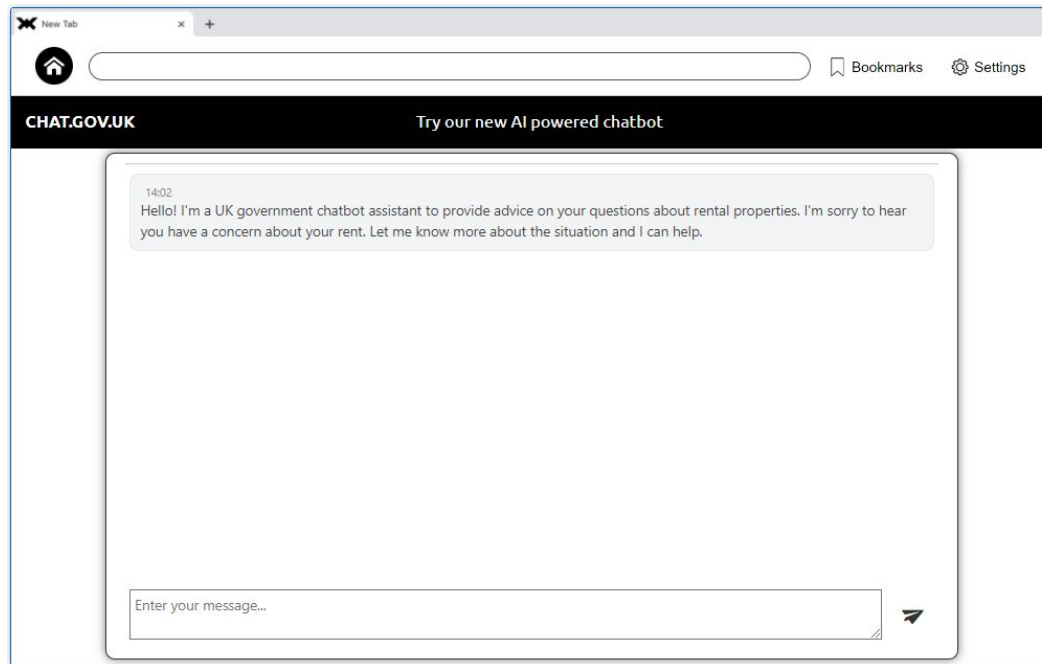


N = 1,014



15 seconds

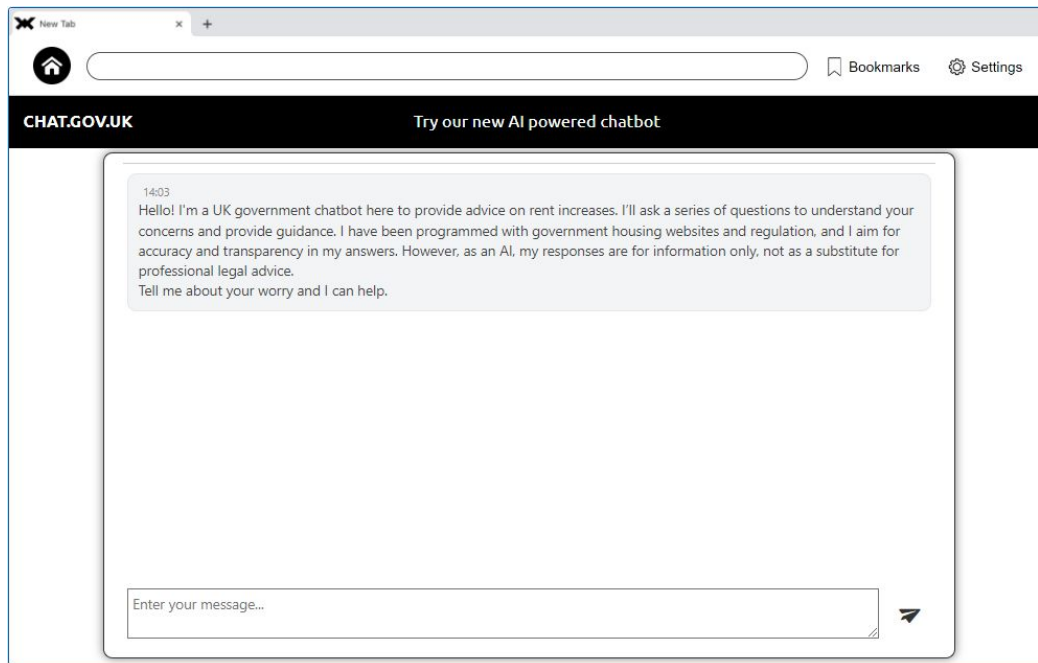
Treatment 3 - Whole page bot



[Click here to see the related gov.uk page about rent](#)

