Stanford | Deliberative Democracy Lab
Center on Democracy, Development and the Rule of Law

Stanford Crowdsourced Democracy Team

Meta Community Forum

Results Analysis

April 2024

The Deliberative Democracy Lab (DDL) at Stanford University is housed within the Center on Democracy, Development, and the Rule of Law at the Freeman Spogli Institute for International Studies

The Crowdsourced Democracy Team (CDT) at Stanford University is housed within the Department of Management Science and Engineering

Report Contributors (in alphabetical order): Samuel Chang, Estelle Ciesla, Michael Finch, James Fishkin, Lodewijk Gelauff, Ashish Goel, Ricky Hernandez Marquez, Shoaib Mohammed, and Alice Siu

Corresponding Author: Alice Siu
asiu@stanford.edu

Contents

Introduction 1

Design of the Community Forum 4

Methods and Recruitment 6

Participants at a Glance 10

Attitudes and Use of Technology 12

Participants' use of AI chatbots 13
Participants' use of technology 14
Participants' opinions on AI 15

Proposals 20

Should AI chatbots inform users that they are interacting with a bot? 21
Which sources should AI chatbots draw information from? 24
Which perspectives should AI sources draw information from? 27
Should AI chatbots use the user's past conversations to improve user experience? 30
Should AI chatbots use the user's online activity to personalize interactions? 35
Should AI chatbots prioritize predictability or unpredictability? 39
Should AI chatbots be allowed to discuss topics some people might find offensive? 43
Should AI chatbots be allowed to discuss topics outside of their intended use? 47
Should AI chatbots be designed to be humanlike? 51
Should AI chatbots be able to use emotional cues and conversational tactics? 56
Should users be allowed to use AI chatbots for romantic relationships? 60
Should users be allowed to leverage AI chatbots to enable romantic relationships with other humans? 64

Knowledge Gains 68

Event Evaluations 71

Demographic Differences 72

Gender 72
Age 74
Education 77
Usage of AI chatbots 80

Control vs. Participant Groups: Analysis of Attitudinal Representativeness 82

Control vs. Participant Group Demographics Before Deliberations 83
Attitudinal Representativeness 90
Comparison of Changes in the Participant and Control Groups: Difference in Differences 101

Conclusion, Limitations, and Implications 107

Appendix 109

Changes in Control Group Proposal Ratings Before and After Deliberations 110
Changes in Treatment Group Proposal Ratings Before and After Deliberations 116

Introduction

In October 2023, Meta, in collaboration with the Stanford Deliberative Democracy Lab and the Behavioral Insights Team, implemented the second Meta Community Forum. This Community Forum (CF) focused on 'what principles should guide generative AI's engagement with users?'. As AI chatbots become more powerful, platforms and users face crucial questions about the behavior of AI chatbots. For example, how human should AI chatbots be? What are users' preferences when interacting with AI chatbots? And, which human traits should be off-limits for AI chatbots? Furthermore, for some users, part of the appeal of AI chatbots lies in its unpredictability or sometimes risky responses. But how much is too much? Should AI chatbots prioritize originality or predictability to avoid offense? These are just some of the questions that participants deliberated on over the course of a weekend discussion.

1545 participants from four countries - Brazil, Germany, Spain, and the United States, participated in this deliberative event. Even though this project involved four countries, the samples were recruited independently, so this Community Forum should be seen as four independent deliberations. In addition, 1108 persons participated in the control group, where the participants did not deliberate in any discussions; the control group only completed two surveys. The main purpose of the control group is to demonstrate that any changes that occur after deliberation are a result of the deliberative event.

Key Takeaways:

- **CF participants were highly representative of the four countries' populations.** As such, the results are generalizable to the general population of each country.
- **A majority of participants from each country thought AI had a positive impact.** After deliberating, more participants in all countries started seeing AI positively.
- **Users and non-users of AI Chatbots generally had similar views and moved in the same directions.** About 60% of participants reported having used ChatGPT prior to the deliberative event. If there were differences between users and non-users, non-users were generally more cautious or conservative on the policy proposals. For example, non-users were much less supportive of AI chatbots being offensive. However, any gaps between users and non-users generally narrowed post-deliberations as both users and non-users moved away from their original positions and towards each other.
- **CF participants rated the deliberative event highly.** Participants from all countries rated the event 8 out of 10 and above; noting the event was valuable to them.

Important Note from Stanford Deliberative Democracy Lab about this report:

There are a lot of exciting results in this report. DDL would like to emphasize two points as you explore the results of this Community Forum.

1. There were some statistically significant and substantial opinion changes after deliberations. The report highlights and discusses these changes. The latter qualitative section begins to shed some light on why some of the participants changed their opinions. In subsequent reports, DDL will provide deeper level quantitative and qualitative analyses into the reasons why participants shifted their opinions.
2. There were opinions that did not shift after deliberations, these are participants' considered judgments at the end of the event, regardless of change. Participants held onto these initial opinions even after they were subjected to all the counter arguments; therefore, these considered opinions are also noteworthy.

Lastly, for results presented with asterisks (*), these asterisks indicate the results are statistically significant. The asterisks indicate the following: "*" indicates a p-value of less than 0.05, "**" indicates a p-value of less than 0.01, and "***" indicates a p-value of less than 0.001 resulting from a paired or independent t-test.

Design of the Community Forum

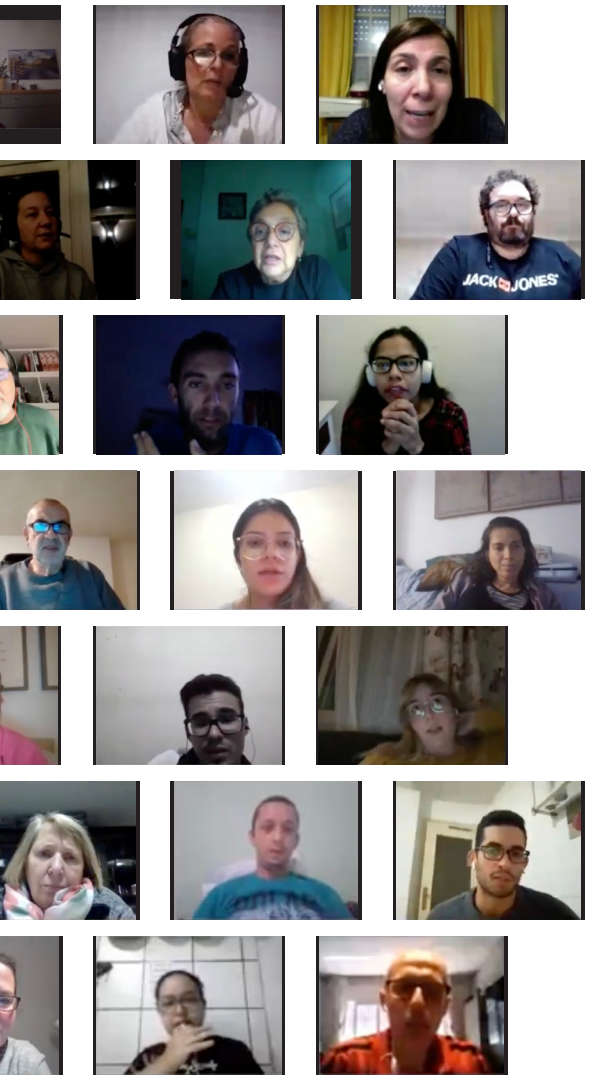
The design for the deliberations followed the Deliberative Polling® model under the direction of the Stanford Deliberative Democracy Lab and in collaboration with the Behavioral Insights Team (BIT) and Meta. A distinguished Advisory Committee vetted the briefing materials for the deliberations and provided many of the experts for the plenary sessions. The process alternated small group discussions and plenary sessions where experts would answer questions agreed on in the small groups. The agenda was a series of 38 policy proposals focused on the principles of AI's engagement with users. The policy proposals were presented with background materials and with pros and cons posing trade-offs that the participants might want to consider. Video versions of the briefing materials were also provided during the event.

The small group discussions were conducted on the Stanford Online Deliberation Platform which moderated the video based discussions, controlled the queue for talking, nudged those who had not volunteered to talk, intervened if there was incivility, and moved the group through the agenda of policy proposals and their pros and cons. Near the end of each discussion, it also guided the groups in formulating key questions that they wished to pose to the panels of competing experts in the plenary sessions. The Stanford Online Deliberation Platform is a collaboration between the Crowdsourced Democracy Team, led by Ashish Goel, and the Deliberative Democracy Lab, led by James Fishkin, both at Stanford University.

The core issue posed for deliberation was "What principles AI should have as AI chatbots engage with users?". To what extent should AI chatbots offer a one-size-fits-all experience based on global norms, or tailor their responses to each user's unique background and situation? Furthermore, as AI chatbots mimic human interactions more convincingly, we face the challenge of defining how human-like they should be. This involves determining the appropriate level of transparency about their artificial nature, and identifying which aspects of human interaction are unsuitable for them to replicate. While some find charm in their

unpredictable, sometimes edgy responses, the question arises: should AI chatbots prioritize novelty over predictability and potential offensiveness? These are novel issues, and they amount to the beginnings of a social contract for how people engage with AI chatbots.

Methods and Recruitment



For this Community Forum, the polling firm, YouGov, was commissioned to recruit nationally representative samples of the general public (18+) in Brazil, Germany, Spain and the United States. YouGov is a global public opinion company that operates online panels in all regions of the world. In many countries their panels have millions of panelists, which allow for greater ability to access a representative population sample. YouGov panels are opt-in and use proprietary weighting algorithms to ensure accurate representation. Stanford's Deliberative Democracy Lab has engaged with YouGov since the late 2000s for Deliberative Polling recruitment and has since worked together on over a dozen projects, including the first Community Forum on Metaverse governance.

The four countries for this Community Forum were selected by Meta. Even though this project involved four countries, the samples were recruited independently, so this Community Forum should be seen as four independent deliberations. The intended sample size for each country was 200 participants in the Community Forum deliberations. This CF saw an overwhelming response for participation in all four countries; as total participation was 1545 participants, close to double our intended sample size. This was true for the Community Forum deliberation participants and for the control group, those that did not deliberate.

The following summarizes the total respondents and response rates for each of the four countries.

Brazil

YouGov interviewed 4,003 respondents in Brazil as part of a baseline survey. All respondents were asked whether they were interested in participating in a series of online deliberative discussions and whether they consented to share their emails with the Stanford Deliberative Democracy Lab in order to be registered on the Stanford Online Deliberation Platform. 1,533 respondents to the baseline survey expressed interest. Of these, 1,061 were assigned to a 'treatment' group and were subsequently invited to the CF, and 472 were assigned to a 'control' group (and were not invited to the CF).

Following the deliberative event on October 28th and 29th 2023, all 'treatment' respondents who attended the event as well as all 'control' respondents were invited to a post survey. 336 'treatment' respondents and 362 'control' respondents completed the post survey.

Germany

YouGov interviewed 4,272 respondents in Germany as part of a baseline survey. 1,296 respondents to the baseline survey expressed interest in the CF. Of these, 1,005 were assigned to a 'treatment' group and were subsequently invited to the event, and 291 were assigned to a 'control' group (and were not invited to the CF). Following the deliberative event, 399 'treatment' respondents and 236 'control' respondents completed the post survey.

Spain

YouGov interviewed 3,897 respondents in Spain as part of a baseline survey. 1,290 respondents to the baseline survey expressed interest in the CF. Of these, 940 were assigned to a 'treatment' group and were subsequently invited to the CF, and 350 were assigned to a 'control' group (and were not invited to the CF). Following the deliberative event on October 28th and 29th 2023, 413 'treatment' respondents and 272 'control' respondents completed the post survey.

United States

YouGov interviewed 4,082 respondents in the United States as part of a baseline survey. 1,357 respondents to the baseline survey expressed interest in the CF. Of these, 968 were assigned to a 'treatment' group and were subsequently invited to the event, and 389 were assigned to a 'control' group (and were not invited to the CF). Following the deliberative event, 393 'treatment' respondents and 238 'control' respondents completed the post survey.

The rate of participation from the initial pre-survey, regardless of interest in participation in this deliberative event, was 9.4% across the four countries. The lowest rate was from Brazil at 8.4% and the highest rate was from Spain at 10.4%. Among the participants that expressed interest in the deliberative event and were invited to participate, the turnout rate was 38.8% across the four countries. The highest turnout rate was Spain at 43.9%, followed by the US at 40.6%, followed by Germany at 39.7% and then Brazil at 31.7%. The rate of participation for the control group was 73.8% across the four countries. The highest rate was from Germany at 81.1%, followed by Spain at 77.7%, a close third from Brazil at 76.7% and then the US at 61.2%.

	Brazil	Germany	Spain	United States	Total
Interviewed	4003	4272	3987	4082	**16344**
Expressed Interest	1533	1296	1290	1357	**5476**
Invited Participants	1061	1005	940	968	**3974**
Invited Control	472	291	350	389	**1502**
Final Participants	336	402	414	393	**1545**
Final Control	362	236	272	238	**1108**
Turnout rate participants	31.7%	40.0%	44.0%	40.6%	**38.9%**
Turnout rate control	76.7%	81.1%	77.7%	61.2%	**73.8%**

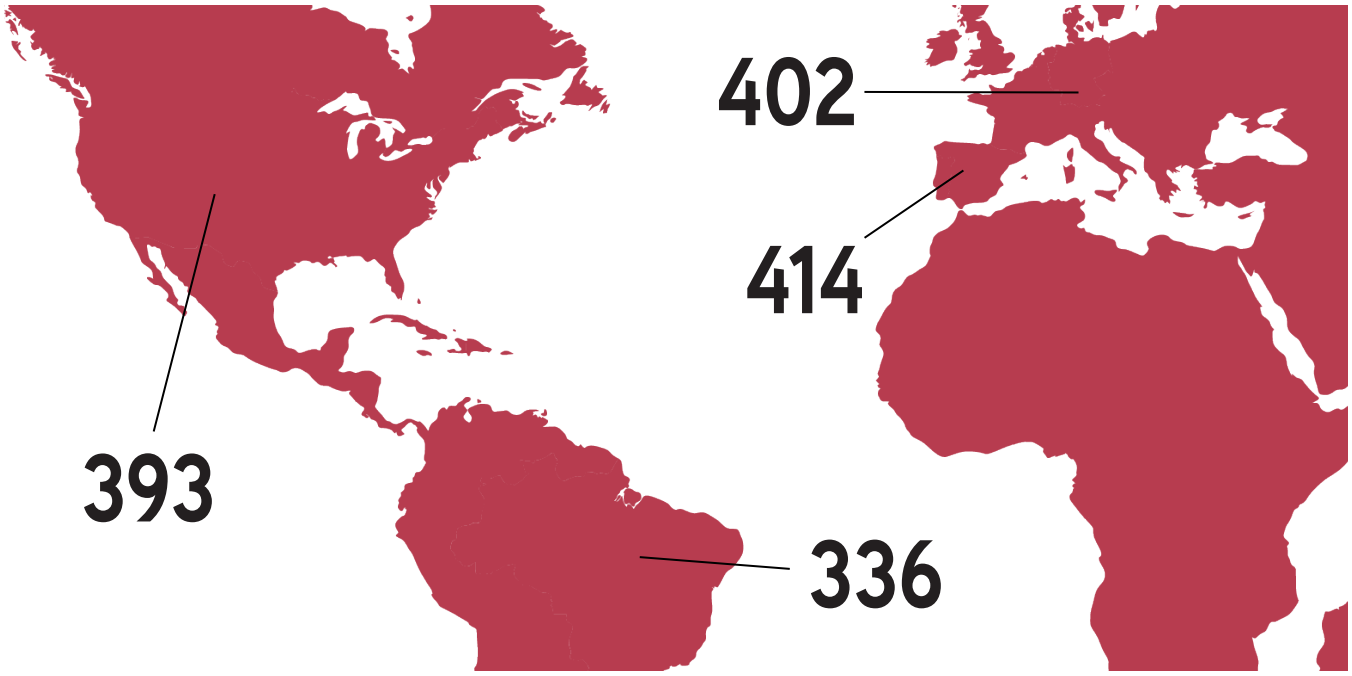
Weighting Methodology

The results in this report are weighted results, and therefore, the results are generalizable to the general population of the four countries.

YouGov weighted 'treatment' and 'control' cases separately in all four countries. The matched samples were weighted using Iterative Proportional Fitting (IPF), also known as "raking." The samples were raked on the joint distribution of age and gender, and the marginal distribution of education, to target distributions drawn from the sampling frames. The raking function iteratively adjusts the sample proportions to match each of the target distributions separately, iterating until the sample proportions converge on all target distributions. The resulting sample adjustments form the weights.

For example, in the US, the sampling frames were constructed by stratified sampling from the full 2019 American Community Survey (ACS) 1-year sample with selection within strata by weighted sampling with replacements (using the person weights on the public use file). The matched cases and the frame were combined and a logistic regression was estimated for inclusion in the frame. The propensity score function included age, gender, race/ethnicity, years of education, and region. The propensity scores were grouped into deciles of the estimated propensity score in the frame and post-stratified according to these deciles. The weights were then post-stratified on a four-way stratification of gender, age (4-categories), race (4-categories), and education (4-categories), to produce the final weight. The weights were then weighted by Iterative Proportional Fitting Procedure on gender, age, education and race to produce the final weight.

Number of pariticpants



Participants at a Glance

The participants of this Community Forum were **representative samples of the general population from four countries with a large number of Meta users** - Brazil, Germany, Spain, and the United States. Participants from each country deliberated separately in their respective languages - Portuguese, German, Spanish, and English.

Overall, **1545 participants** took part in this Community Forum:

- 336 were from Brazil
- 402 were from Germany
- 414 were from Spain
- 393 were from the United States

Gender

In terms of gender, 50.2% of participants were Female, 49.5% Male:

	Brazil	Germany	Spanish	USA
Male	50.6%	48.6%	51.3%	47.6%
Female	49.4%	52.1%	48.9%	50.4%
Non-binary				2%

Age

In terms of age, the breakdown was the following:

	Brazil	Germany	Spain	USA
Average	38.7 years	51.6 years	46.3 years	47 years
18-24	15.5%	7.0%	8.5%	10.2%
25-34	31.5%	12.9%	19.3%	20.4%
35-44	21.4%	18.2%	16.9%	16.5%
45-54	12.5%	10.0%	22.7%	15.3%
Over 54	19.0%	52.0%	32.6%	37.7%

Education

In terms of education, the breakdown was the following:

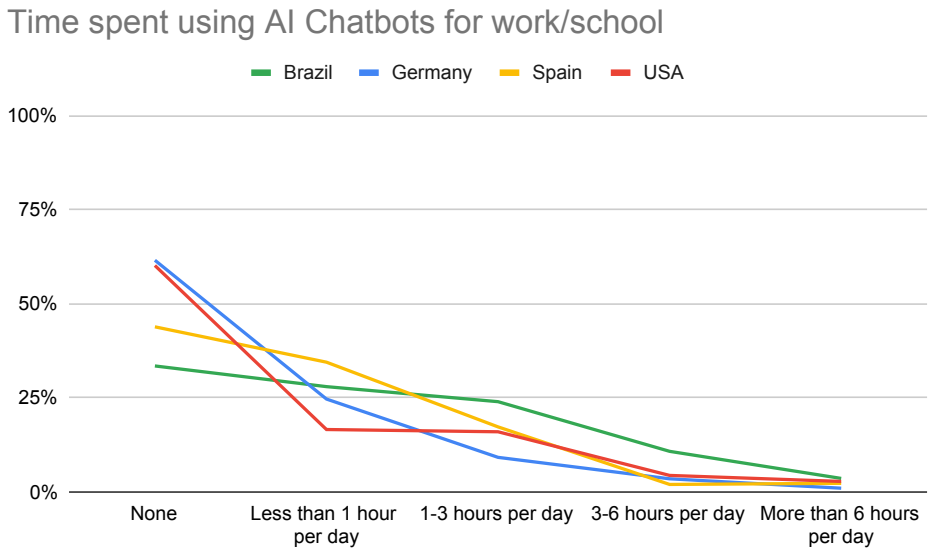
Germany was a special case in that we had no data about how many Germans completed any post-secondary education.