N = 975

Treatment 4 - Whole page 'transparent bot' (flagging AI risks)

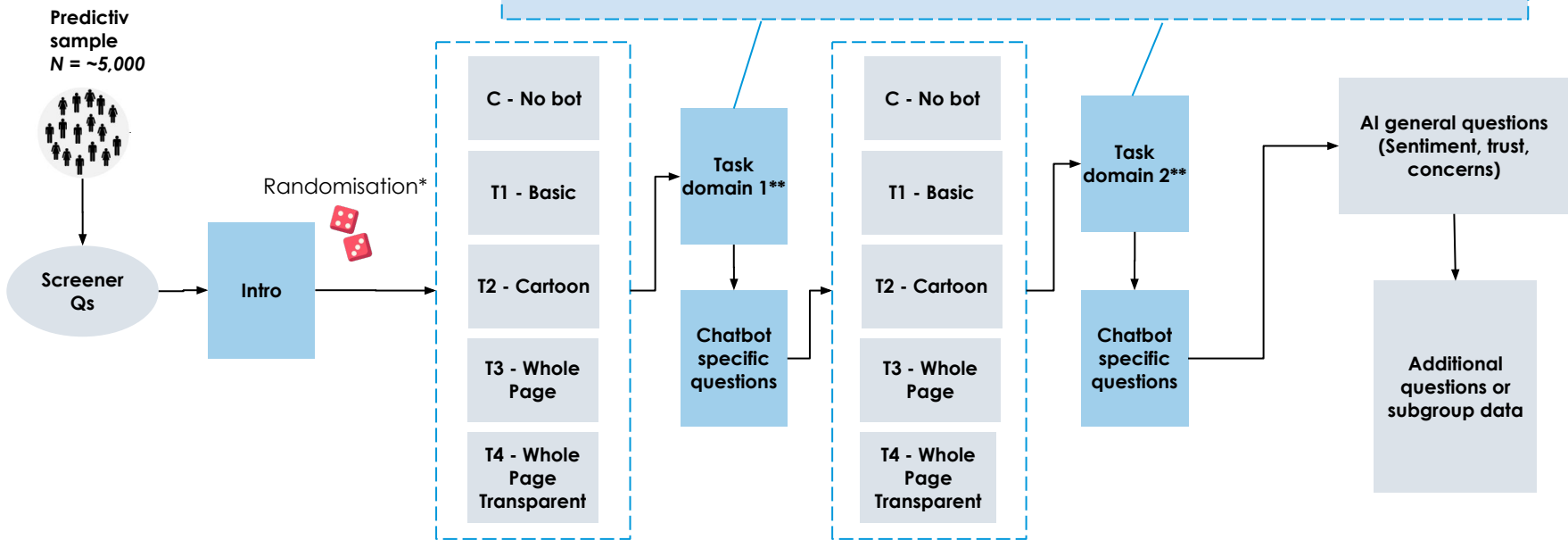


[Click here to see the related gov.uk page about rent](#) 

N = 914

User journey

Experimental flow



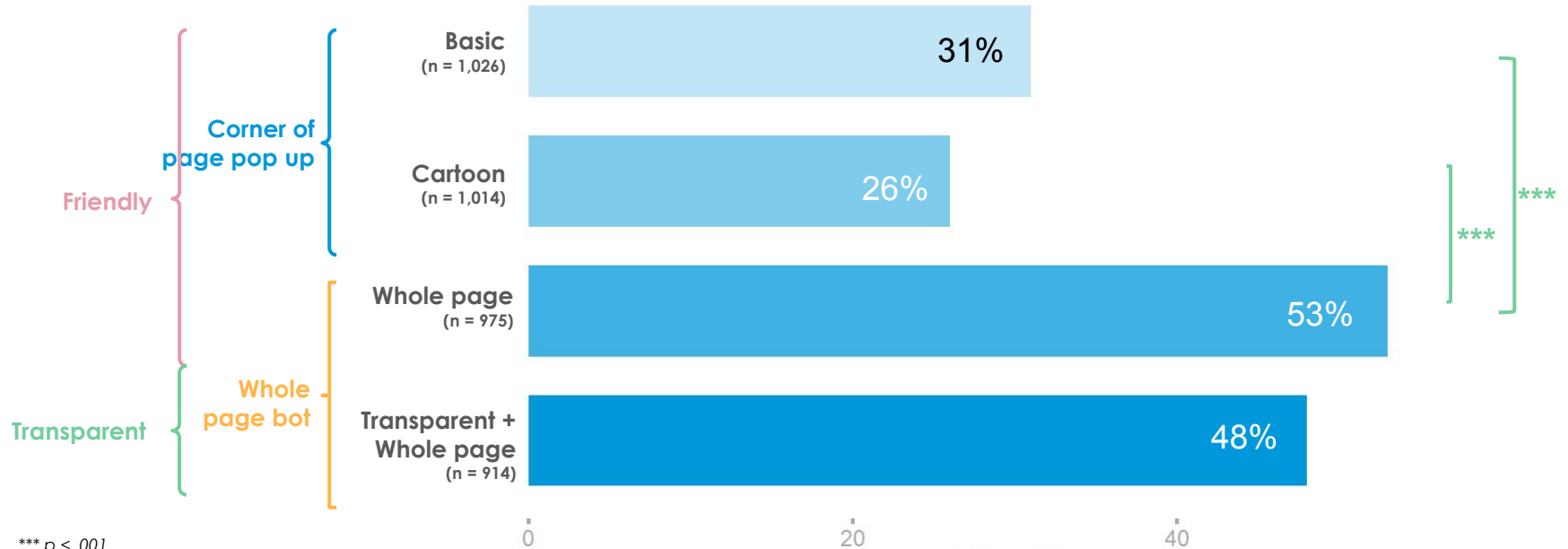
*Randomisation occurs before respondents see the first task but respondents will see the same treatment throughout

**Task order will be assigned randomly to account for ordering effects

Key Findings: Engagement

40% of those who saw a bot chose to message it.

Whole page bots had higher rates of engagement at around 50%.



*** p < .001

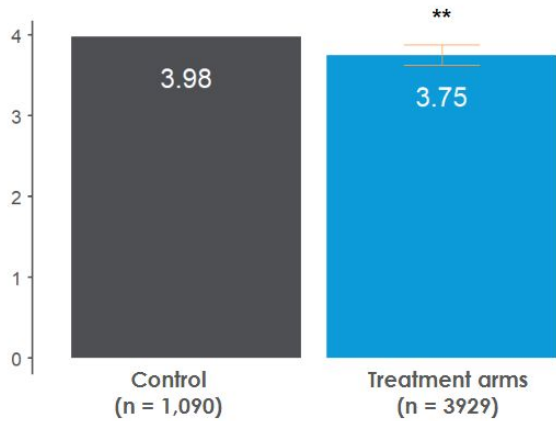
Analysis controlling for age, gender, location, ethnicity, education, urbanicity, income, and order of tasks.