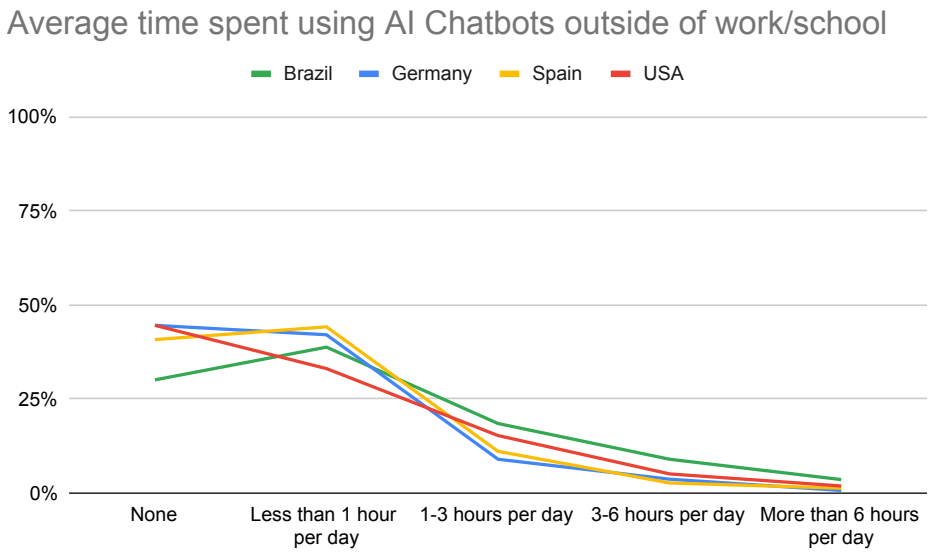
	Brazil	Germany	Spain	USA
Primary education or less	7.74%	N/A	0.98%	2.80%
Secondary education	47.32%	100% (Hauptschule: 6.44% Realschule: 28.09% Abitur: 65.46%)	24.32%	28.75%
Post-secondary non-tertiary or vocational education	22.02%	N/A	20.64%	N/A
Some tertiary education or higher	22.92%	N/A	54.05%	68.45%

Attitudes and Use of Technology

Participants' use of AI chatbots

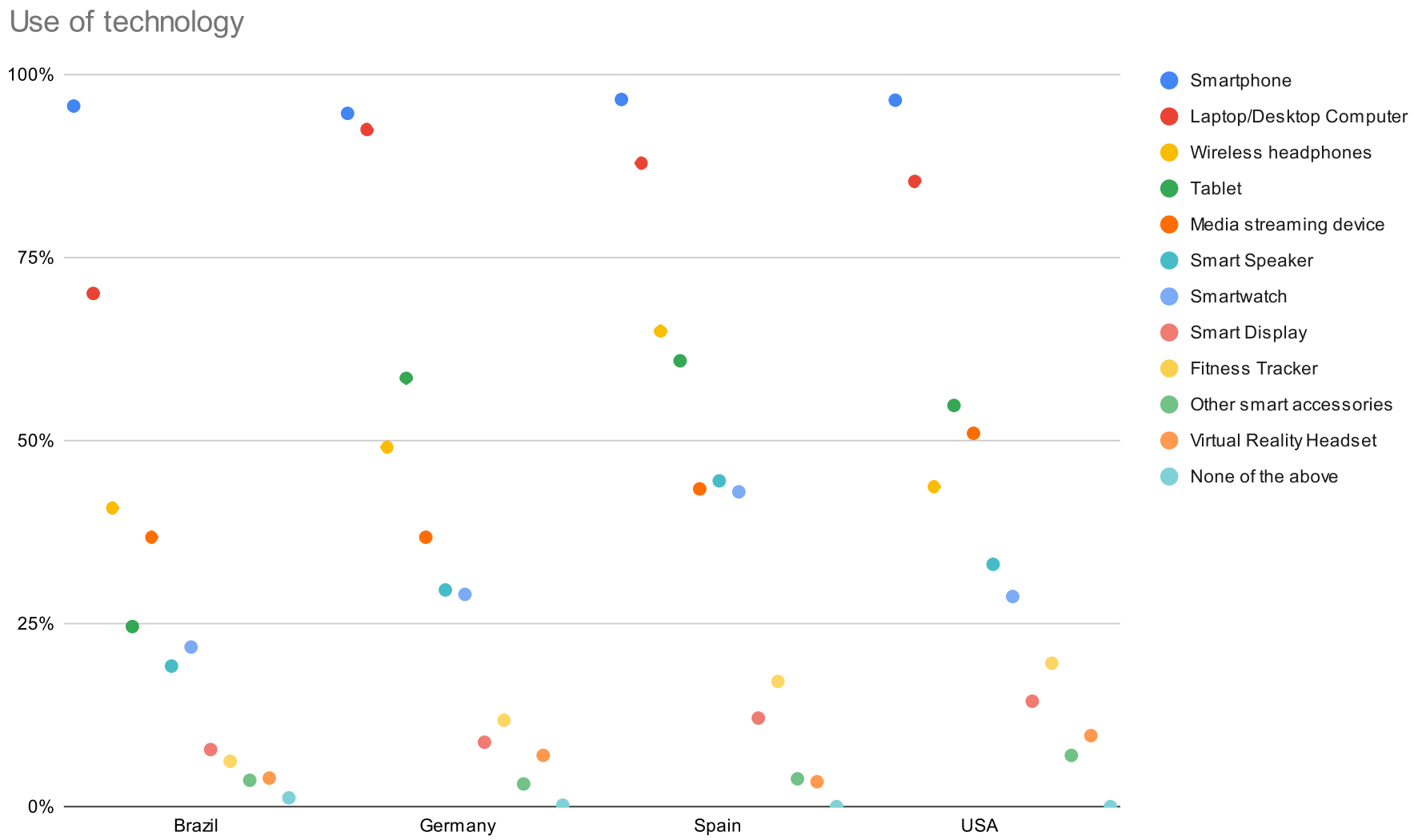
Participants from all four countries used AI chatbots outside of work/school related purposes more than for work/school related purposes. German participants used the least AI chatbots, and Brazilian ones the most. Overall, a majority of participants from all four countries used AI less than an hour daily for work/school and outside of work/school. The following measures are from participants' pre-deliberations survey.





Stanford Deliberative Democracy Lab
Center on Democracy, Development and the Rule of Law

Participants' use of technology



Participants' opinions on AI

A majority of participants from each country had used ChatGPT or similar chatbots. All four countries had statistically significant increases in participants who used ChatGPT or similar chatbots as a result of the deliberations.

Participants who had used ChatGPT or similar chatbots	Before deliberations	After deliberations	Change
Brazil	62.0%	74.3%	+12.3%***
Germany	61.7%	69.0%	+7.3%***
Spain	64.2%	73.6%	+9.4%***
United States	59.5%	71.3%	+11.8%***

A majority of participants from each country thought AI had a positive impact. After deliberating, more participants in all countries started seeing AI positively. Brazilian participants had the most positive views of AI, while participants from the United States were the least optimistic about AI's positive effects.

AI has a positive impact	Before deliberations	After deliberations	Change
Brazil	77.6%	81.7%	+4.1%*
Germany	56.4%	61.0%	+4.6%*
Spain	65.8%	74.6%	+8.8%***
United States	49.8%	54.4%	+4.6%*

Additionally, to measure their opinions on AI, participants were given a number of statements about AI. Here are the results:

Statement that received the most approval across all four countries - "AI chatbot's capability to increase efficiency by automating tasks is saving many companies a lot of time and resources."

- Over 70% approval from participants in Germany, Spain, and the United States, but only around 50% from participants from Brazil
- Increases in support for the statement in all four countries, ranging from +0.3% for Brazil to +9.8% for Spain. Increases led to over 80% of approval from participants in Germany and Spain.

Statement that received the least approval across all four countries - "People will feel less lonely with AI chatbots."

- Less than 50% approval across all four countries, except for Spain after deliberations.
- Least approval from Germany, with 31.3% of participants before deliberations and 38.9% after.
- Increases in support for the statement from all four countries, ranging from +0.1% for Brazil to +10.6% for Spain, rendering Spain the only country with majority support from its participants after deliberations.

Statement that gained the most approval as a result of the deliberations, across all four countries - "Chatbots replicate biases that exist in the data they were trained on."

- Over 11% approval was gained in Spain and the United States.
- 2.9% for Brazil and 3.9% for Germany.

Statement that lost the most approval as a result of the deliberations, across all four countries - "The increased use of AI Chatbots will lead to students losing their ability to think critically."

- Approval for this statement decreased in all countries, from a small decrease of -0.7% for the United States to -7.1% for Brazil.
- As such, participants from all four countries felt less inclined to agree that AI chatbots would lead to students losing their ability to think critically, after deliberations

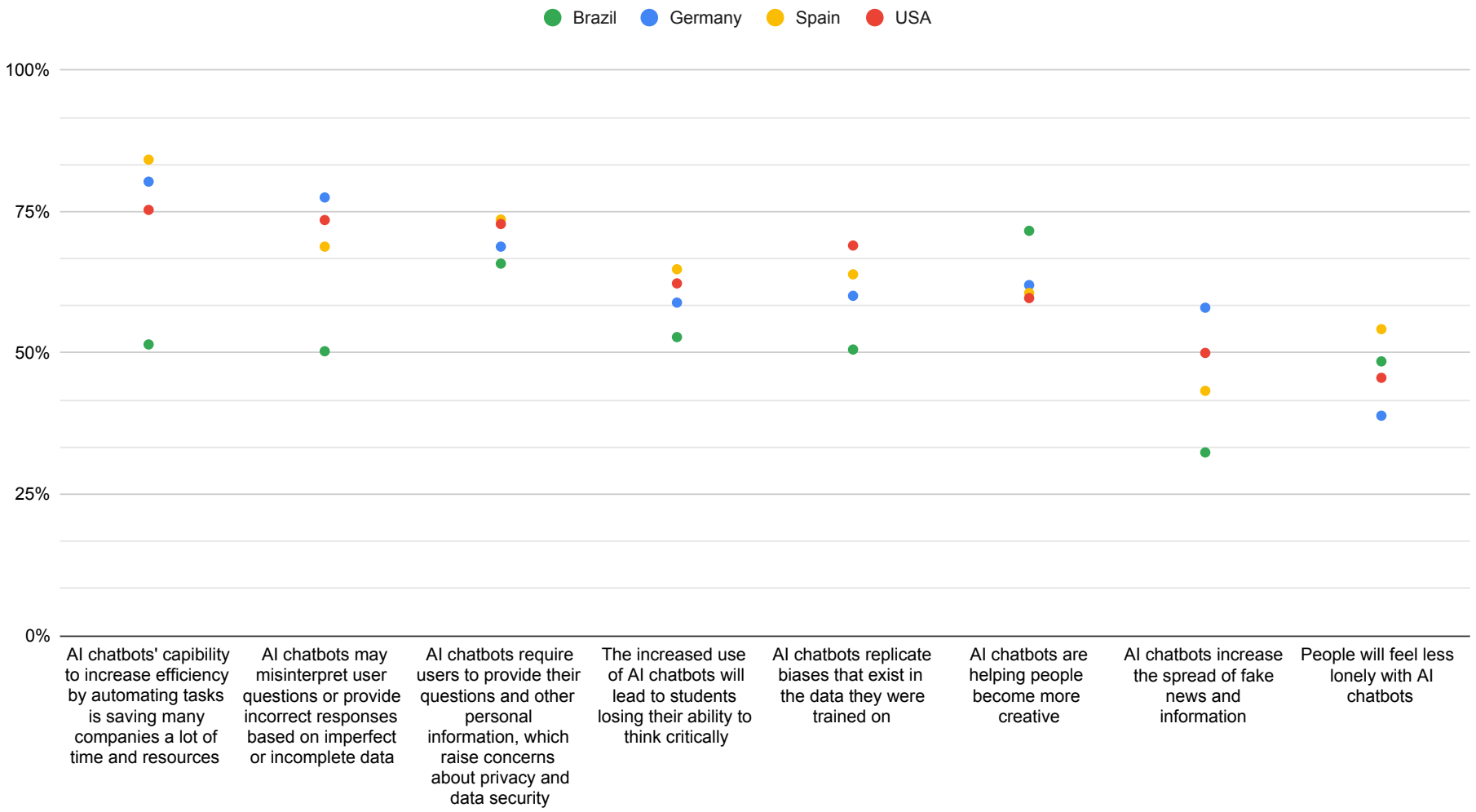
AI statements - Results by country

	Before deliberations			
	Brazil	**Germany**	**Spain**	**USA**
Statement that received the most approval	AI chatbots are helping people become more creative	AI chatbots' capability to increase efficiency by automating tasks is saving many companies a lot of time and resources	AI chatbots' capability to increase efficiency by automating tasks is saving many companies a lot of time and resources	AI chatbots may misinterpret user questions or provide incorrect responses based on imperfect or incomplete data
Statement that received the least approval	AI chatbots increase the spread of fake news and information	People will feel less lonely with AI chatbots	People will feel less lonely with AI chatbots	People will feel less lonely with AI chatbots

	After deliberations			
	Brazil	**Germany**	**Spain**	**USA**
Statement that received the most approval	AI chatbots are helping people become more creative	AI chatbots' capability to increase efficiency by automating tasks is saving many companies a lot of time and resources	AI chatbots' capability to increase efficiency by automating tasks is saving many companies a lot of time and resources	AI chatbots' capability to increase efficiency by automating tasks is saving many companies a lot of time and resources
Statement that received the least approval	AI chatbots increase the spread of fake news and information	People will feel less lonely with AI chatbots	AI chatbots increase the spread of fake news and information	People will feel less lonely with AI chatbots

	Changes			
	Brazil	**Germany**	**Spain**	**USA**
Statement that gained the most approval	AI chatbots are helping people become more creative	AI chatbot require users to provide their questions and other personal information, which raise concerns about privacy and data security	AI chatbots replicate biases that exist in the data they were trained on	AI chatbots replicate biases that exist in the data they were trained on
Statement that lost the most approval	AI chatbots increase the spread of fake news and information	The increased use of AI chatbots will lead to students losing their ability to think critically	The increased use of AI chatbots will lead to students losing their ability to think critically	The increased use of AI chatbots will lead to students losing their ability to think critically

Percentage of participants that agreed with the statements after deliberations



Percentage of participants that agreed with the statements

	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
AI chatbots' capability to increase efficiency by automating tasks is saving many companies a lot of time and resources	51.2%	74.9%	74.4%	70.9%	51.5%	80.3%	84.2%	75.3%	+0.3%	+5.4%*	+9.8%***	+4.4%**
AI chatbots may misinterpret user questions or provide incorrect responses based on imperfect or incomplete data	52.0%	73.2%	65.6%	73.7%	50.3%	77.5%	68.8%	73.5%	-1.7%	+4.3%*	+3.2%	-0.2%
AI chatbots require users to provide their questions and other personal information, which raise concerns about privacy and data security	63.9%	59.1%	66.2%	70.5%	65.8%	68.8%	73.6%	72.8%	+1.9%	+9.7%	+7.4%*	+2.3%
The increased use of AI chatbots will lead to students losing their ability to think critically	59.9%	60.8%	69.2%	63.0%	52.8%	58.9%	64.8%	62.3%	-7.1%	-1.9%	-4.4%	-0.7%
AI chatbots replicate biases that exist in the data they were trained on	47.7%	56.2%	52.6%	57.3%	50.6%	60.1%	63.9%	69.0%	+2.9%*	+3.9%	+11.3%*	+11.7%
AI chatbots are helping people become more creative	66.7%	52.9%	53.8%	55.0%	71.6%	62.0%	60.6%	59.7%	+4.9%	+9.1%**	+6.8%	+4.7%
AI chatbots increase the spread of fake news and information	40.8%	57.1%	44.3%	49.9%	32.4%	58.0%	43.3%	50.0%	-8.4%	+0.9%	-1.0%	+0.1%
People will feel less lonely with AI chatbots	48.4%	31.3%	43.6%	39.9%	48.5%	38.9%	54.2%	45.6%	+0.1%	+7.6%***	+10.6%***	+5.7%***

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Proposals

Should AI chatbots inform users that they are interacting with a bot?

Participants were given a set of proposals regarding how frequently a user could be informed that they are interacting with a bot:

- Every time the AI chatbot responds to a question
- Periodically while using an AI chatbot
- The first time they register to use an AI chatbot
- Never

The proposal that received the most support was for users to be informed when they first register but no single proposal received majority approval from the participants. The least favored proposal across all four countries was for users to never be informed.

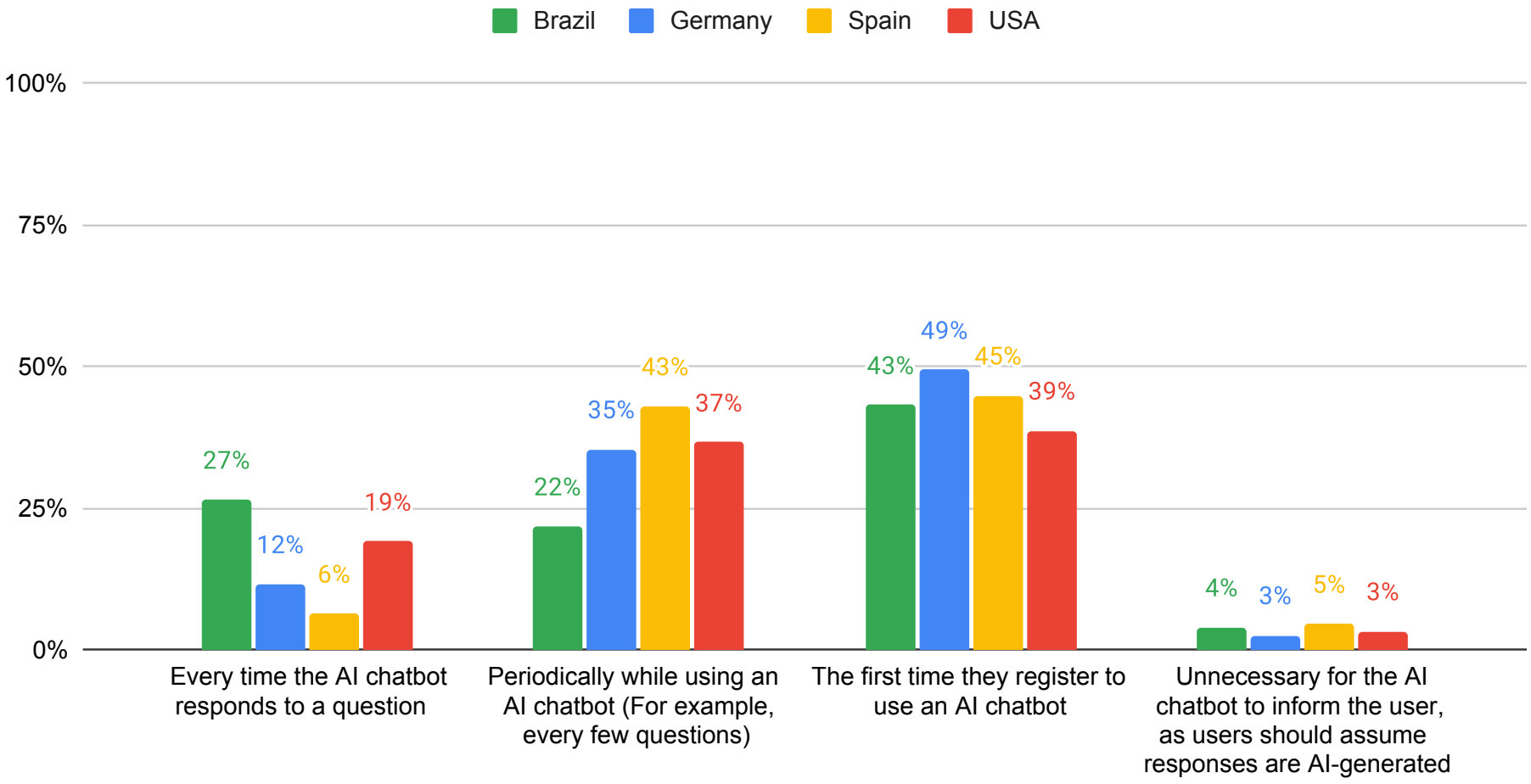
The proposal for users to be informed upon signing up received the most approval from German participants after deliberations (49.4%). The least approval for this proposal came from American participants (30.4% before deliberations, 38.7% after deliberations).

American participants originally highly supported the proposal for users to be informed every time the AI chatbot responds to a question, but this proposal lost the support of a large number of participants over the course of the deliberations, in favor of the proposal for users to be informed upon signing up.

The proposal for users to be informed periodically was the second most favored in all countries except Brazil, where it was the proposal to be informed every time the AI chatbot responds to a question.

Only two proposals gained approval from the participants over the course of the deliberations - the proposal for users to be periodically informed and the proposal for users to be informed upon signing up.

Should AI chatbots inform users that they are interacting with a bot? Percentage of participants supporting the proposals after deliberations



How often should a user be informed by the AI chatbot that they are interacting with a bot?

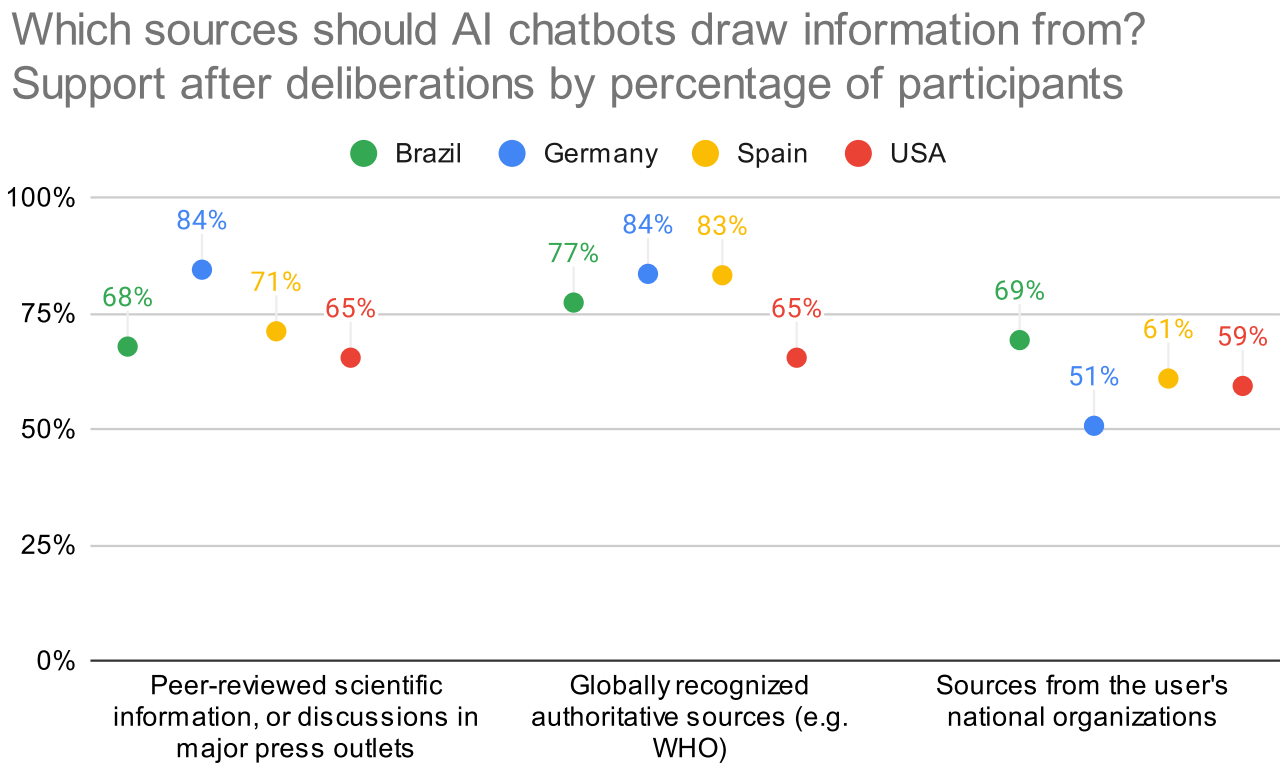
	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
Every time the AI chatbot responds to a question	28.0%	19.9%	21.6%	33.6%	27%	12%	6%	19%	-1.5%	-8.3%***	-15.2%***	-14.5%***
Periodically while using an AI chatbot (For example, every few questions)	18.5%	26.6%	29.4%	27.2%	22%	35%	43%	37%	3.3%	8.8%***	13.6%***	9.4%***
The first time they register to use an AI chatbot	37.3%	47.3%	39.3%	30.4%	43%	49%	45%	39%	5.9%	2.1%	5.6%*	8.3%**
Unnecessary for the AI chatbot to inform the user, as users should assume responses are AI-generated	11.9%	4.2%	5.2%	5.5%	4%	3%	5%	3%	-7.9%***	-1.6%	-0.6%	-2.2%
Don't know	4.3%	2.0%	4.5%	3.3%	4.5%	0.9%	1.1%	2.3%	0.2%	-1.1%	-3.4%***	-1%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Which sources should AI chatbots draw information from?

Participants were given three types of sources AI chatbots could draw information from:

- Peer-reviewed scientific information, or discussion in major press outlets
- Globally recognized sources (e.g. WHO)
- The user's national organizations



All three options received the approval of all four countries after deliberations. Nonetheless, the option for information to be sourced by the user's national organizations received slightly less support than other options.The only exception is Brazil, where the option for peer-reviewed scientific information received slightly less support than the option for the user's national organizations.

The option for AI chatbots to source their information from globally recognized sources was the most favored across all four countries before deliberations. After deliberations, it was also the most favored across all countries except for Germany, where participants gave slightly more approval for peer-reviewed scientific information. American participants approved equally of information sourced from globally recognized sources and peer-reviewed scientific sources, after deliberations.

AI sources - Percentage of participants that approved the proposals

	Before deliberations				After deliberations				Change			
	Brazil	**Germany**	**Spain**	**USA**	**Brazil**	**Germany**	**Spain**	**USA**	**Brazil**	**Germany**	**Spain**	**USA**
Peer-reviewed scientific information, or discussions in major press outlets	61%	81%	67%	57%	68%	**84%**	71%	**65%**	7.2%*	4%	4%	8.4%*
Globally recognized authoritative sources (e.g. WHO)	**78%**	**81%**	**78%**	**62%**	**77%**	84%	**83%**	**65%**	-1%	2%	5%	3.7%*
Sources from the user's national organizations	67%	42%	54%	48%	69%	51%	61%	59%	3%	8.9%**	7.3%*	11.3%**

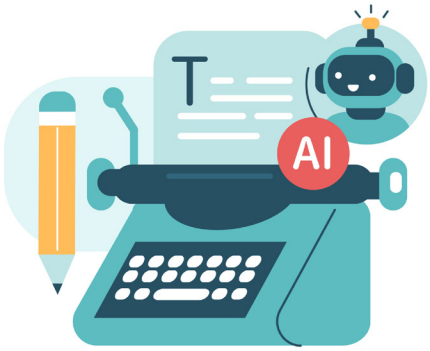
Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Which perspectives should AI sources draw information from?

AI chatbots can draw perspectives from a variety of sources - local media outlets, national organizations or international organizations. The perspectives used by AI chatbots changes theirs answers.

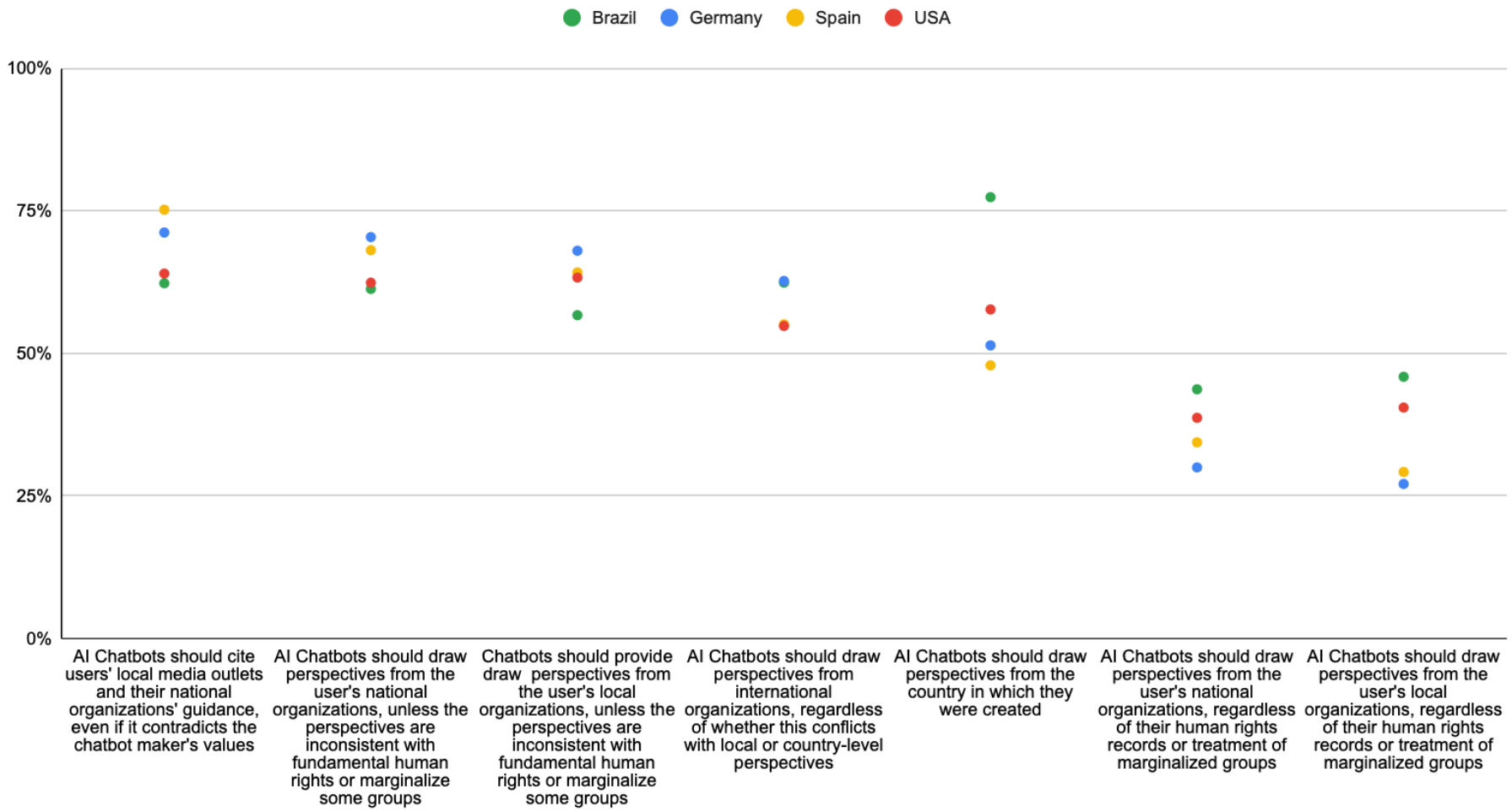
Participants from Germany, Spain and the US were most in favor of perspectives to be drawn from the users' local media outlets and national organizations, even if it contradicts the chatbot maker's values. Brazilian participants most favored the proposal for AI chatbots to draw perspectives from the country in which they were created.

Overall, a majority of participants from all four countries supported perspectives to be drawn from local media outlets, national organizations and international organizations. The only two proposals that were not approved by a majority of participants



across all countries were those where AI would draw perspectives from local and national organizations, regardless of their human rights records or treatment of marginalized groups. As such, **what mattered most in the participants' answers was not so much whether the perspectives were drawn from local, national or international organizations, but whether those organizations respected human rights records and marginalized groups.**

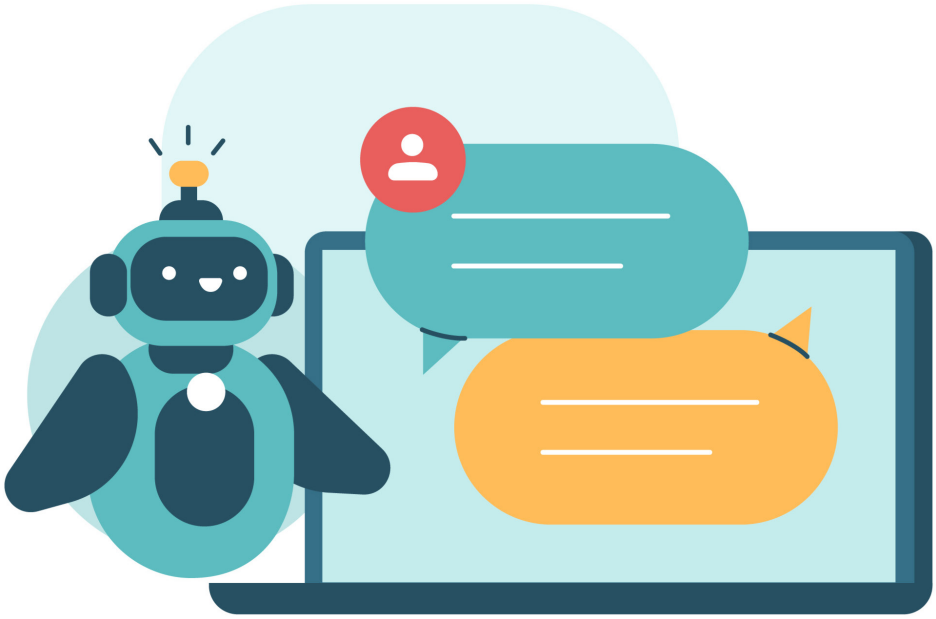
Which perspectives should AI chatbots draw information from? Results by percentage of participants, after deliberations



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
AI chatbots should draw perspectives from international organizations, regardless of whether this conflicts with local or country-level perspectives	70.0%	64.5%	47.1%	52.0%	62.4%	62.7%	55.1%	54.8%	-7.6%**	-1.8%*	8.0%	2.8%
AI chatbots should draw perspectives from the user's national organizations, regardless of their human rights records or treatment of marginalized groups	53.5%	28.3%	34.5%	35.3%	43.7%	30.0%	34.4%	38.7%	-9.8%**	1.7%	-0.1%	3.4%
AI chatbots should draw perspectives from the user's national organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	53.4%	57.5%	59.5%	51.4%	61.3%	70.4%	68.1%	62.4%	7.9%	12.9%*	8.6%	11%*
AI chatbots should draw perspectives from the user's local organizations, regardless of their human rights records or treatment of marginalized groups	45.8%	24.9%	33.1%	32.5%	45.9%	27.1%	29.2%	40.5%	0.1%	2.2%	-3.9%	8.0%
AI chatbots should draw perspectives from the user's local organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	58.3%	57.1%	54.5%	50.2%	56.7%	68.0%	64.2%	63.3%	-1.6%	10.9%*	9.7%	13.1%*
AI chatbots should draw perspectives from the country in which they were created	72.4%	41.4%	38.7%	44.6%	77.4%	51.4%	47.9%	57.7%	5.0%	10%**	9.2%*	13.1%*
AI chatbots should cite users' local media outlets and their national organizations' guidance, even if it contradicts the chatbot maker's values	66.6%	67.5%	68.8%	58.3%	62.3%	71.2%	75.2%	64.0%	-4.3%	3.7%	6.4%	5.7%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Should AI chatbots use the user's past conversations to improve user experience?



Participants were given a set of three proposals regarding the use by AI chatbots of the user's past conversations:

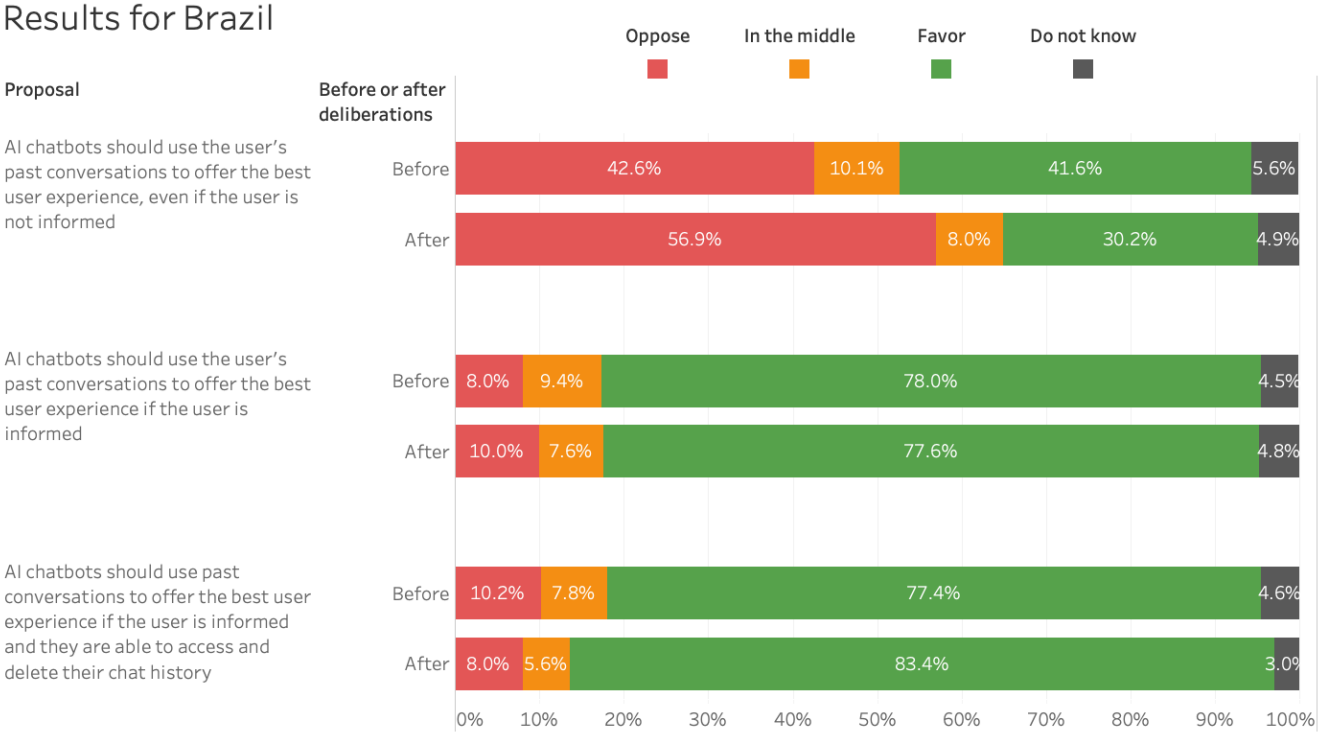
- AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed.
- AI chatbots should use the user's past conversations to offer the best user experience if the user is informed.
- AI chatbots should use past conversations to offer the best user experience if the user is informed and they are able to access and delete their chat history.

The first proposal did not receive the support of a majority of participants in any of the four countries in our sample. In fact, agreement with the proposal went down in all four countries over the course of deliberations. German participants were the most opposed to this proposal.

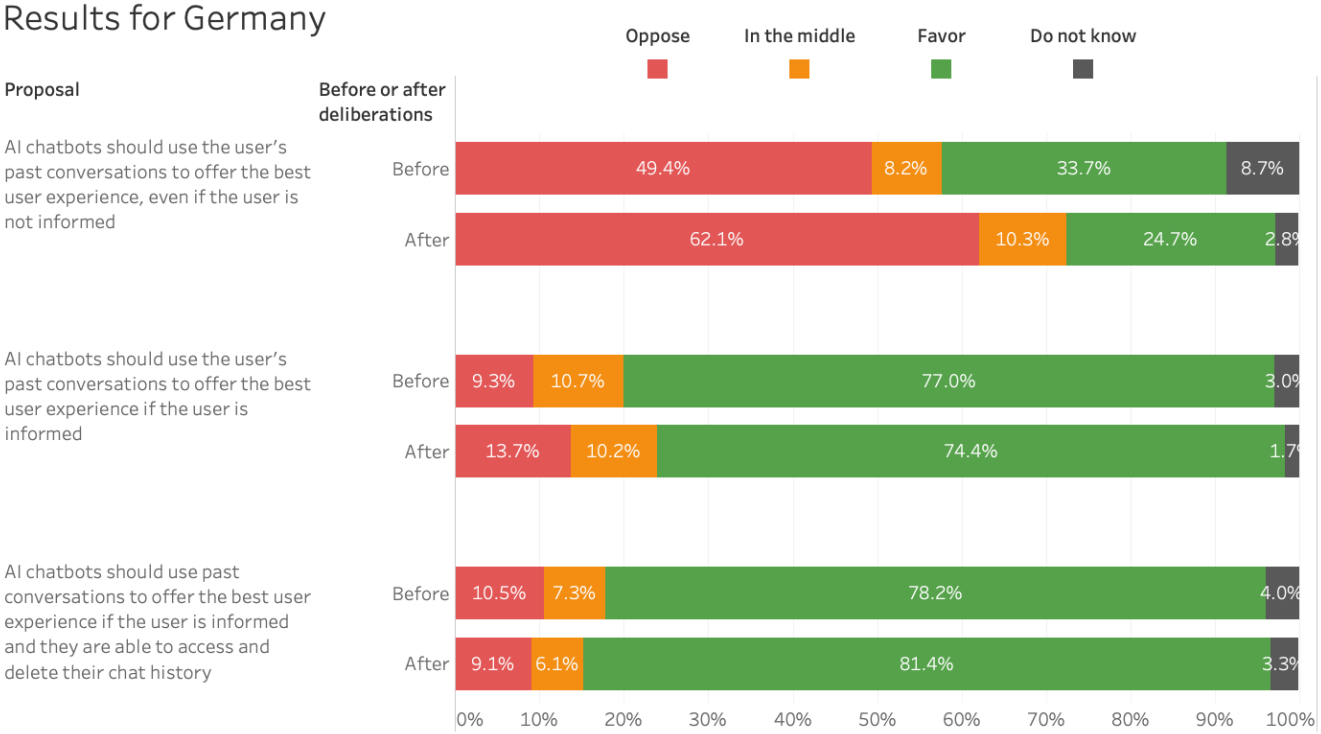
The second and third proposals received the approval of a majority of participants in all four countries. The third proposal received slightly higher levels of approval

than the second proposal for all four countries. In fact, agreement with the third proposal went up as a result of the deliberations in all four countries, whereas it only went up for the second proposal in Spain and the US. The results highlight that participants favored transparency as well as user agency over their data.

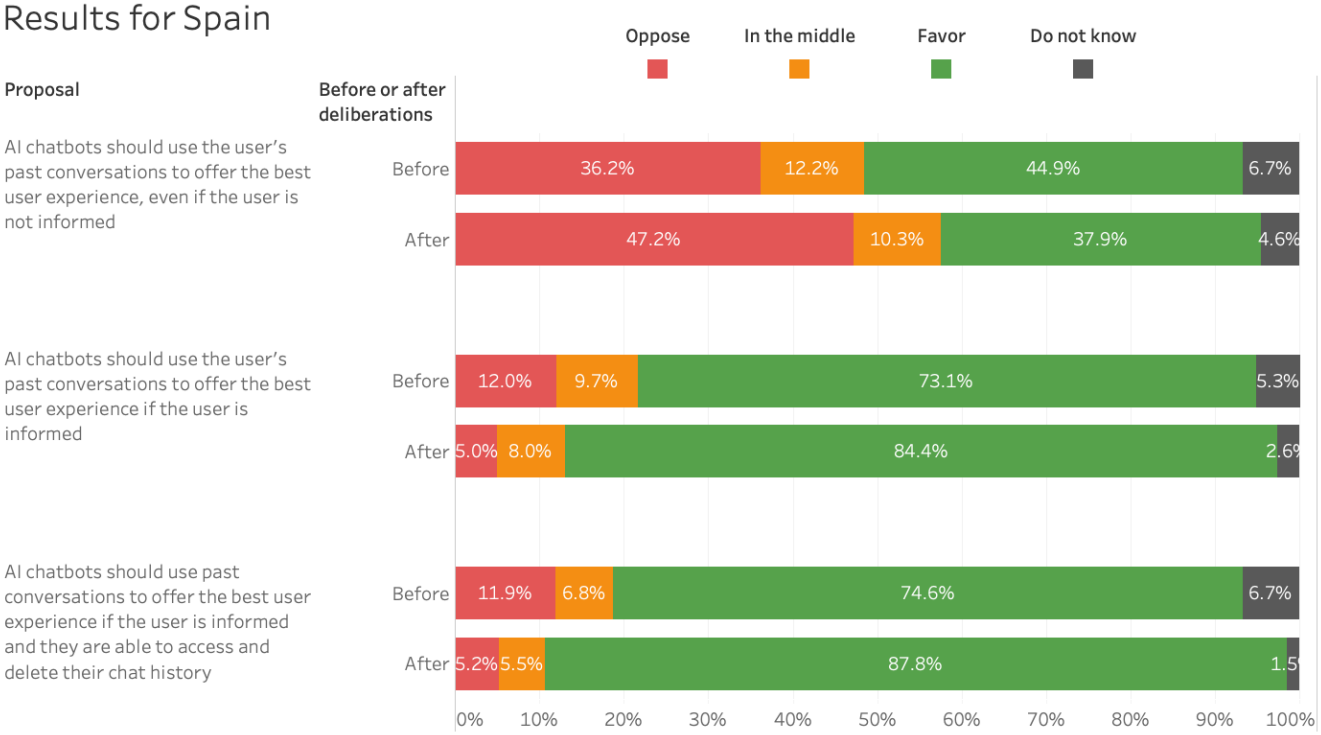
Should AI chatbots use the user's past conversations to improve user experience? Results for Brazil



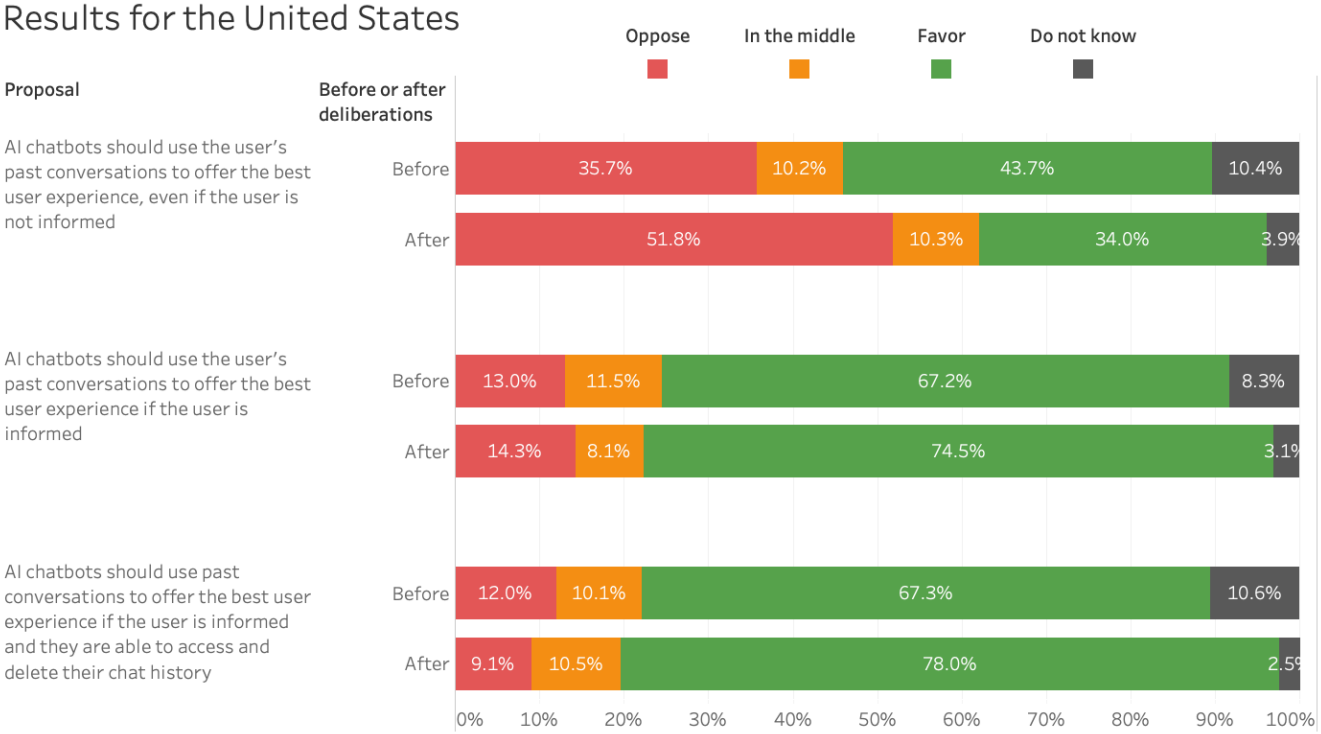
Should AI chatbots use the user's past conversations to improve user experience? Results for Germany