Significance stars indicate difference from best performer (53%)

Data collected by BIT from 6th - 28th Nov 2023

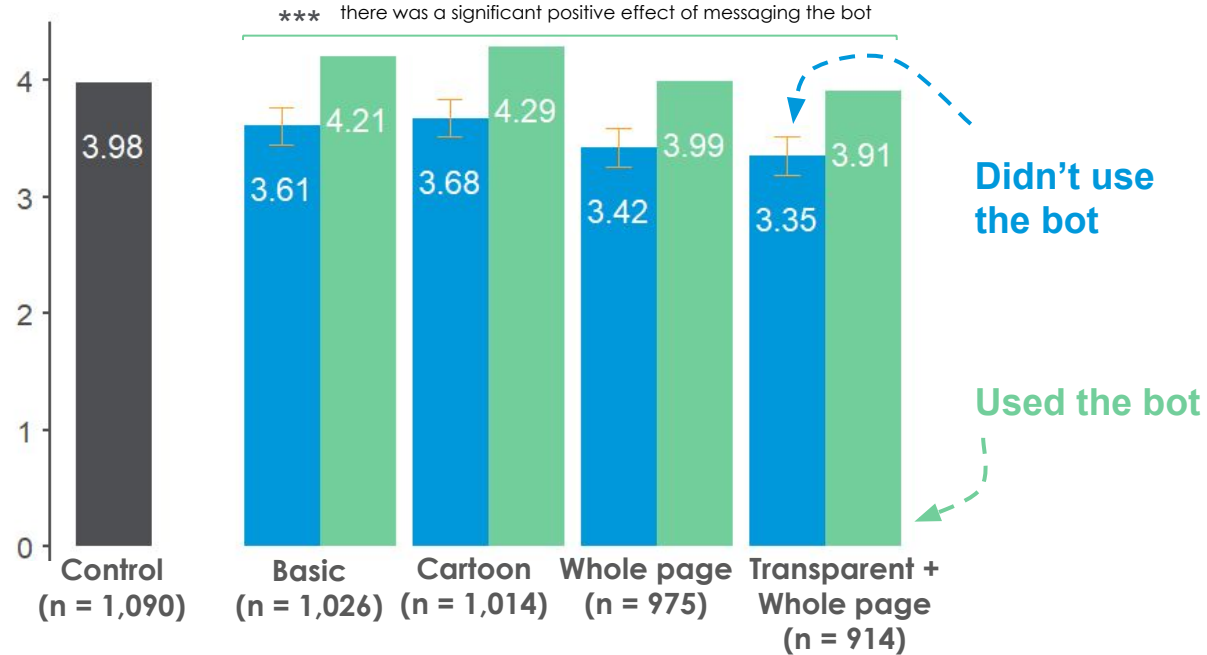
Key Findings - Task Performance

Overall, accuracy was lower when participants had access to a bot, but higher for those that actually used the bot.



Participants were given 3 multiple choice questions in each of the health task and rental task. The answers could be found in the webpage or by asking the chatbot.

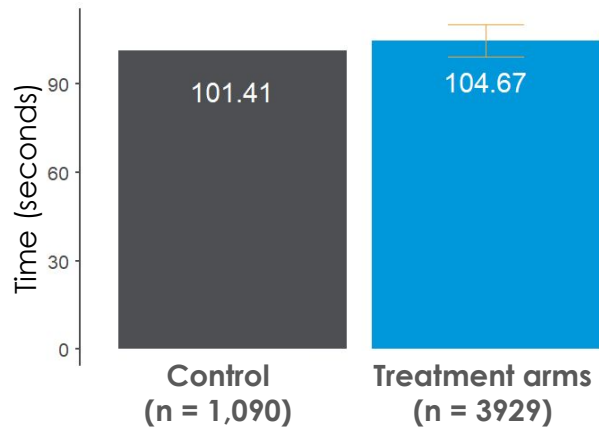
Accuracy is the sum of these scores (scale from 0 to 6)



The fact that bot users are more accurate, but overall those that saw the bot were less accurate, may be due to benefits of bot use, but could also be due to self-selection; the types of people who used the chatbot may have been more accurate even without the bot (for example, very engaged participants).