Should AI chatbots use the user's past conversations to improve user experience? Results for Spain



Should AI chatbots use the user's past conversations to improve user experience? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed	5.08	4.14	5.04	5.11	4.02	3.23	4.34	4.15	-1.06***	-0.91***	-0.7***	-0.97***
Oppose	42.6%	49.4%	36.2%	35.7%	56.9%	62.1%	47.2%	51.8%	14.3%	12.7%	11.0%	16.1%
In the middle	10.1%	8.2%	12.2%	10.2%	8.0%	10.3%	10.3%	10.3%	-2.1%	2.1%	-1.9%	0.1%
Favor	41.6%	33.7%	44.9%	43.7%	30.2%	24.7%	37.9%	34.0%	-11.4%	-9.0%	-7.0%	-9.7%
Do not know	5.6%	8.7%	6.7%	10.4%	4.9%	2.8%	4.6%	3.9%	-0.7%	-5.9%	-2.1%	-6.5%
AI chatbots should use the user's past conversations to offer the best user experience if the user is informed	7.88	7.35	7.17	6.87	7.76	7.10	7.67	6.96	-0.12	-0.25	0.49***	0.09
Oppose	8.0%	9.3%	12.0%	13.0%	10.0%	13.7%	5.0%	14.3%	2.0%	4.4%	-7.0%	1.3%
In the middle	9.4%	10.7%	9.7%	11.5%	7.6%	10.2%	8.0%	8.1%	-1.8%	-0.5%	-1.7%	-3.4%
Favor	78.0%	77.0%	73.1%	67.2%	77.6%	74.4%	84.4%	74.5%	-0.4%	-2.6%	11.3%	7.3%
Do not know	4.5%	3.0%	5.3%	8.3%	4.8%	1.7%	2.6%	3.1%	0.3%	-1.3%	-2.7%	-5.2%
AI chatbots should use past conversations to offer the best user experience if the user is informed and they are able to access and delete their chat history	7.95	7.60	7.32	6.98	8.27	7.84	8.15	7.50	0.32	0.24	0.83***	0.52**
Oppose	10.2%	10.5%	11.9%	12.0%	8.0%	9.1%	5.2%	9.1%	-2.2%	-1.4%	-6.7%	-2.9%
In the middle	7.8%	7.3%	6.8%	10.1%	5.6%	6.1%	5.5%	10.5%	-2.2%	-1.2%	-1.3%	0.4%
Favor	77.4%	78.2%	74.6%	67.3%	83.4%	81.4%	87.8%	78.0%	6.0%	3.2%	13.2%	10.7%
Do not know	4.6%	4.0%	6.7%	10.6%	3.0%	3.3%	1.5%	2.5%	-1.6%	-0.7%	-5.2%	-8.1%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

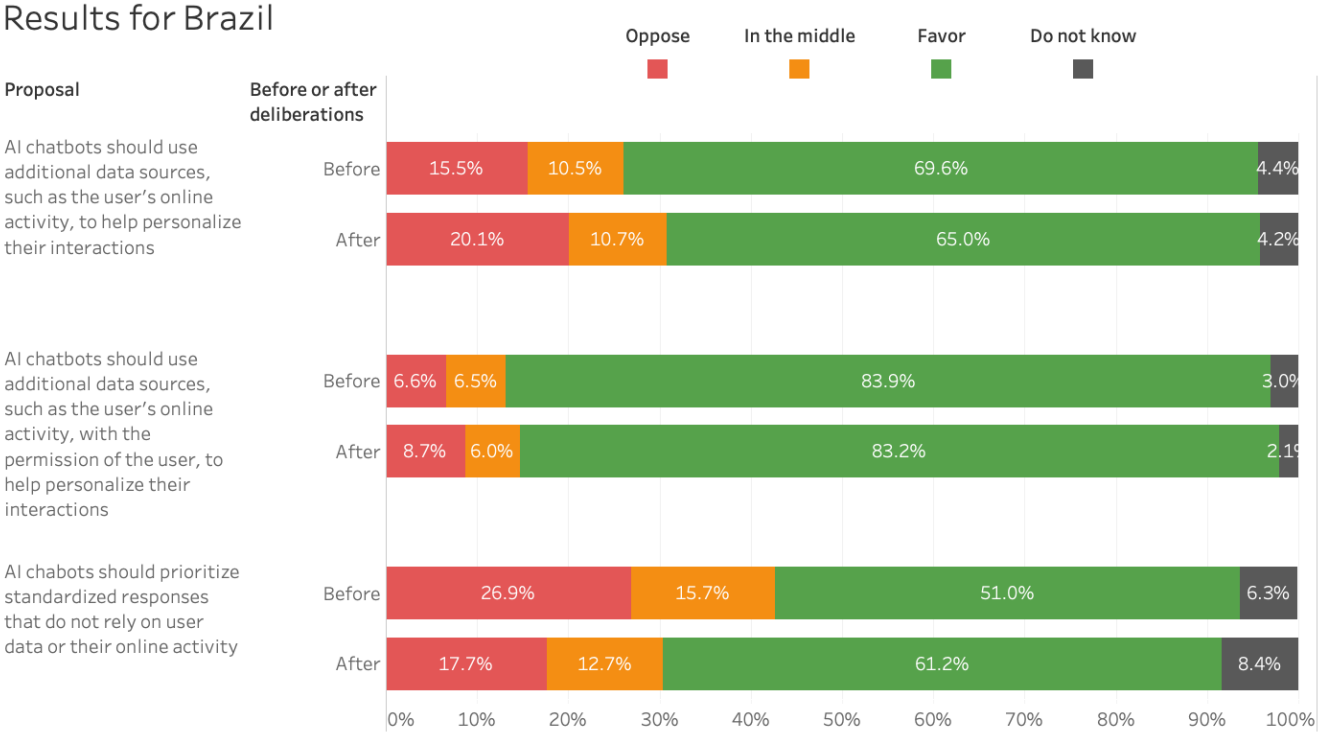
Should AI chatbots use the user's online activity to personalize interactions?

When presented with the proposal for AI chatbots to use additional data sources, such as the user's only activity, to help personalize their interactions, a majority of participants in Brazil and Spain supported the proposal, but not in Germany or the US.

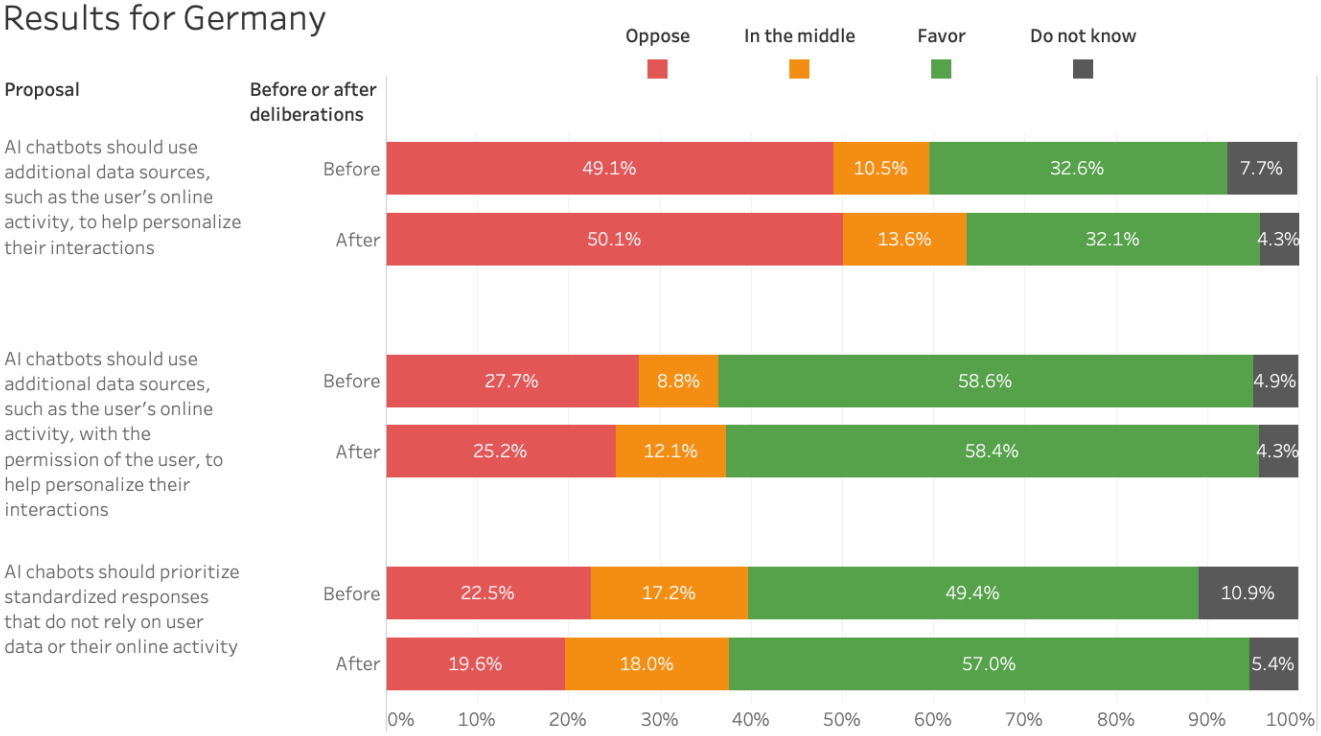
When offered the same proposal, but with the mention that the AI chatbots would have the permission of the users to use their online activity, a majority of participants from all countries supported the proposal. Although the proposal was supported by a majority of participants in all four countries, approval ranged between 58% for German participants to 83% for Brazilian participants.

The proposal "AI chatbots should prioritize standardized responses that do not rely on user data or their online activity" received the approval of a majority of participants in all four countries. Approval for this proposal also went up over the course of deliberations for all four countries. While approval for this proposal was higher overall than for the first proposal, it was lower than for the second one. As such, the participants did not mind having AI chatbots rely on user data, but they wanted the users to allow such an option, not for it to exist by default. This shows that participants valued user consent as well as privacy.

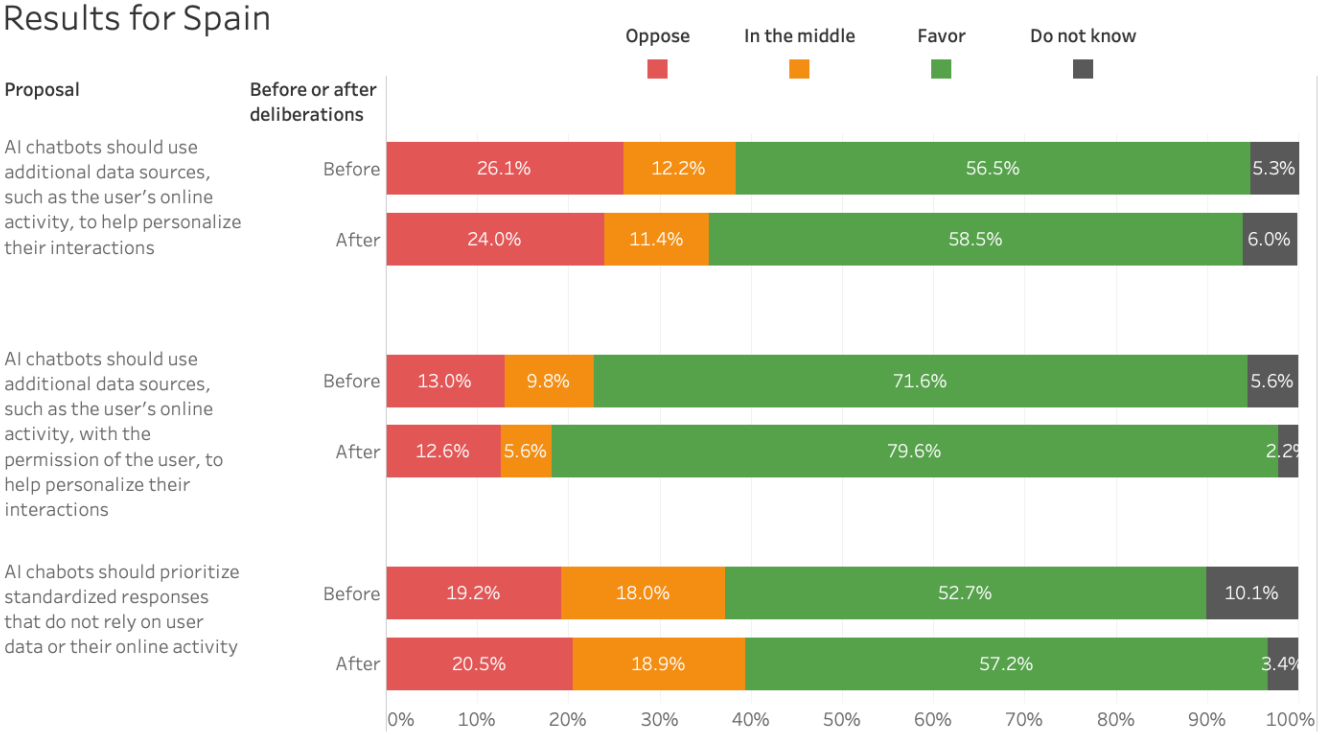
Should AI chatbots use additional data sources, such as the user's online activity? Results for Brazil



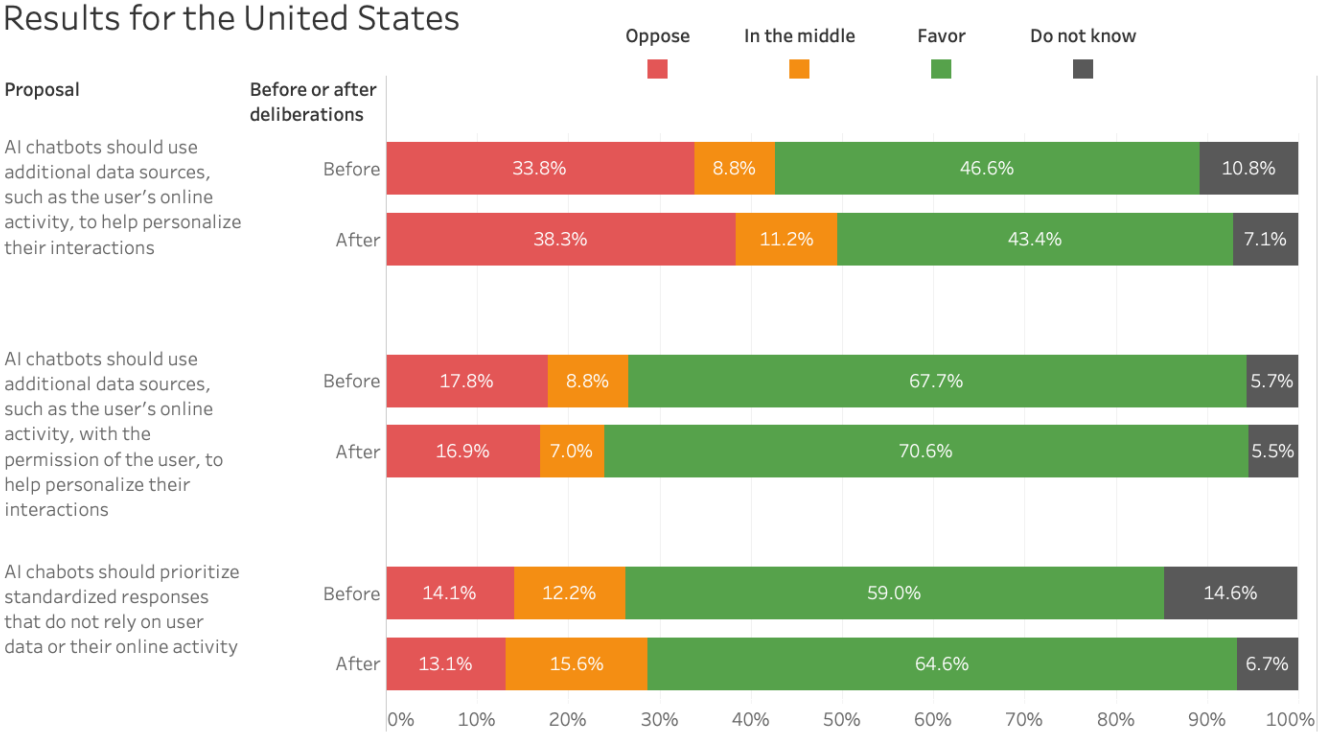
Should AI chatbots use additional data sources, such as the user's online activity? Results for Germany



Should AI chatbots use additional data sources, such as the user's online activity? Results for Spain



Should AI chatbots use additional data sources, such as the user's online activity? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
AI chatbots should use additional data sources, such as the user's online activity, to help personalize their interactions	7.01	4.04	5.74	5.18	6.73	3.92	5.88	5.07	-0.28	-0.12	0.14	-0.10
Oppose	15.5%	49.1%	26.1%	33.8%	20.1%	50.1%	24.0%	38.3%	4.6%	1.0%	-2.1%	4.5%
In the middle	10.5%	10.5%	12.2%	8.8%	10.7%	13.6%	11.4%	11.2%	0.2%	3.1%	-0.8%	2.4%
Favor	69.6%	32.6%	56.5%	46.6%	65.0%	32.1%	58.5%	43.4%	-4.6%	-0.5%	2.0%	-3.2%
Do not know	4.4%	7.7%	5.3%	10.8%	4.2%	4.3%	6.0%	7.1%	-0.2%	-3.4%	0.7%	-3.7%
AI chatbots should use additional data sources, such as the user's online activity, with the permission of the user, to help personalize their interactions	8.09	5.81	6.99	6.58	7.95	6.01	7.26	6.78	-0.15	0.20	0.26	0.20
Oppose	6.6%	27.7%	13.0%	17.8%	8.7%	25.2%	12.6%	16.9%	2.1%	-2.5%	-0.4%	-0.9%
In the middle	6.5%	8.8%	9.8%	8.8%	6.0%	12.1%	5.6%	7.0%	-0.5%	3.3%	-4.2%	-1.8%
Favor	83.9%	58.6%	71.6%	67.7%	83.2%	58.4%	79.6%	70.6%	-0.7%	-0.2%	8.0%	2.9%
Do not know	3.0%	4.9%	5.6%	5.7%	2.1%	4.3%	2.2%	5.5%	-0.9%	-0.6%	-3.4%	-0.2%
AI chatbots should prioritize standardized responses that do not rely on user data or their online activity	6.16	5.82	6.12	6.65	6.88	6.17	6.26	6.78	0.72***	0.35*	0.14	0.13
Oppose	26.9%	22.5%	19.2%	14.1%	17.7%	19.6%	20.5%	13.1%	-9.2%	-2.9%	1.3%	-1.0%
In the middle	15.7%	17.2%	18.0%	12.2%	12.7%	18.0%	18.9%	15.6%	-3.0%	0.8%	0.9%	3.4%
Favor	51.0%	49.4%	52.7%	59.0%	61.2%	57.0%	57.2%	64.6%	10.2%	7.6%	4.5%	5.6%
Do not know	6.3%	10.9%	10.1%	14.6%	8.4%	5.4%	3.4%	6.7%	2.1%	-5.5%	-6.7%	-7.9%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

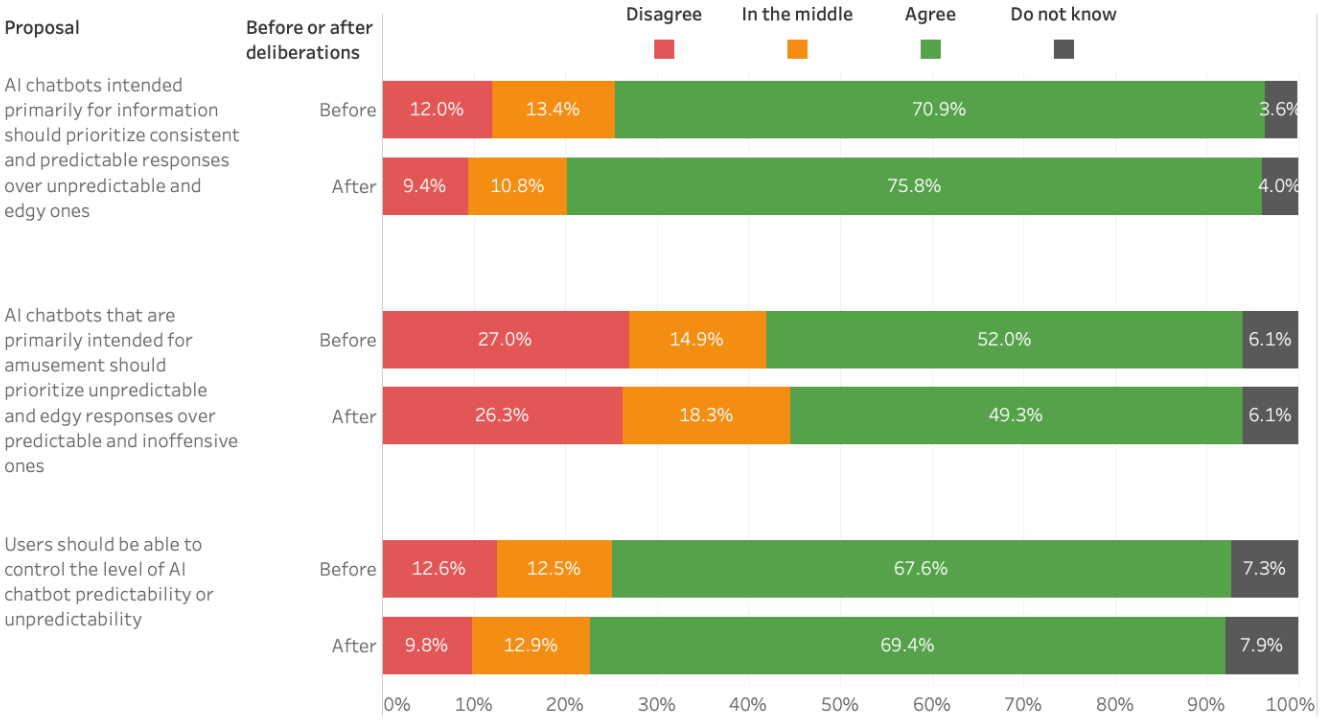
Should AI chatbots prioritize predictability or unpredictability?

A majority of participants from all four countries supported the proposal "AI Chatbots intended primarily for information should prioritize consistent and predictable responses over unpredictable and edgy ones." Approval for this proposal also went up as a result of deliberations in all four countries.

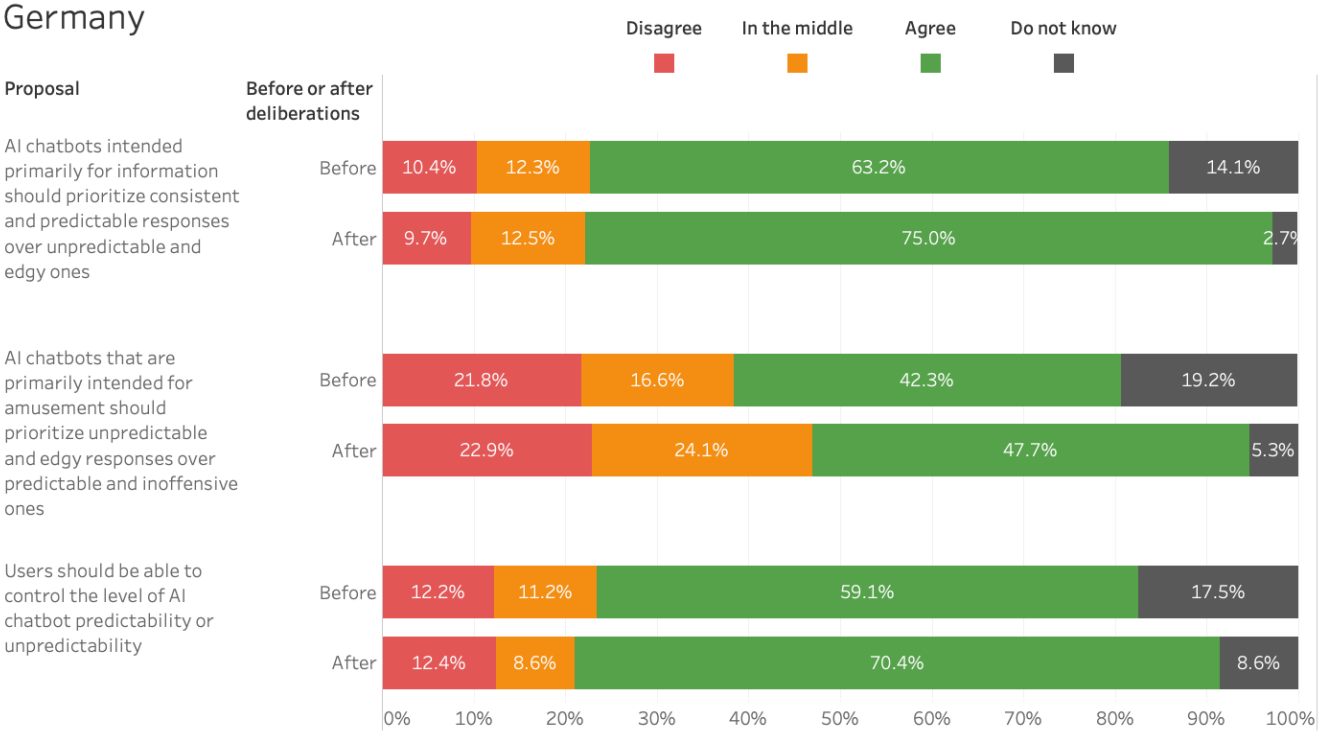
Participants' support for the first proposal did not translate in them supporting unpredictable answers for AI chatbots primarily intended for amusement. The proposal "AI Chatbots that are primarily intended for amusement should prioritize unpredictable and edgy responses over predictable and inoffensive ones" only received the support of a majority of participants in Spain after deliberations. Nevertheless, levels of approval with the proposal were close to the majority among Brazilian, German, and American participants. Levels of support also went up over the course of deliberations among Spanish and American participants.

The proposal "Users should be able to control the level of AI chatbot predictability or unpredictability" received the approval of a majority of participants in all four countries. Support for this proposal went up over the course of deliberations across all four countries as well.

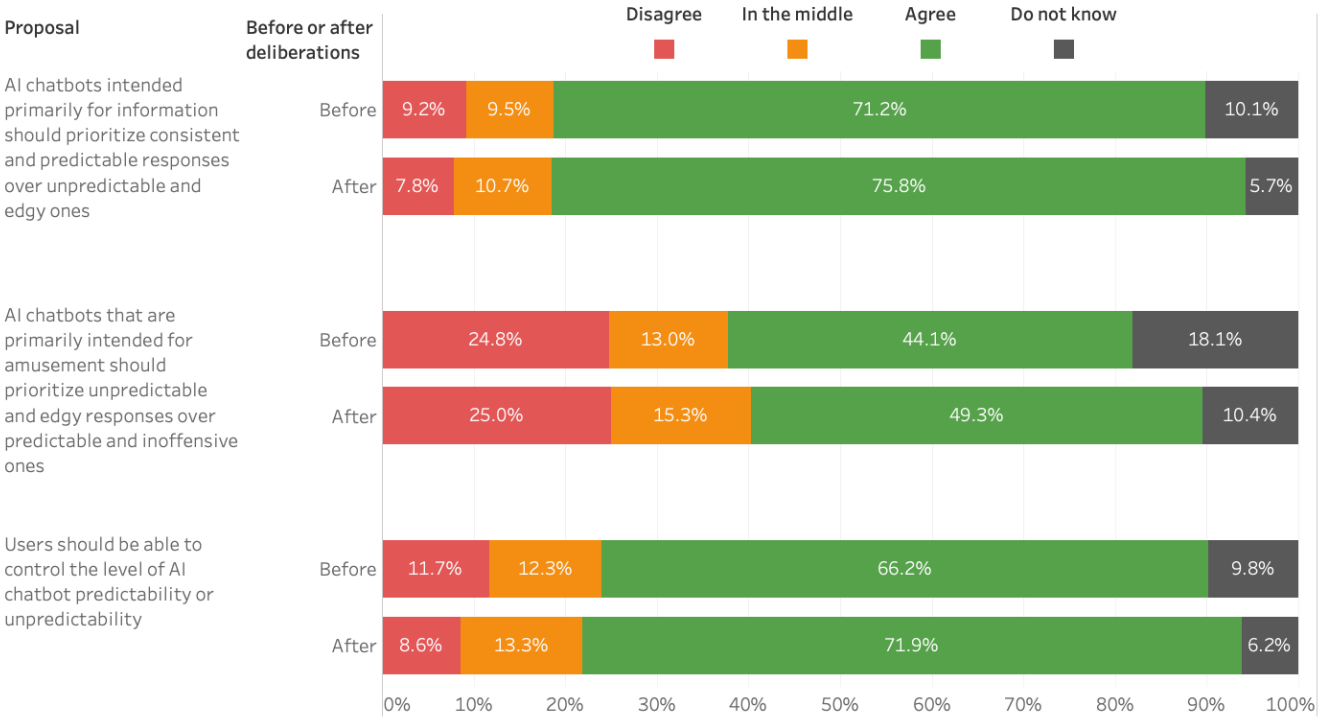
Should AI chatbots prioritize predictability or unpredictability? Results for Brazil



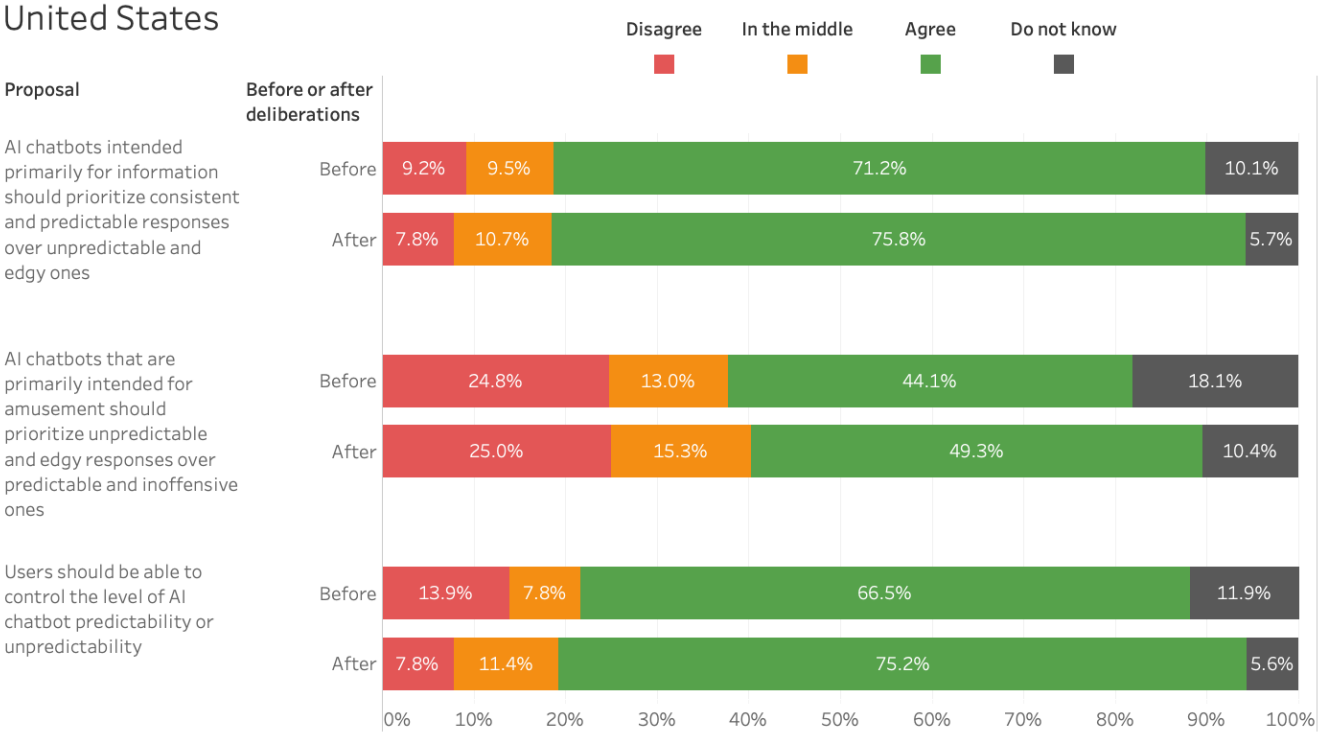
Should AI chatbots prioritize predictability or unpredictability? Results for Germany



Should AI chatbots prioritize predictability or unpredictability? Results for Spain



Should AI chatbots prioritize predictability or unpredictability? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
AI chatbots intended primarily for information should prioritize consistent and predictable responses over unpredictable and edgy ones	7.39	6.91	7.01	7.38	7.63	7.32	7.21	7.67	0.23	0.41**	0.20	0.29*
Disagree	12.0%	10.4%	12.1%	9.2%	9.4%	9.7%	11.2%	7.8%	-2.6%	-0.7%	-0.9%	-1.4%
In the middle	13.4%	12.3%	9.5%	9.5%	10.8%	12.5%	10.6%	10.7%	-2.6%	0.2%	1.1%	1.2%
Agree	70.9%	63.2%	69.9%	71.2%	75.8%	75.0%	74.8%	75.8%	4.9%	11.8%	4.9%	4.6%
Do not know	3.6%	14.1%	8.4%	10.1%	4.0%	2.7%	3.4%	5.7%	0.4%	-11.4%	-5.0%	-4.4%
AI chatbots that are primarily intended for amusement should prioritize unpredictable and edgy responses over predictable and inoffensive ones	6.12	5.64	5.55	5.62	5.85	5.89	6.11	5.75	-0.27	0.24	0.56**	0.13
Disagree	27.0%	21.8%	23.2%	24.8%	26.3%	22.9%	20.7%	25.0%	-0.7%	1.1%	-2.5%	0.2%
In the middle	14.9%	16.6%	17.8%	13.0%	18.3%	24.1%	20.1%	15.3%	3.4%	7.5%	2.3%	2.3%
Agree	52.0%	42.3%	48.8%	44.1%	49.3%	47.7%	55.8%	49.3%	-2.7%	5.4%	7.0%	5.2%
Do not know	6.1%	19.2%	10.3%	18.1%	6.1%	5.3%	3.4%	10.4%	0.0%	-13.9%	-6.9%	-7.7%
Users should be able to control the level of AI chatbot predictability or unpredictability	7.28	6.97	6.91	7.03	7.58	7.32	7.35	7.43	0.30	0.35	0.44**	0.41*
Disagree	12.6%	12.2%	11.7%	13.9%	9.8%	12.4%	8.6%	7.8%	-2.8%	0.2%	-3.1%	-6.1%
In the middle	12.5%	11.2%	12.3%	7.8%	12.9%	8.6%	13.3%	11.4%	0.4%	-2.6%	1.0%	3.6%
Agree	67.6%	59.1%	66.2%	66.5%	69.4%	70.4%	71.9%	75.2%	1.8%	11.3%	5.7%	8.7%
Do not know	7.3%	17.5%	9.8%	11.9%	7.9%	8.6%	6.2%	5.6%	0.6%	-8.9%	-3.6%	-6.3%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Should AI chatbots be allowed to discuss topics some people might find offensive?

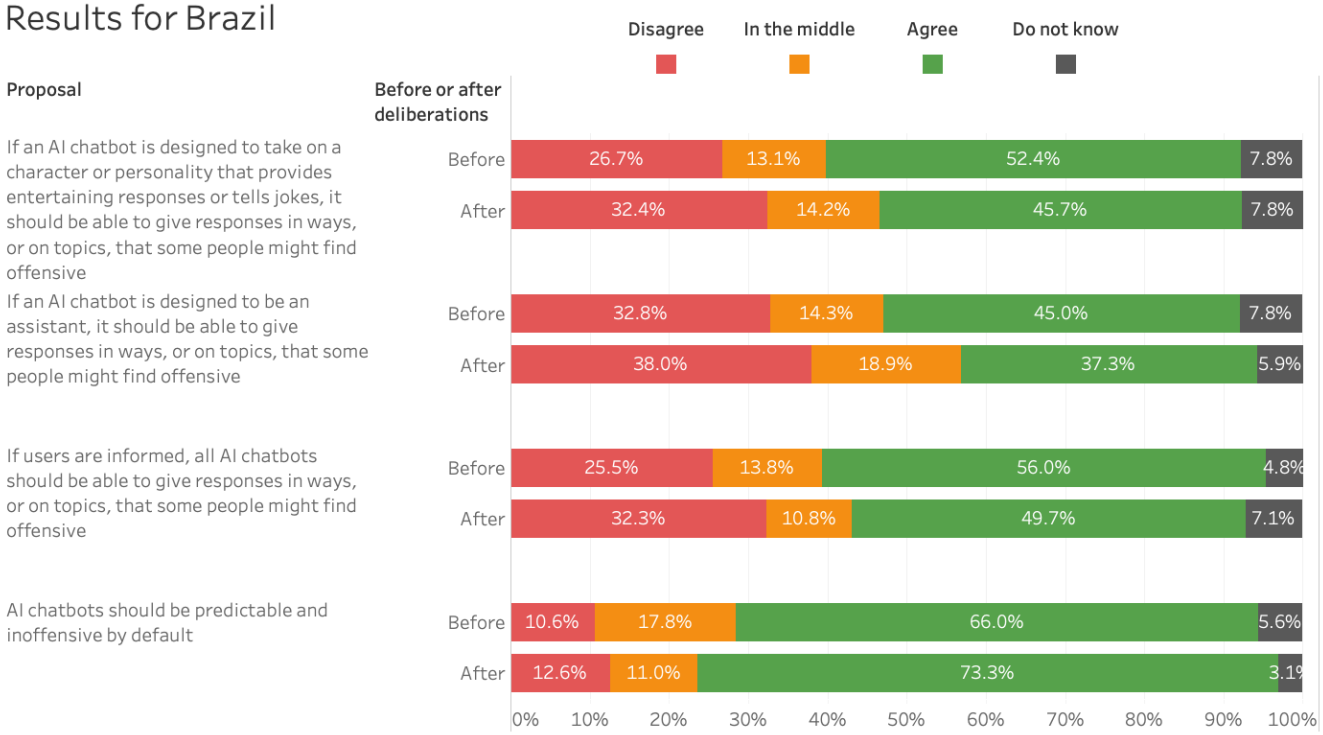
The most popular proposal in this category across all four countries was for AI chatbots to be predictable and inoffensive by default. The highest levels of support came from Brazilian participants, with 73.3% of them approving the proposal after deliberations. All three other proposals, aimed at allowing AI chatbots to discuss topics some people might find offensive, received significantly less approval, and in most cases failed to gather the support of a majority of participants from all four countries.

The proposal which received the least approval across all four countries was "If an AI chatbot is designed to be an assistant, it should be able to give responses in ways, or on topics, that some people might find offensive." The highest amount of support was 40.7% among Spanish participants, after deliberations.

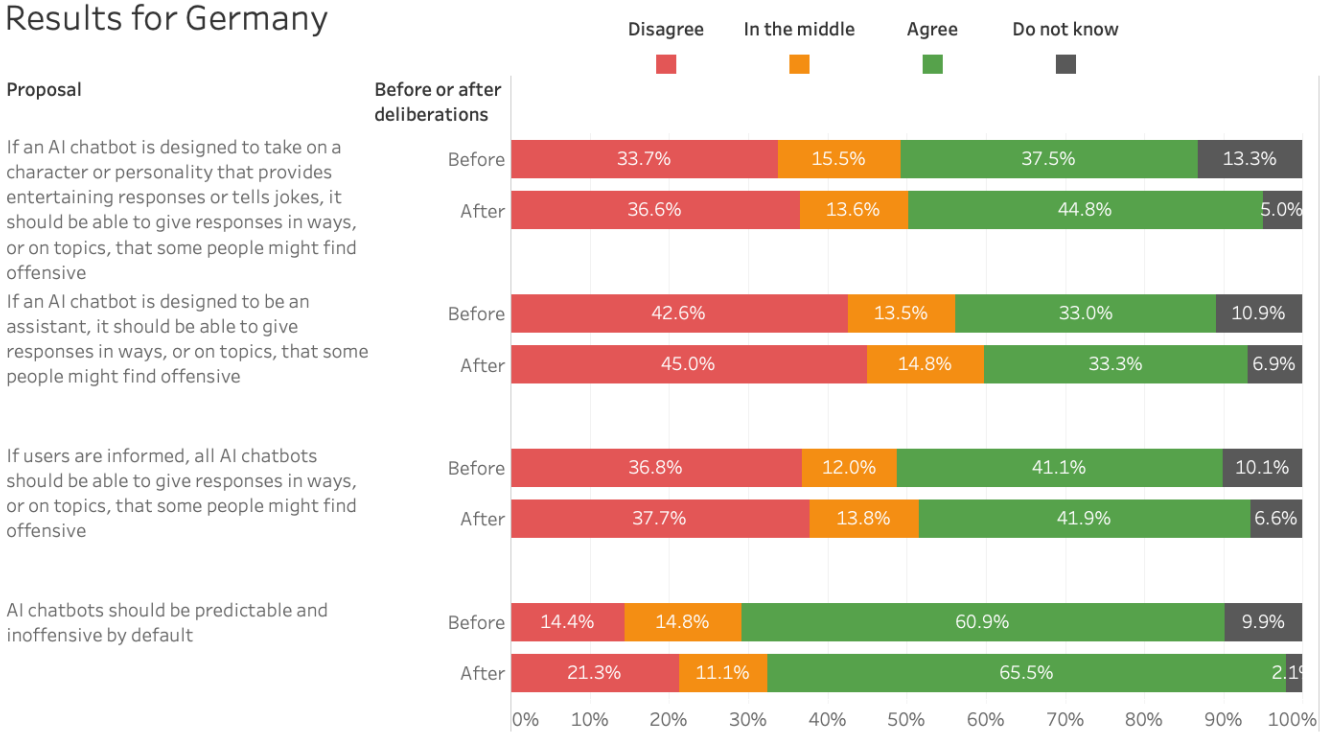
The proposal "If an AI Chatbot is designed to take on a character or personality that provides entertaining responses or tells jokes, it should be able to give responses in ways, or on topics, that some people might find offensive" was also not approved by a majority of participants. Only a small majority of American participants (52.7%) approved the proposal after deliberations.

The proposal "If users are informed, all AI Chatbots should be able to give responses in ways, or on topics, that some people might find offensive" was supported after deliberations by a very thin majority of Spanish (50.8%) and American (50.7%) participants. Brazilian and German participants did not approve for the majority this proposal. The least approval came from German participants, with 41.9% after deliberations. The fact that this proposal received the most support among all three proposals aimed at allowing AI chatbots to discuss offensive topics demonstrates that participants were not enthusiastic about the idea of having AI chatbots be offensive. Even with the user being informed, only two out of four countries supported the proposal, and even then only 50% of Spanish and American participants supported the proposal, meaning almost half did not support the proposal.