Key Findings - Task Performance

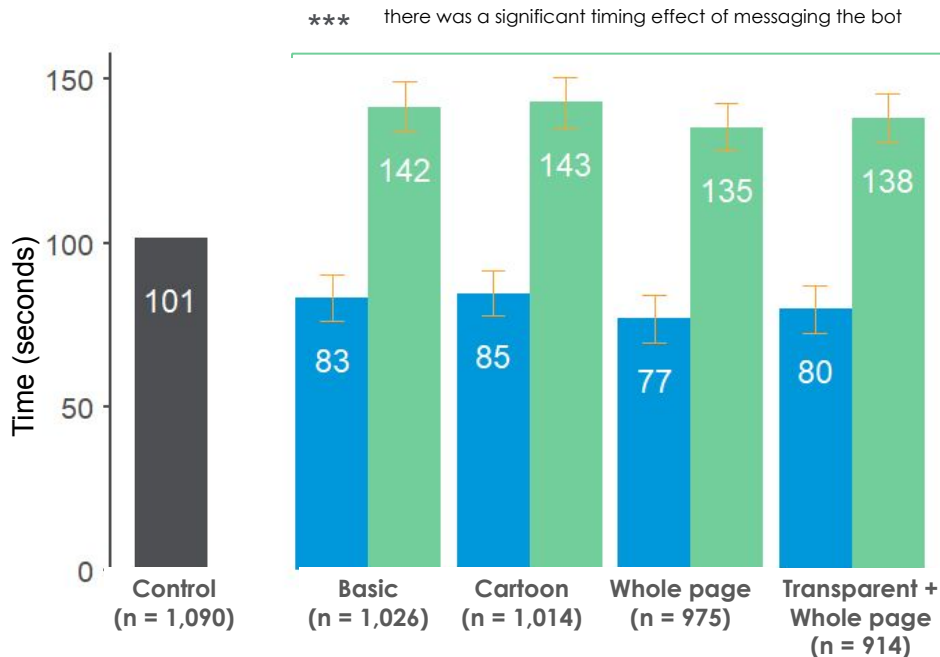
Overall, the interventions made no difference to the amount of time taken, however those who chose to use a bot were *slower* than those who did not.



** $p < .01$; *** $p < .001$

Analysis controlling for age, gender, location, ethnicity, education, urbanicity, income, and order of tasks.

Data collected by BIT from 6th - 28th Nov 2023



As with accuracy, this finding may also be due to self-selection, as those who chose to use the bot may have taken longer anyway (due to being more conscientious or more engaged).

Key Findings - Task Performance

Participants with access to a bot found the task easier, but thought the information was less trustworthy.

% who say they...	Control (n = 1,090)	Basic (n = 1,026)	Cartoon (n = 1,014)	Basic whole page (n = 975)	Transparent Whole page (n = 914)
... found the information easy to understand	90%	92%	91%	91%	90%
... found the information helpful	92%	91%	91%	91%	90%
... found the information friendly	87%	89%	88%	87%	88%
... found the information trustworthy	92%	92%	92%	90%	89%
... are confident in the accuracy of their answers to the questions	78%	79%	77%	76%	75%
... did not find the experience challenging	62%	67%	64%	69%	68%

Green shading indicates significantly higher than control ($p < .05$)
Red shading indicated significantly lower than control ($p < .05$)

Data collected by BIT from 6th - 28th Nov
2023

Key Findings - Task Performance

Exposure to the bot led to considerably higher support for government use of AI on similar tasks