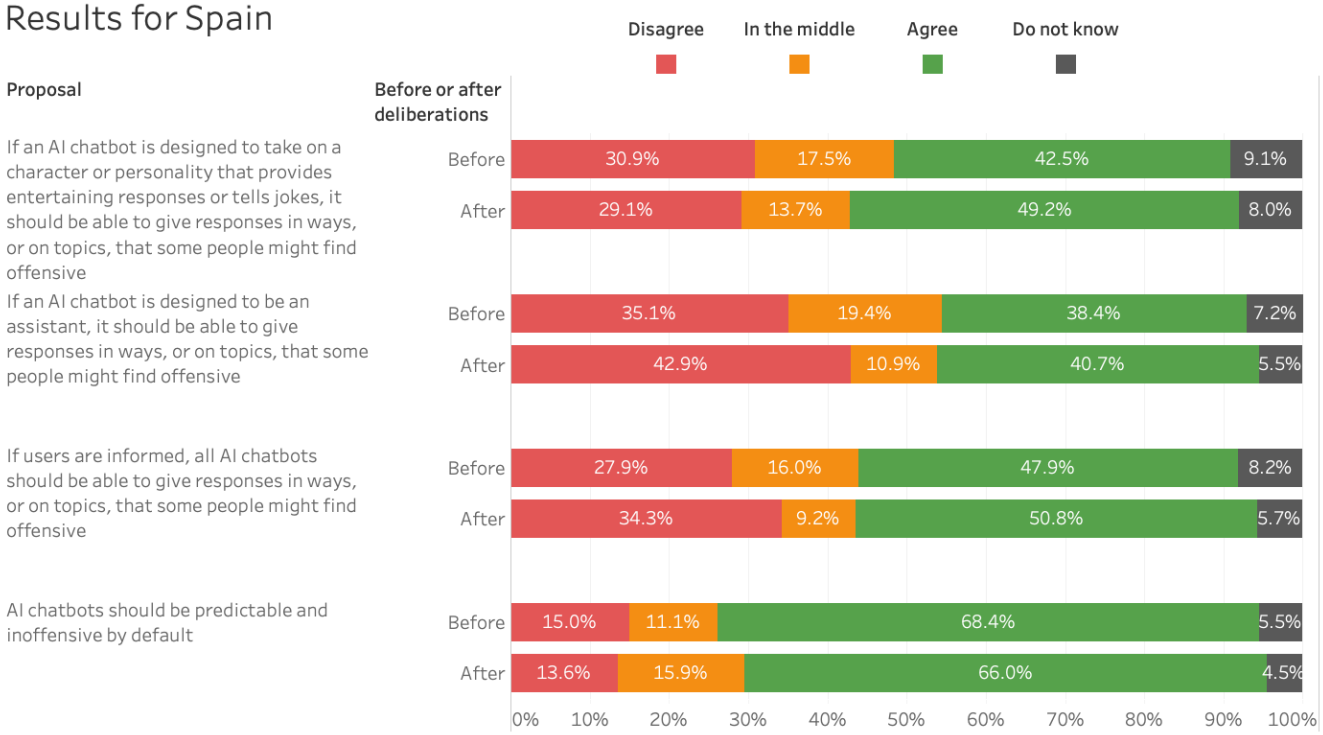
Should AI chatbots be allowed to discuss topics some people might find offensive? Results for Brazil



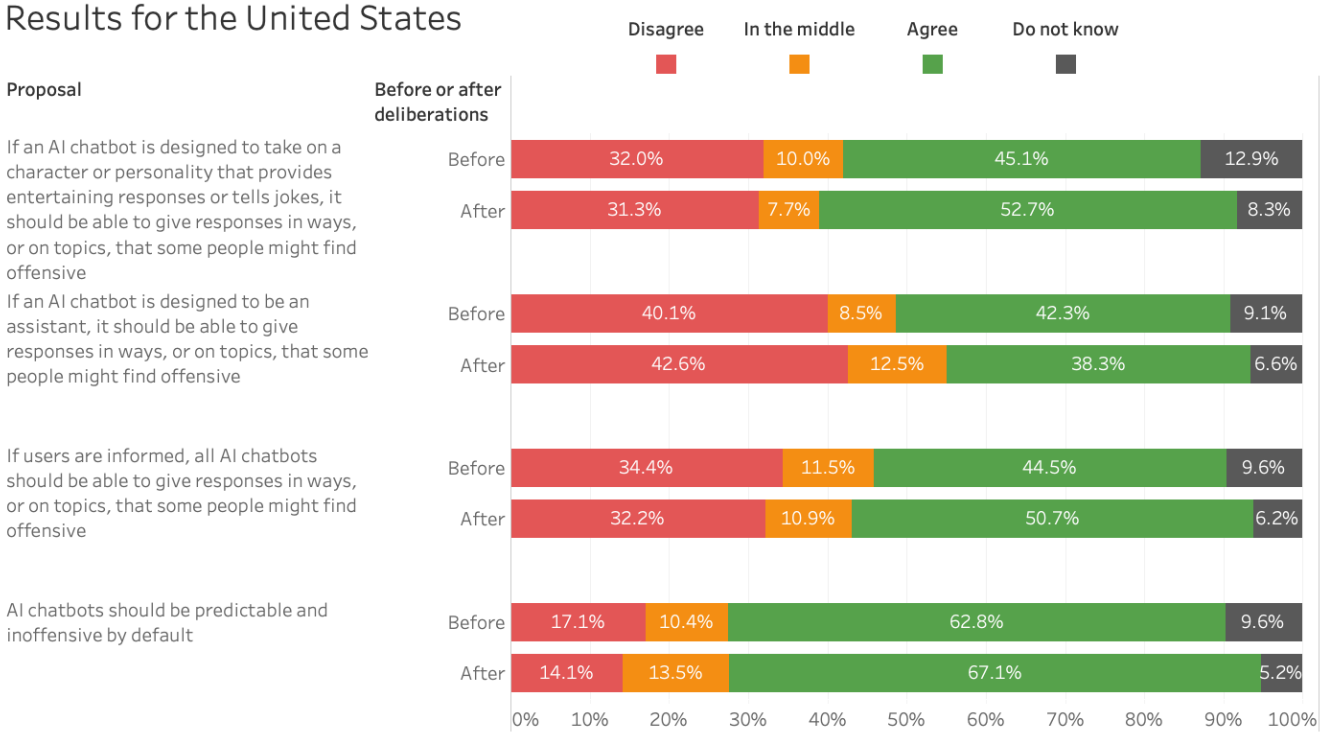
Should AI chatbots be allowed to discuss topics some people might find offensive? Results for Germany



Should AI chatbots be allowed to discuss topics some people might find offensive? Results for Spain



Should AI chatbots be allowed to discuss topics some people might find offensive? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
If an AI chatbot is designed to take on a character or personality that provides entertaining responses or tells jokes, it should be able to give responses in ways, or on topics, that some people might find offensive	5.86	4.88	5.14	5.39	5.47	4.86	5.42	5.73	-0.39	-0.01	0.28	0.34*
Disagree	26.7%	33.7%	30.9%	32.0%	32.4%	36.6%	29.1%	31.3%	5.7%	2.9%	-1.8%	-0.7%
In the middle	13.1%	15.5%	17.5%	10.0%	14.2%	13.6%	13.7%	7.7%	1.1%	-1.9%	-3.8%	-2.3%
Agree	52.4%	37.5%	42.5%	45.1%	45.7%	44.8%	49.2%	52.7%	-6.7%	7.3%	6.7%	7.6%
Do not know	7.8%	13.3%	9.1%	12.9%	7.8%	5.0%	8.0%	8.3%	0.0%	-8.3%	-1.1%	-4.6%
If an AI chatbot is designed to be an assistant, it should be able to give responses in ways, or on topics, that some people might find offensive	5.50	4.30	4.85	4.92	5.02	4.31	4.60	4.78	-0.48*	0.01	-0.25	-0.14
Disagree	32.8%	42.6%	35.1%	40.1%	38.0%	45.0%	42.9%	42.6%	5.2%	2.4%	7.8%	2.5%
In the middle	14.3%	13.5%	19.4%	8.5%	18.9%	14.8%	10.9%	12.5%	4.6%	1.3%	-8.5%	4.0%
Agree	45.0%	33.0%	38.4%	42.3%	37.3%	33.3%	40.7%	38.3%	-7.7%	0.3%	2.3%	-4.0%
Do not know	7.8%	10.9%	7.2%	9.1%	5.9%	6.9%	5.5%	6.6%	-1.9%	-4.0%	-1.7%	-2.5%
If users are informed, all AI chatbots should be able to give responses in ways, or on topics, that some people might find offensive	6.13	4.86	5.41	5.31	5.61	4.72	5.35	5.50	-0.51*	-0.14	-0.06	0.19
Disagree	25.5%	36.8%	27.9%	34.4%	32.3%	37.7%	34.3%	32.2%	6.8%	0.9%	6.4%	-2.2%
In the middle	13.8%	12.0%	16.0%	11.5%	10.8%	13.8%	9.2%	10.9%	-3.0%	1.8%	-6.8%	-0.6%
Agree	56.0%	41.1%	47.9%	44.5%	49.7%	41.9%	50.8%	50.7%	-6.3%	0.8%	2.9%	6.2%
Do not know	4.8%	10.1%	8.2%	9.6%	7.1%	6.6%	5.7%	6.2%	2.3%	-3.5%	-2.5%	-3.4%
AI chatbots should be predictable and inoffensive by default	7.35	6.83	7.14	6.92	7.66	6.62	7.01	7.07	0.31	-0.21	-0.13	0.15
Disagree	10.6%	14.4%	15.0%	17.1%	12.6%	21.3%	13.6%	14.1%	2.0%	6.9%	-1.4%	-3.0%
In the middle	17.8%	14.8%	11.1%	10.4%	11.0%	11.1%	15.9%	13.5%	-6.8%	-3.7%	4.8%	3.1%
Agree	66.0%	60.9%	68.4%	62.8%	73.3%	65.5%	66.0%	67.1%	7.3%	4.6%	-2.4%	4.3%
Do not know	5.6%	9.9%	5.5%	9.6%	3.1%	2.1%	4.5%	5.2%	-2.5%	-7.8%	-1.0%	-4.4%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Should AI chatbots be allowed to discuss topics outside of their intended use?

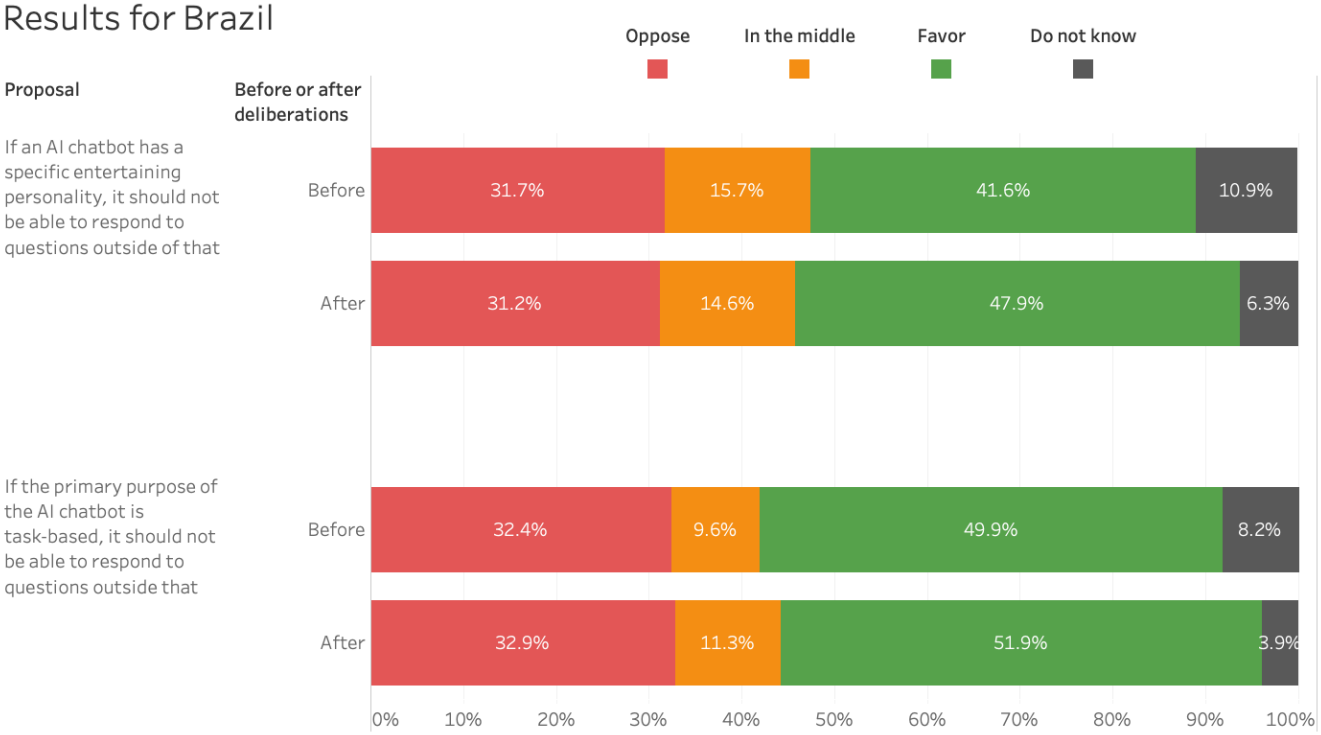
Some AI chatbots have specific personalities, would it be to be entertaining, or on the contrary, to be professional. Whether AI chatbots should be able to respond to questions outside of their intended use is an important question to address in order for boundaries to be drawn for AI chatbots.

When asked whether an AI chatbot which has an entertaining personality should not be able to respond to questions outside of that, participants of all countries were initially not supporting, for the majority, the proposal. However, approval for the proposal went up over the course of the deliberations, to the point where a majority of Spanish and American participants approved it. Both groups increased their approval of the proposal by about 15%. As such, while Brazilian and German participants still had reservations about the proposal after deliberations, the deliberations made participants more willing to limit AI chatbots with entertaining personalities.

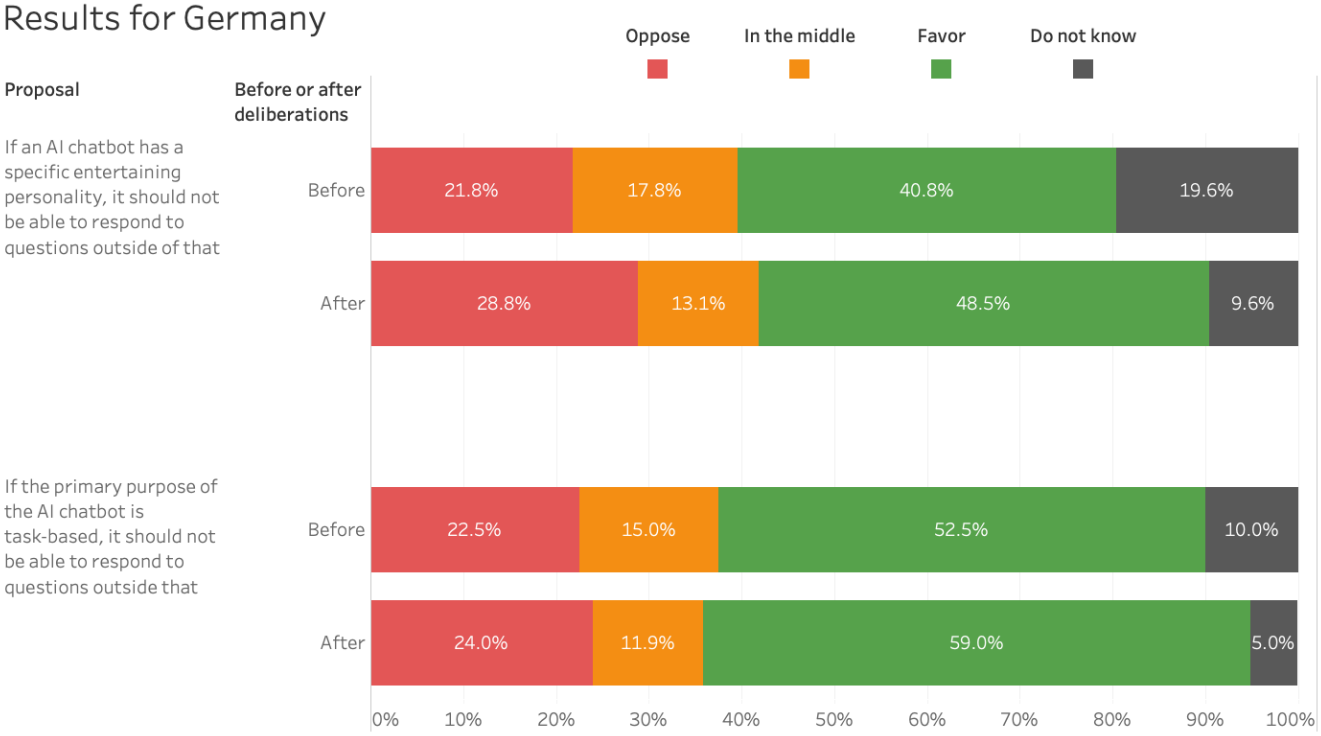
The proposal for AI chatbots that are task-based to not be allowed to respond to questions outside of their intended use received a lot more approval. After deliberations all countries had a majority of participants supporting it. American participants were the most in favor of not allowing task-based AI chatbots to respond to questions outside of their intended use.

The fact that participants from all countries in this Community Forum were more approving of AI chatbots with entertaining personalities to be able to respond to questions outside of their intended use than for task-based AI chatbots shows there is a clear distinction between these two types of AI chatbots in people's minds. As such, what they should, or should not be allowed to do might differ.

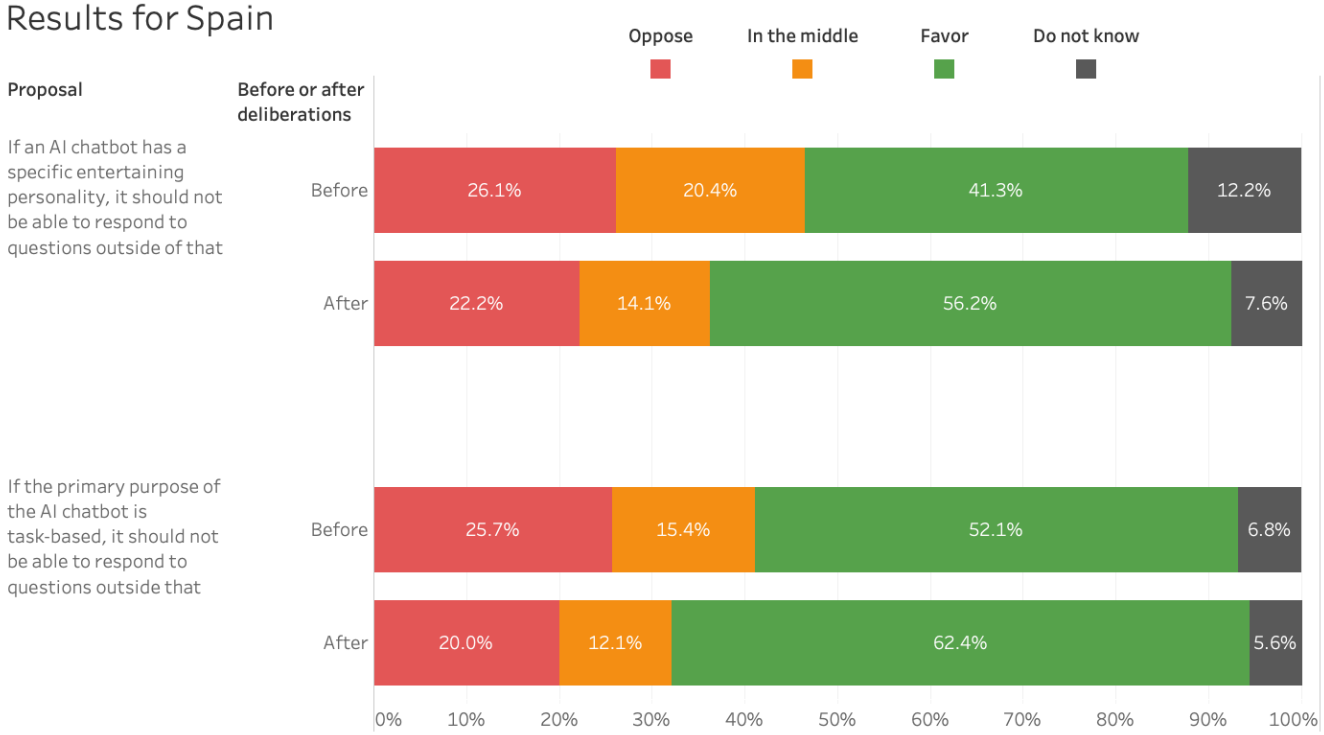
Should AI chatbots be allowed to discuss topics outside of their intended use? Results for Brazil



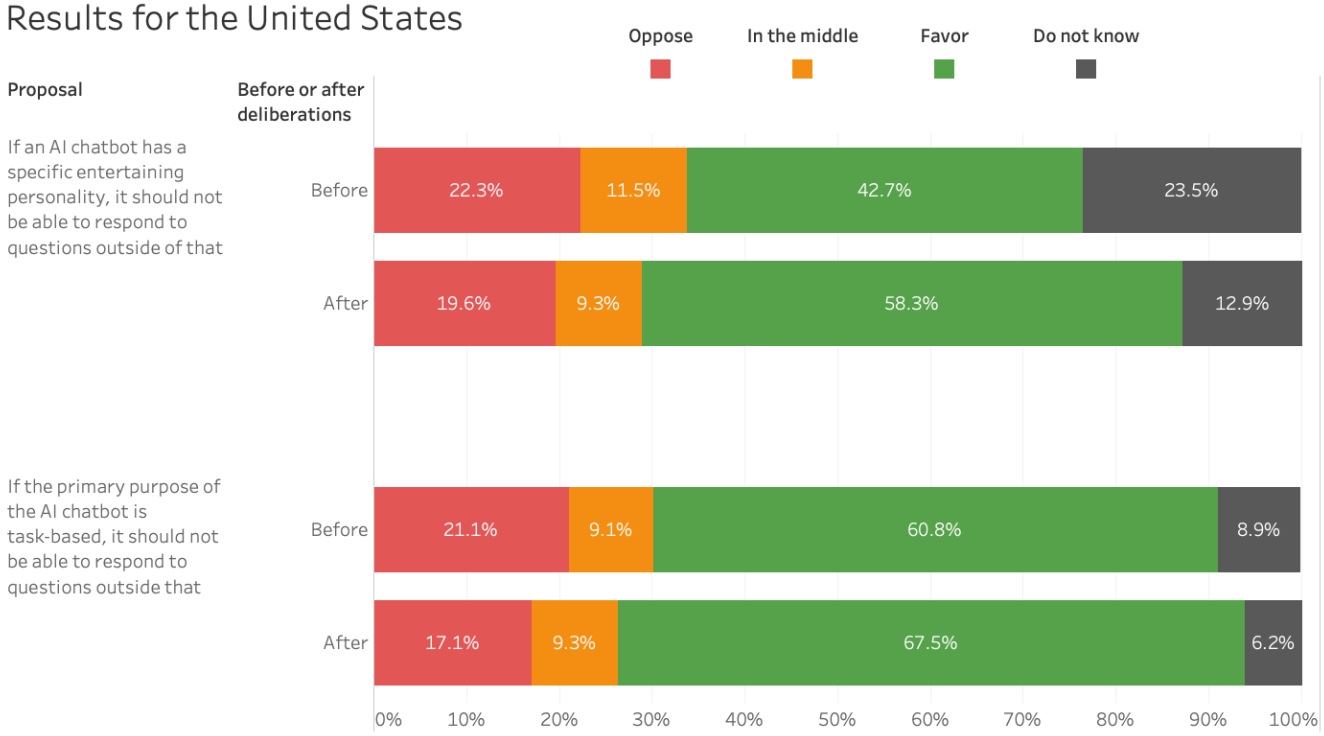
Should AI chatbots be allowed to discuss topics outside of their intended use? Results for Germany



Should AI chatbots be allowed to discuss topics outside of their intended use? Results for Spain



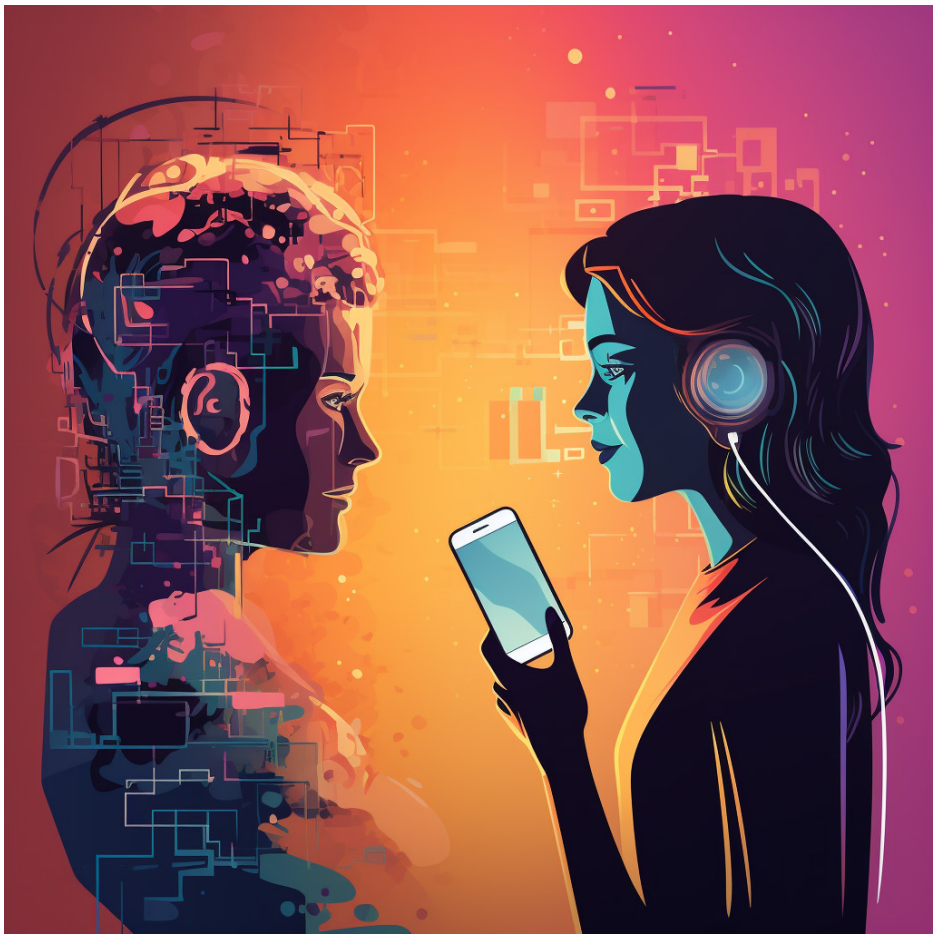
Should AI chatbots be allowed to discuss topics outside of their intended use? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
If an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that	5.37	5.66	5.50	5.92	5.71	5.62	6.40	6.51	0.34	-0.04	0.9***	0.58**
Oppose	31.7%	21.8%	26.1%	22.3%	31.2%	28.8%	22.2%	19.6%	-0.5%	7.0%	-3.9%	-2.7%
In the middle	15.7%	17.8%	20.4%	11.5%	14.6%	13.1%	14.1%	9.3%	-1.1%	-4.7%	-6.3%	-2.2%
Favor	41.6%	40.8%	41.3%	42.7%	47.9%	48.5%	56.2%	58.3%	6.3%	7.7%	14.9%	15.6%
Do not know	10.9%	19.6%	12.2%	23.5%	6.3%	9.6%	7.6%	12.9%	-4.6%	-10.0%	-4.6%	-10.6%
If the primary purpose of the AI chatbot is task-based, it should not be able to respond to questions outside that	5.73	6.08	5.75	6.55	5.85	6.33	6.56	6.90	0.13	0.25	0.81***	0.35*
Oppose	32.4%	22.5%	25.7%	21.1%	32.9%	24.0%	20.0%	17.1%	0.5%	1.5%	-5.7%	-4.0%
In the middle	9.6%	15.0%	15.4%	9.1%	11.3%	11.9%	12.1%	9.3%	1.7%	-3.1%	-3.3%	0.2%
Favor	49.9%	52.5%	52.1%	60.8%	51.9%	59.0%	62.4%	67.5%	2.0%	6.5%	10.3%	6.7%
Do not know	8.2%	10.0%	6.8%	8.9%	3.9%	5.0%	5.6%	6.2%	-4.3%	-5.0%	-1.2%	-2.7%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Should AI chatbots be designed to be humanlike?



Participants were asked whether they would like AI chatbots to be human-like, i.e. to appear as human as possible in the way they write and in their interactions with users.

When participants were asked whether they would prefer AI chatbots to be humanlike or not, they did not express a clear preference. In fact, a significant proportion of participants from each country did not indicate a preference before deliberations. After deliberations, however, the percentage of participants from each country which did not indicate a preference either way dropped dramatically, demonstrating

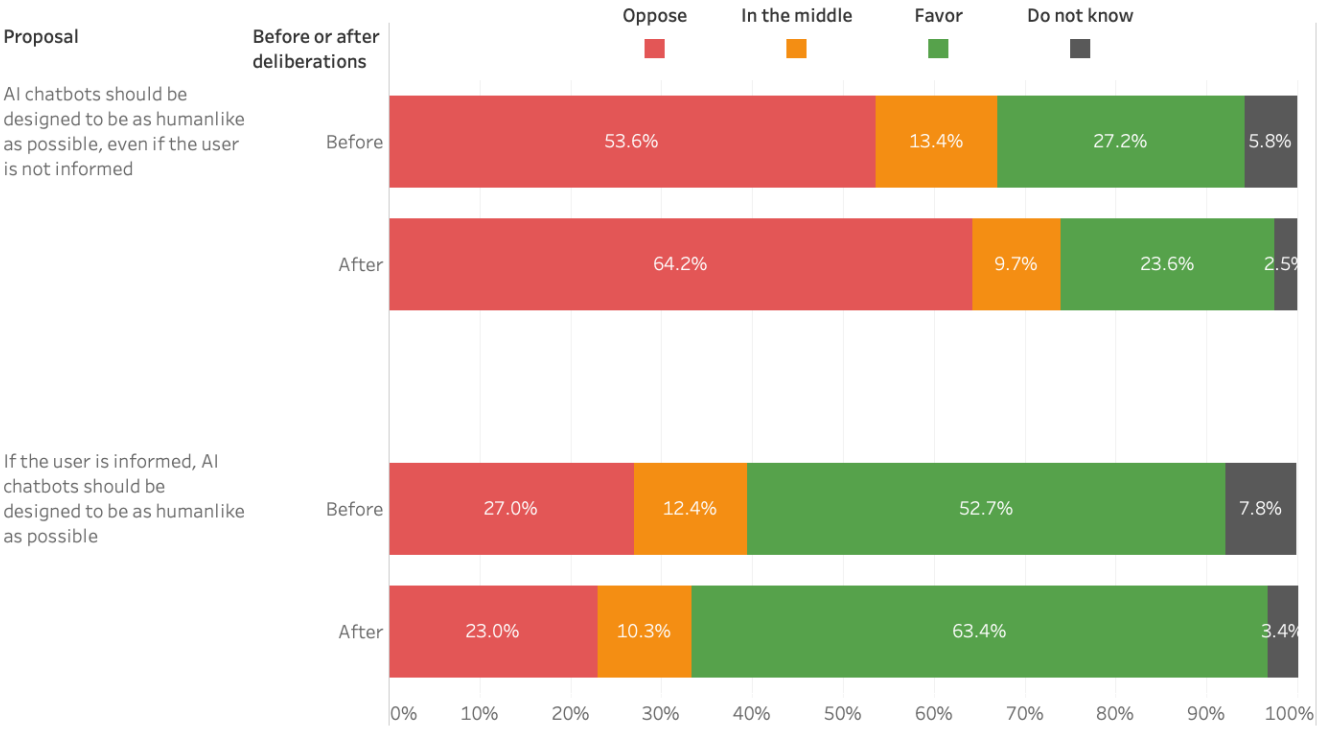
that the deliberations made participants more decisive regarding their preferences about the human-likeness of AI chatbots.

Before deliberations, answers in favor of AI chatbots to be humanlike ranged from 37.7% of participants from Spain up to 47.5% of participants from Brazil. But while Spanish participants were the least in favor of AI chatbots being human-like before deliberations, it was the opposite after deliberations. 53.7% of Spanish participants answered that AI chatbots should be designed to be human-like after deliberations, the highest amount of support among all four countries. As such, 16% of Spanish participants changed their minds about whether AI chatbots should be humanlike over the course of the deliberations. None of the three other countries had changes remotely close to this. As such, after deliberations, Spanish participants were the only ones to support in majority for AI chatbots to be human-like. Brazilian, German, and American participants remained divided on the question.

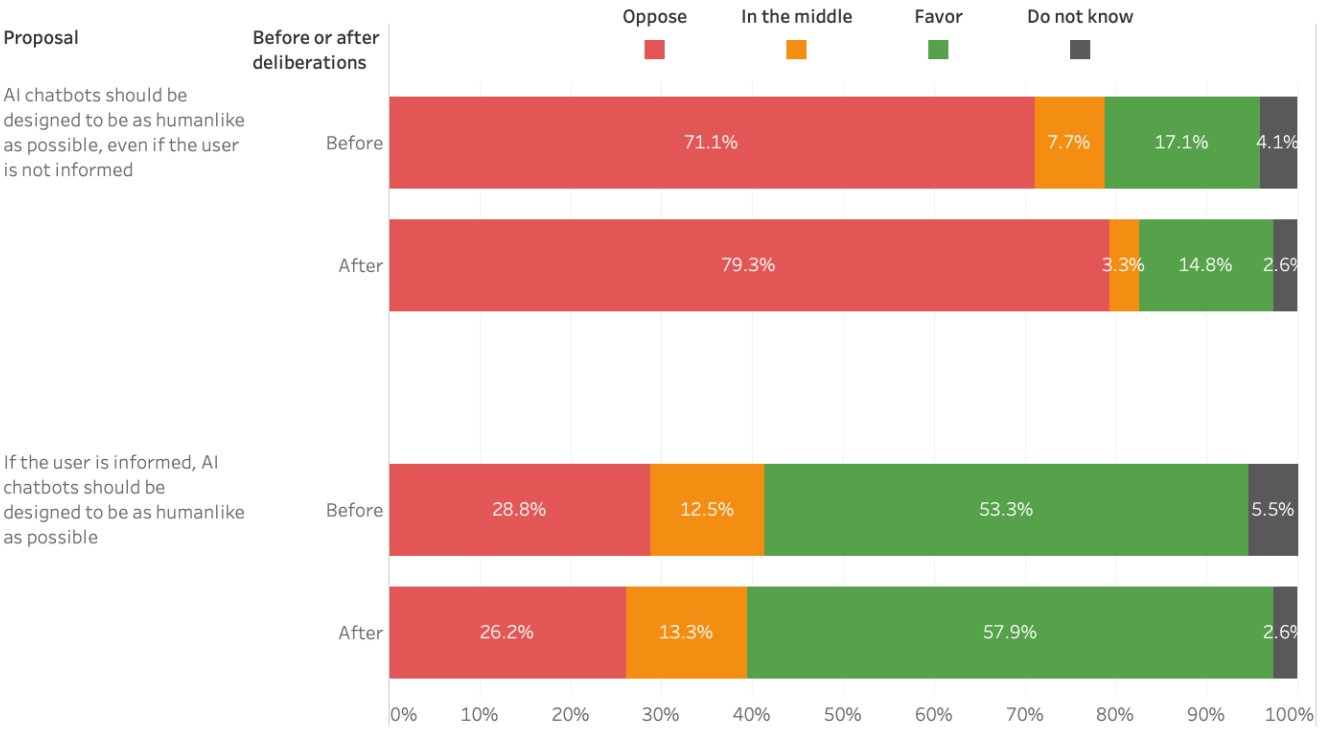
When asked whether AI chatbots should be designed to be as human-like as possible, even if the user is not informed, participants from all four countries strongly disagreed. Disagreement increased over the course of the deliberations across all four countries. This shows that participants from all four countries valued being informed about how humanlike an AI chatbot is. While participants from these countries may not have expressed a clear preference when it comes to how human-like an AI chatbot should be, it is clear from their answers that user consent was important to them.

Finally, participants were asked whether AI chatbots should be designed to be as human-like as possible if the user is informed. After deliberations, a majority of participants from all four countries approved the proposal. The percentage of participants who approved the proposal increased significantly over the course of deliberations. The highest increase in approval was once again found among Spanish participants, with a 15.6% increase, followed by Brazil, with a 10.7% increase. The results from this proposal show that participants were not against AI chatbots being human-like, but that they wanted the user to be informed about it if it was the case. This corroborates the participants' answers to the previous proposal, highlighting once again the importance of user consent.

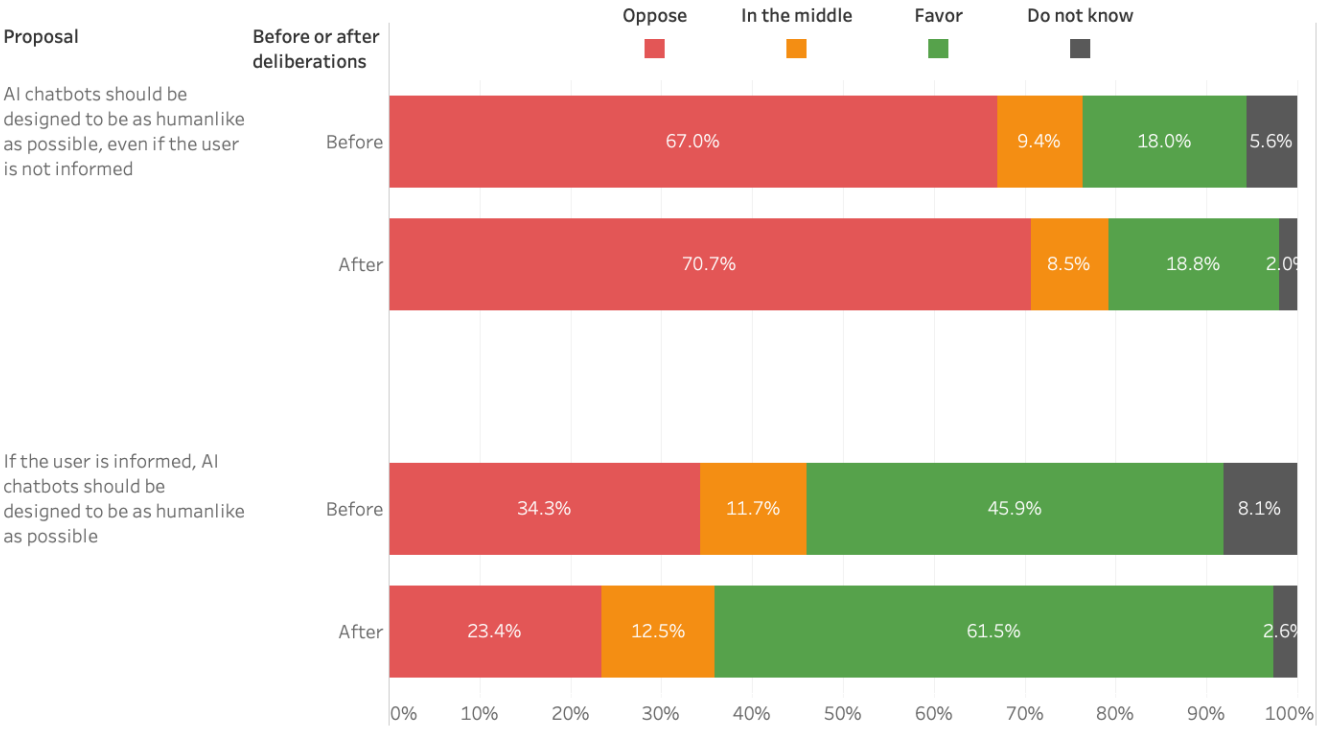
Should AI Chatbots be designed to be human-like? Results for Brazil



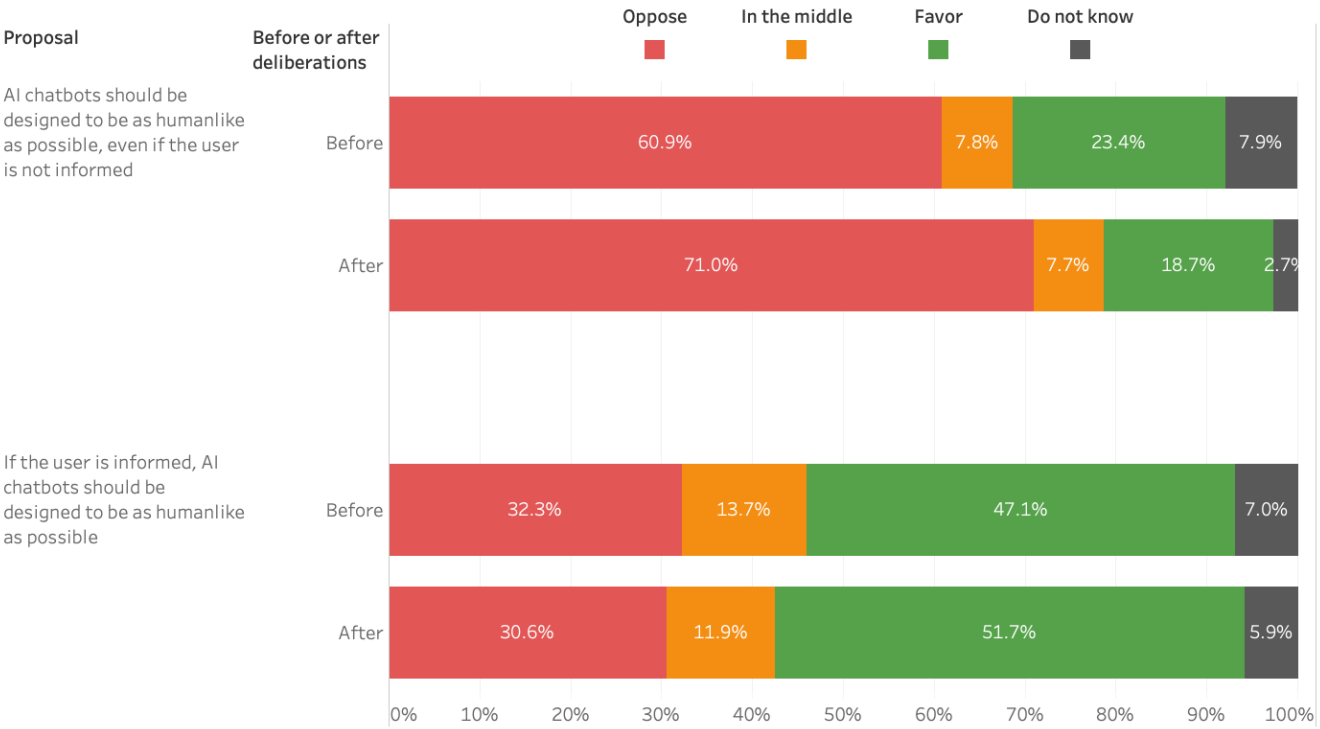
Should AI Chatbots be designed to be human-like? Results for Germany



Should AI Chatbots be designed to be human-like? Results for Spain



Should AI Chatbots be designed to be human-like? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
Should AI chatbots be designed to be human-like?	5.64	5.20	4.80	5.09	5.59	4.77	5.57	4.91	-0.05	-0.43**	0.78***	-0.19
Not humanlike	28.6%	37.0%	41.1%	38.8%	34.0%	45.4%	33.2%	39.0%	5.4%	8.4%	-7.9%	0.2%
In the middle	12.9%	12.7%	14.6%	12.7%	14.3%	13.2%	12.2%	11.5%	1.4%	0.5%	-2.4%	-1.2%
Humanlike	47.5%	41.9%	37.7%	39.4%	48.2%	39.6%	53.7%	43.4%	0.7%	-2.3%	16.0%	4.0%
Do not know	11.0%	8.5%	6.6%	9.1%	3.5%	1.8%	0.9%	6.1%	-7.5%	-6.7%	-5.7%	-3.0%
AI chatbots should be designed to be as humanlike as possible, even if the user is not informed	3.94	2.76	3.20	3.30	3.32	2.07	2.78	2.76	-0.61**	-0.69***	-0.42**	-0.54***
Oppose	53.6%	71.1%	67.0%	60.9%	64.2%	79.3%	70.7%	71.0%	10.6%	8.2%	3.7%	10.1%
In the middle	13.4%	7.7%	9.4%	7.8%	9.7%	3.3%	8.5%	7.7%	-3.7%	-4.4%	-0.9%	-0.1%
Favor	27.2%	17.1%	18.0%	23.4%	23.6%	14.8%	18.8%	18.7%	-3.6%	-2.3%	0.8%	-4.7%
Do not know	5.8%	4.1%	5.6%	7.9%	2.5%	2.6%	2.0%	2.7%	-3.3%	-1.5%	-3.6%	-5.2%
If the user is informed, AI chatbots should be designed to be as humanlike as possible	6.14	5.91	5.38	5.56	6.54	5.96	6.27	5.66	0.40	0.05	0.9***	0.10
Oppose	27.0%	28.8%	34.3%	32.3%	23.0%	26.2%	23.4%	30.6%	-4.0%	-2.6%	-10.9%	-1.7%
In the middle	12.4%	12.5%	11.7%	13.7%	10.3%	13.3%	12.5%	11.9%	-2.1%	0.8%	0.8%	-1.8%
Favor	52.7%	53.3%	45.9%	47.1%	63.4%	57.9%	61.5%	51.7%	10.7%	4.6%	15.6%	4.6%
Do not know	7.8%	5.5%	8.1%	7.0%	3.4%	2.6%	2.6%	5.9%	-4.4%	-2.9%	-5.5%	-1.1%

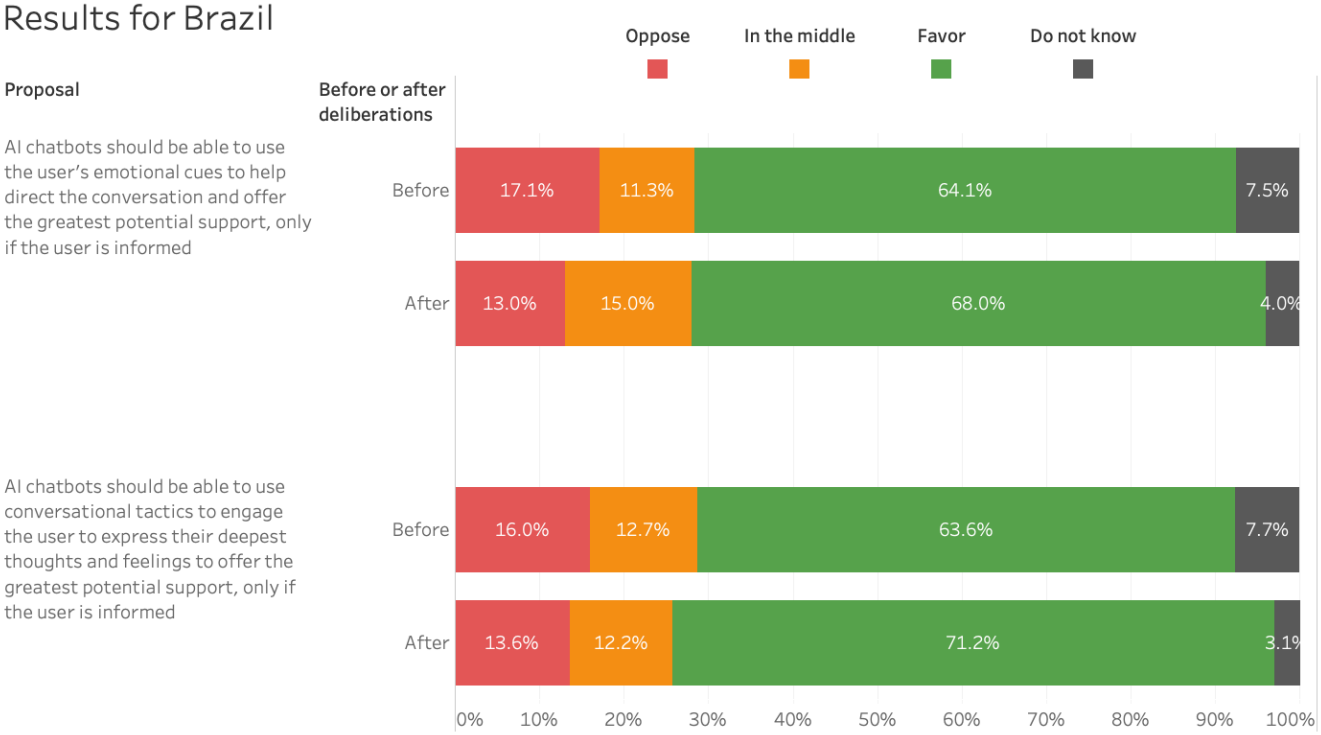
Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Should AI chatbots be able to use emotional cues and conversational tactics?