% who say they...	Control (n = 1,090)	Basic (n = 1,026)	Cartoon (n = 1,014)	Basic whole page (n = 975)	Transparent Whole page (n = 914)
... found the information easy to understand	90%	92%	91%	91%	90%
... found the information helpful	92%	91%	91%	91%	90%
... found the information friendly	87%	89%	88%	87%	88%
... found the information trustworthy	92%	92%	92%	90%	89%
... are confident in the accuracy of their answers to the questions	78%	79%	77%	76%	75%
... did not find the experience challenging	62%	67%	64%	69%	68%
... think the government should use AI tools to help provide information on problems like those in the tasks	63%	75%	72%	77%	77%

Green shading indicates significantly higher than control ($p < .05$)
 Red shading indicated significantly lower than control ($p < .05$)

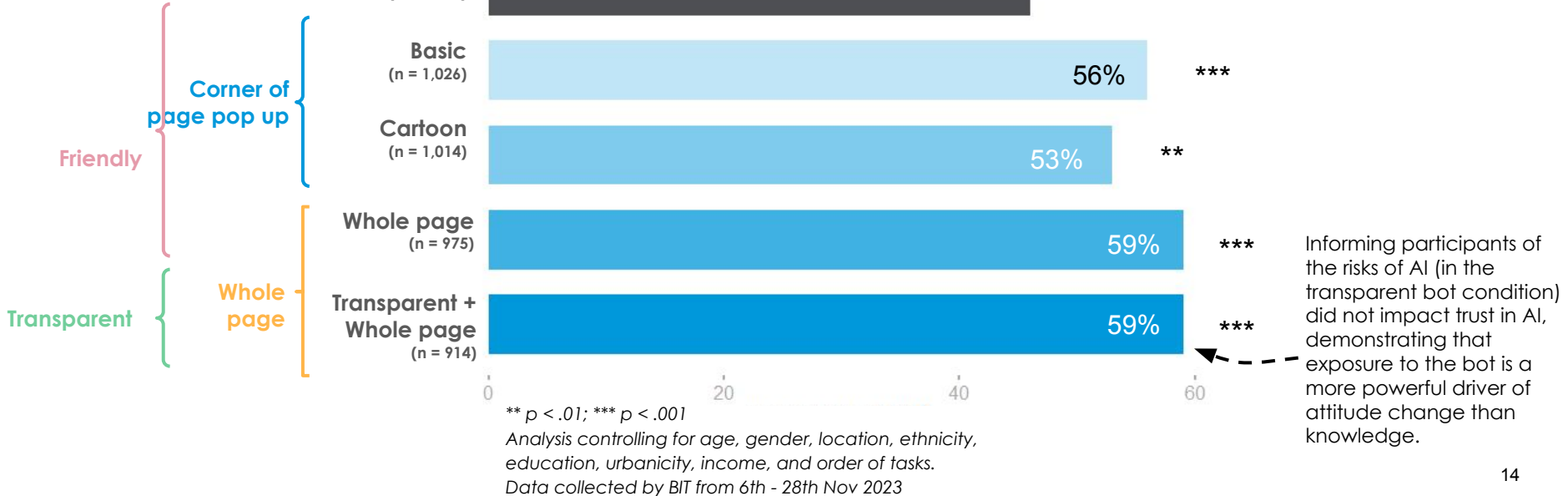
Data collected by BIT from 6th - 28th Nov 2023

Key Findings I

All bot interventions increased trust in AI, with the whole page bots increasing trust by 13pp. Mentioning the risks of AI in the transparent bot design did not impact trust in AI.

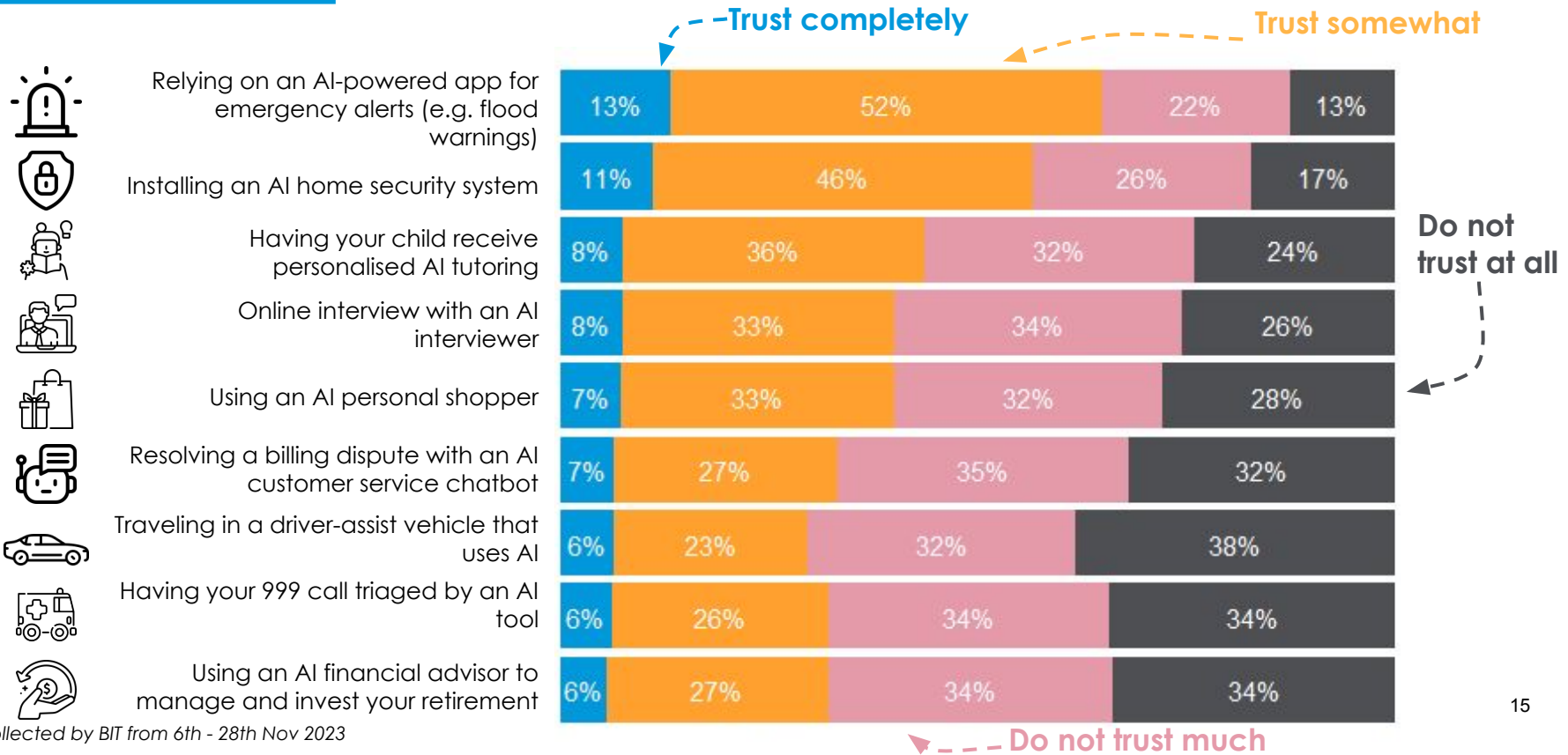
“To what extent do you agree that AI is trustworthy?”

The percentage of people who rated the bot as trustworthy was lowest in the control group, implying that exposure to AI can increase trust.



Key Findings: Trust

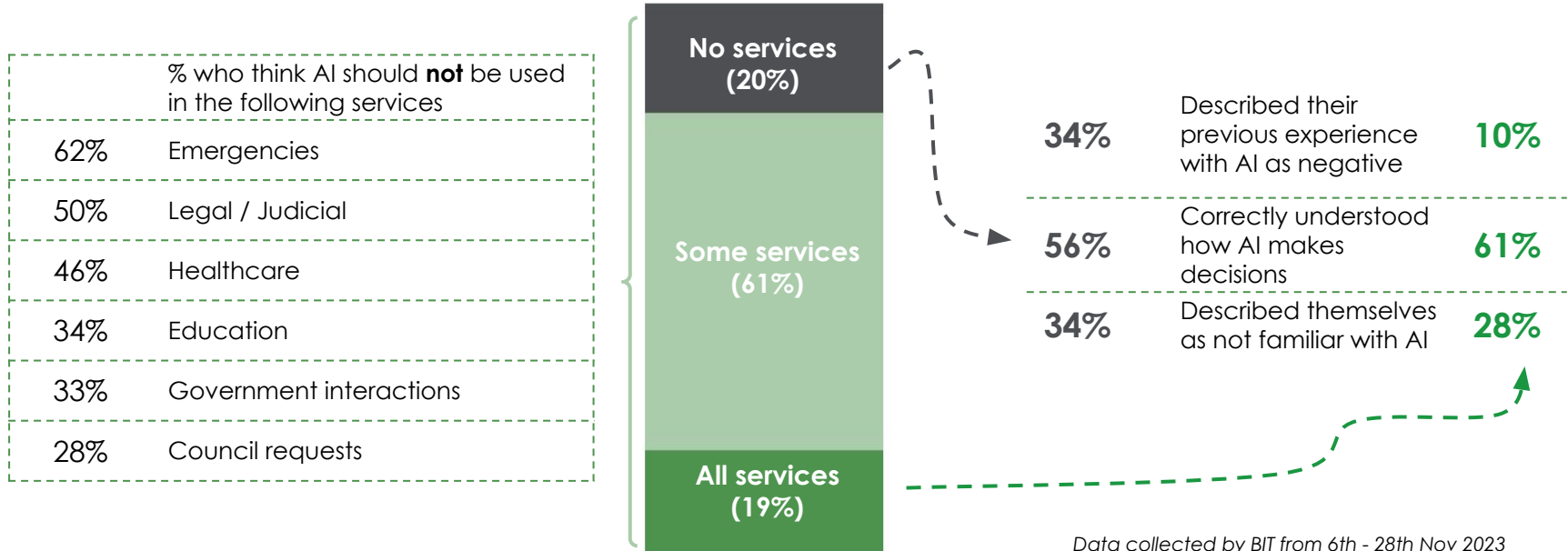
People were most likely to trust AI being used for emergency alerts, and least like to trust AI for financial advice, in a healthcare setting, or in transportation.



Key Findings: Trust

8 in 10 think AI could be used in government services. However, at least 1 in 2 thought AI should not be applied to emergency services or legal services.

People who thought AI chatbots could be used in ...
(n = 5,019)



Recommendations

- **Expanding the use of simple AI tools in government services is likely to help build trust and familiarity with AI.** Council requests, low stakes government interactions, and education are where the public is most supportive of use of AI at this stage.
- **There is reasonably high appetite for the use of AI-powered chatbots to navigate government information.** Uptake of the bot was moderate (40%), but participants reported finding the task easier, and were subsequently highly supportive of AI being used to support with similar tasks.
- **Chatbots are likely to be more useful for navigating more complex challenges or gov.uk as a whole than rephrasing existing content.** We expect bigger benefits of chatbots when users need help finding information across all of gov.uk or nhs.uk websites, warranting further research.
- **Using experiments to test comprehension and trust should form a core part of beta testing for new AI tools.** Even if information provided by a bot is accurate, it can still decrease user comprehension. Online experiments like these are a cheap and quick way to test comprehension at scale, and ensure that chatbots do not harm user outcomes.