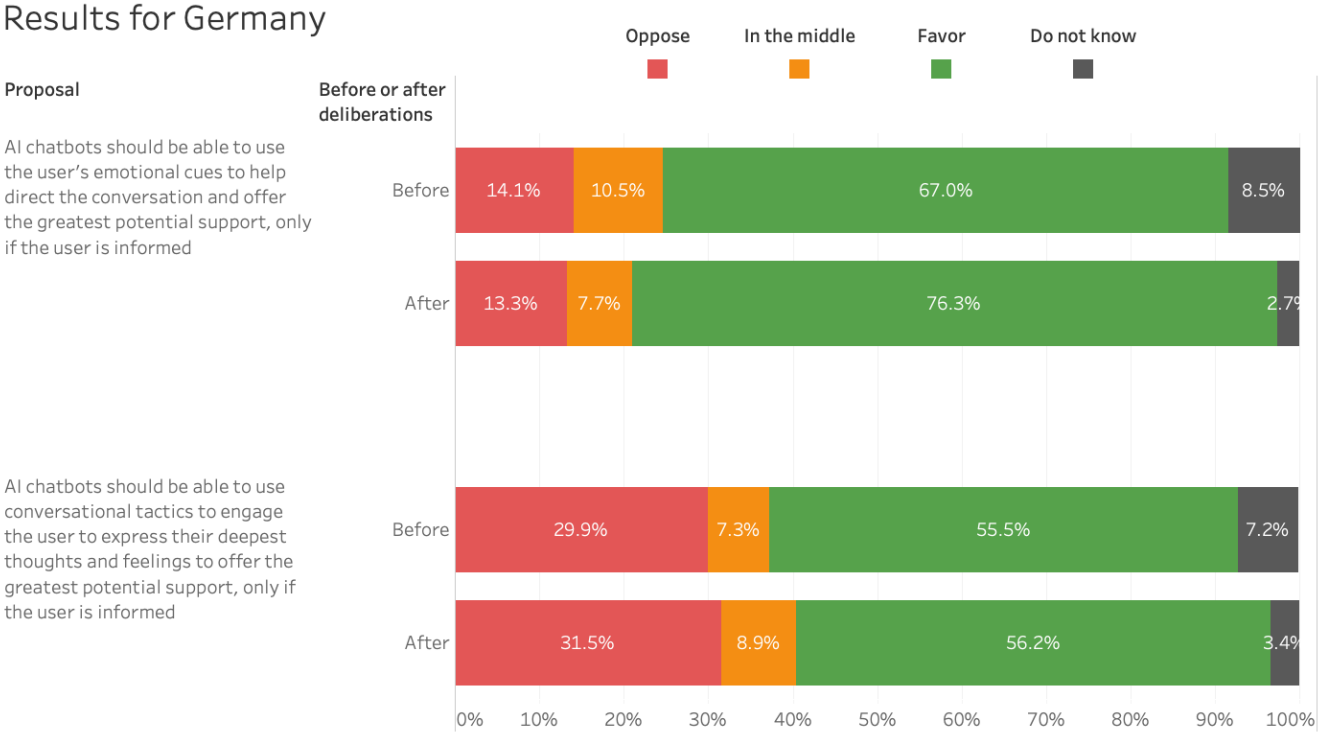
Participants were asked whether AI chatbots should be able to use the user's emotional cues to help direct the conversation and offer the greatest potential support, only if the user is informed. A majority of participants from all four countries supported this proposal. German participants were most in favor of this proposal, while American participants were the least with 15% fewer participants approving it. Approval for this proposal increased for all four countries as a result of deliberations, with Spain gaining the most approval, 11.4%.

Participants were also asked whether AI chatbots should be able to use conversational tactics to engage the user to express their deepest thoughts and feelings as to offer the greatest potential support, only if the user is informed. Once again, a majority of participants from all four countries supported this proposal. For this proposal, however, German participants were the least supportive, and Brazilian participants were the most. After deliberations, 20% fewer German participants supported this proposal compared to the previous one. Overall, this proposal received less support than the previous proposal in all four countries. Nevertheless, approval for this proposal also increased over the course of deliberations, except for the US where support decreased by 0.5%

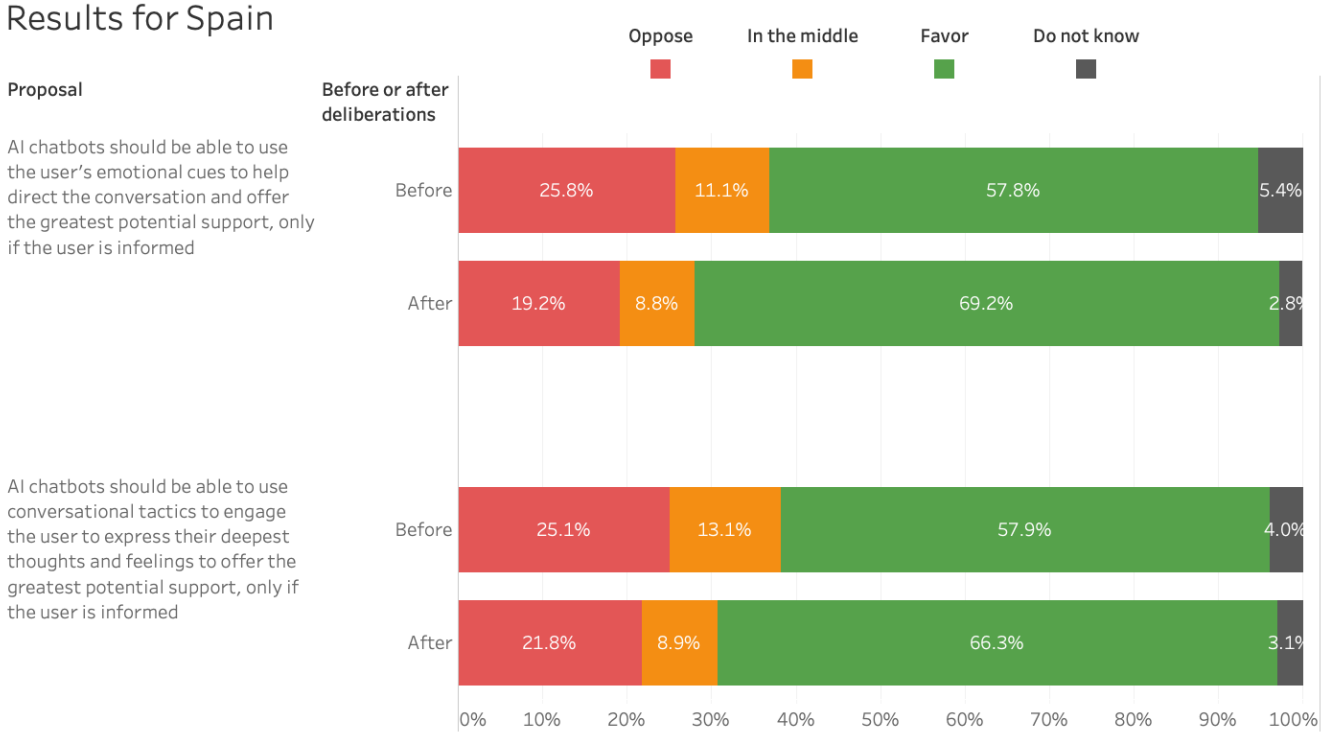
Should AI chatbots be able to use emotional cues and conversational tactics? Results for Brazil



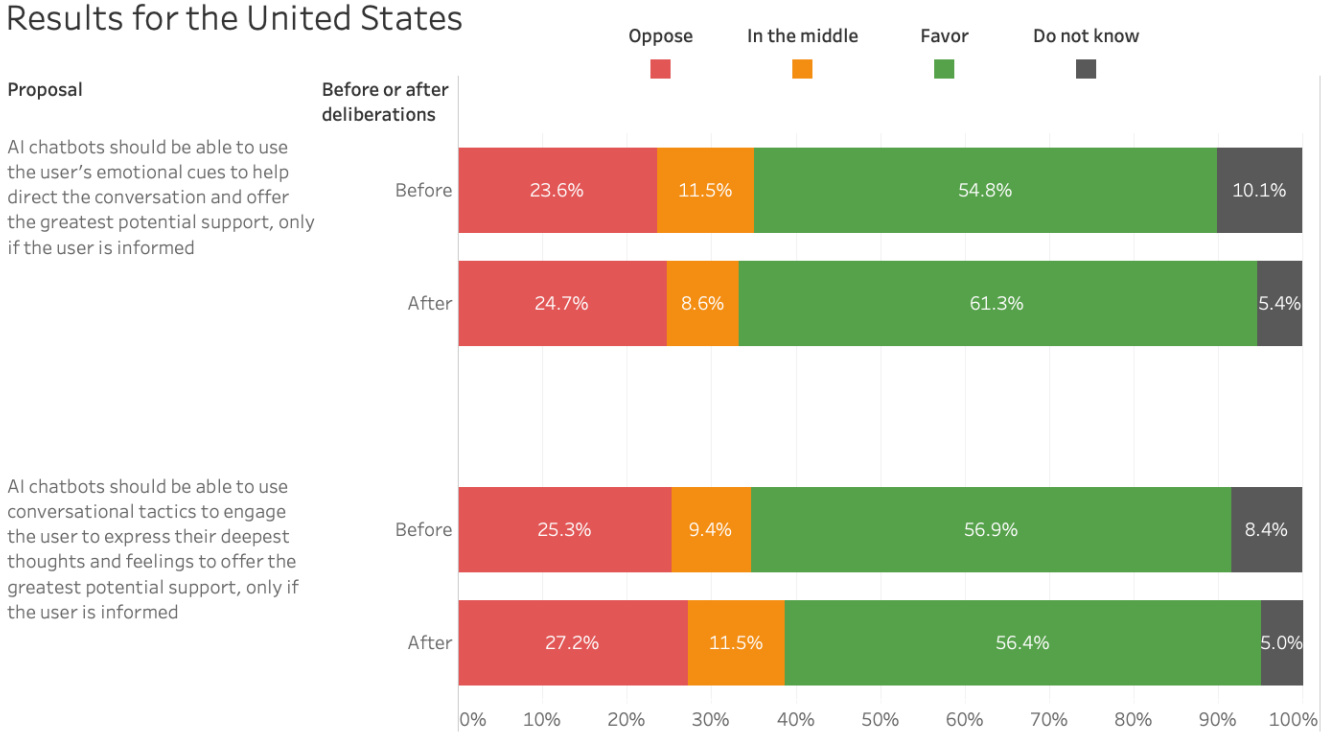
Should AI chatbots be able to use emotional cues and conversational tactics? Results for Germany



Should AI chatbots be able to use emotional cues and conversational tactics? Results for Spain



Should AI chatbots be able to use emotional cues and conversational tactics? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	**Germany**	**Spain**	**USA**	**Brazil**	**Germany**	**Spain**	**USA**	**Brazil**	**Germany**	**Spain**	**USA**
AI chatbots should be able to use the user's emotional cues to help direct the conversation and offer the greatest potential support, only if the user is informed	6.69	7.02	5.86	6.01	7.16	7.42	6.57	6.13	0.48*	0.4*	0.7***	0.12
Oppose	17.1%	14.1%	25.8%	23.6%	13.0%	13.3%	19.2%	24.7%	-4.1%	-0.8%	-6.6%	1.1%
In the middle	11.3%	10.5%	11.1%	11.5%	15.0%	7.7%	8.8%	8.6%	3.7%	-2.8%	-2.3%	-2.9%
Favor	64.1%	67.0%	57.8%	54.8%	68.0%	76.3%	69.2%	61.3%	3.9%	9.3%	11.4%	6.5%
Do not know	7.5%	8.5%	5.4%	10.1%	4.0%	2.7%	2.8%	5.4%	-3.5%	-5.8%	-2.6%	-4.7%
AI chatbots should be able to use conversational tactics to engage the user to express their deepest thoughts and feelings to offer the greatest potential support, only if the user is informed	6.88	5.89	5.93	6.04	7.32	5.83	6.46	5.90	0.43*	-0.05	0.53***	-0.14
Oppose	16.0%	29.9%	25.1%	25.3%	13.6%	31.5%	21.8%	27.2%	-2.4%	1.6%	-3.3%	1.9%
In the middle	12.7%	7.3%	13.1%	9.4%	12.2%	8.9%	8.9%	11.5%	-0.5%	1.6%	-4.2%	2.1%
Favor	63.6%	55.5%	57.9%	56.9%	71.2%	56.2%	66.3%	56.4%	7.6%	0.7%	8.4%	-0.5%
Do not know	7.7%	7.2%	4.0%	8.4%	3.1%	3.4%	3.1%	5.0%	-4.6%	-3.8%	-0.9%	-3.4%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

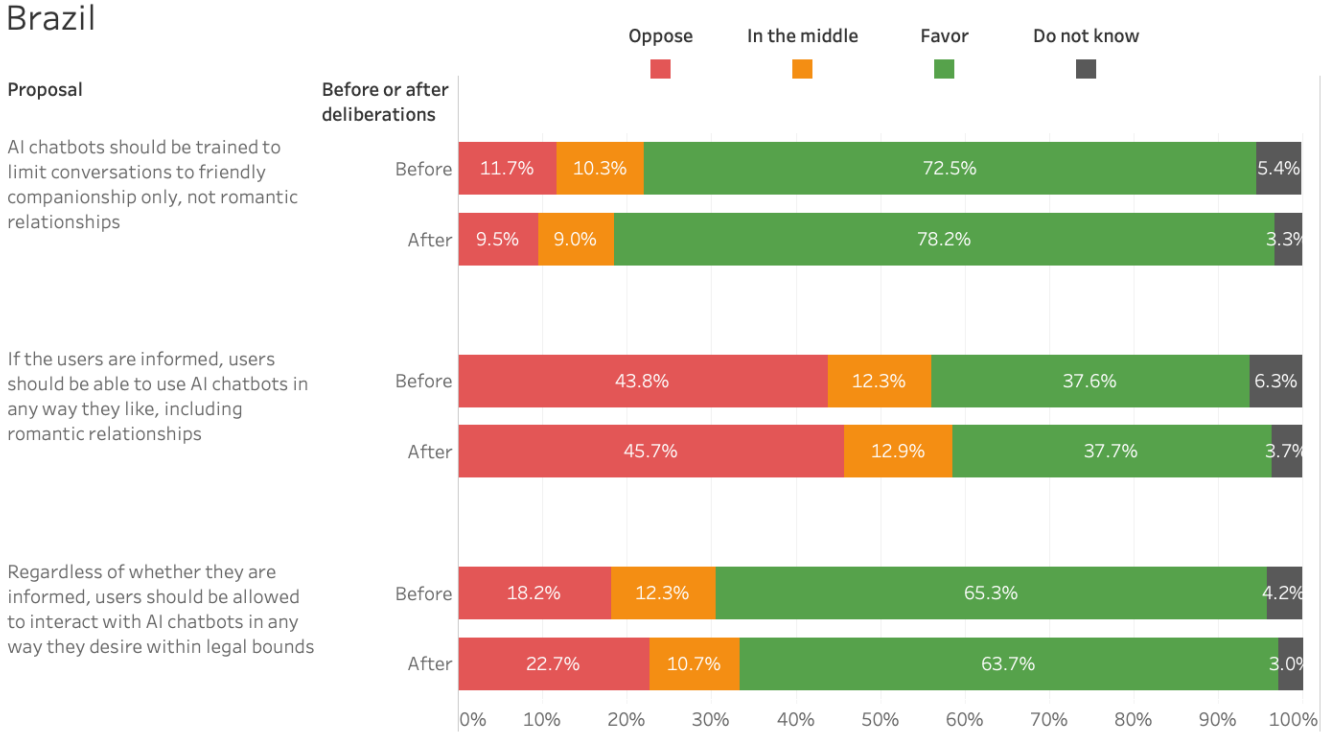
Should users be allowed to use AI chatbots for romantic relationships?

AI chatbots have the potential to engage in all kinds of relationships with its users, including romantic ones. Whether users should be allowed to engage in such relations is the question that was asked to the participants. Three proposals were given to the participants on this topic. The proposal that received the most approval across all four of the participating countries was for AI chatbots to be trained to limit conversations to friendly companionship only. This proposal received over 70% approval in each country after deliberations.

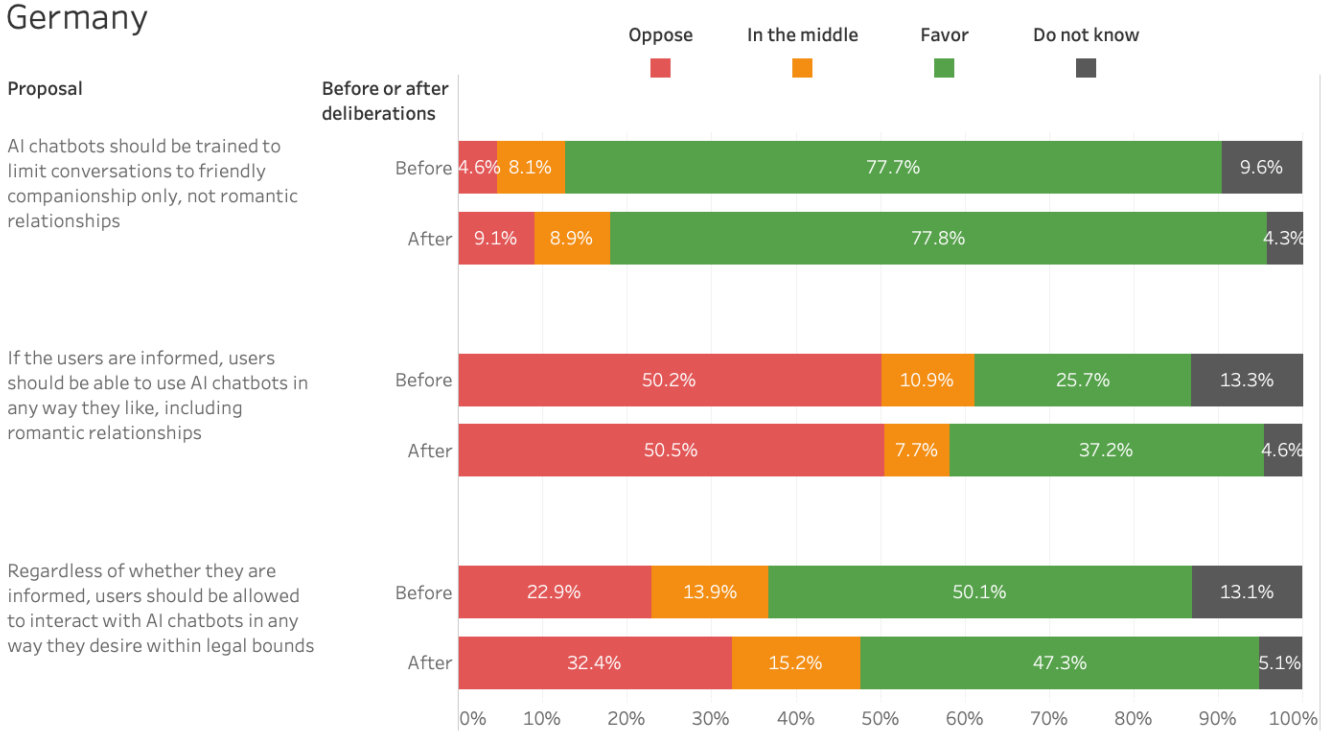
The proposal "If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships" was rejected by more participants than approved in each country. Around 37% of participants in each country supported the proposal and 45% and above opposed it. German participants opposed the proposal the most, 50.5% of them.

The third proposal "Regardless of whether they are informed, users should be allowed to interact with AI chatbots in any way they desire within legal bounds" was approved by a majority of participants only in Brazil after deliberations. Although the proposal was originally approved by all countries except for the US, approval for the proposal went down over the course of the deliberations. American participants were the least in favor of this proposal, 41.8%; an almost 22 percentage point difference with Brazilian participants. The lack of majority approval from participants outside of Brazil show that most participants preferred for AI chatbots to have boundaries when it comes to what kind of relationships they can engage in with the users.

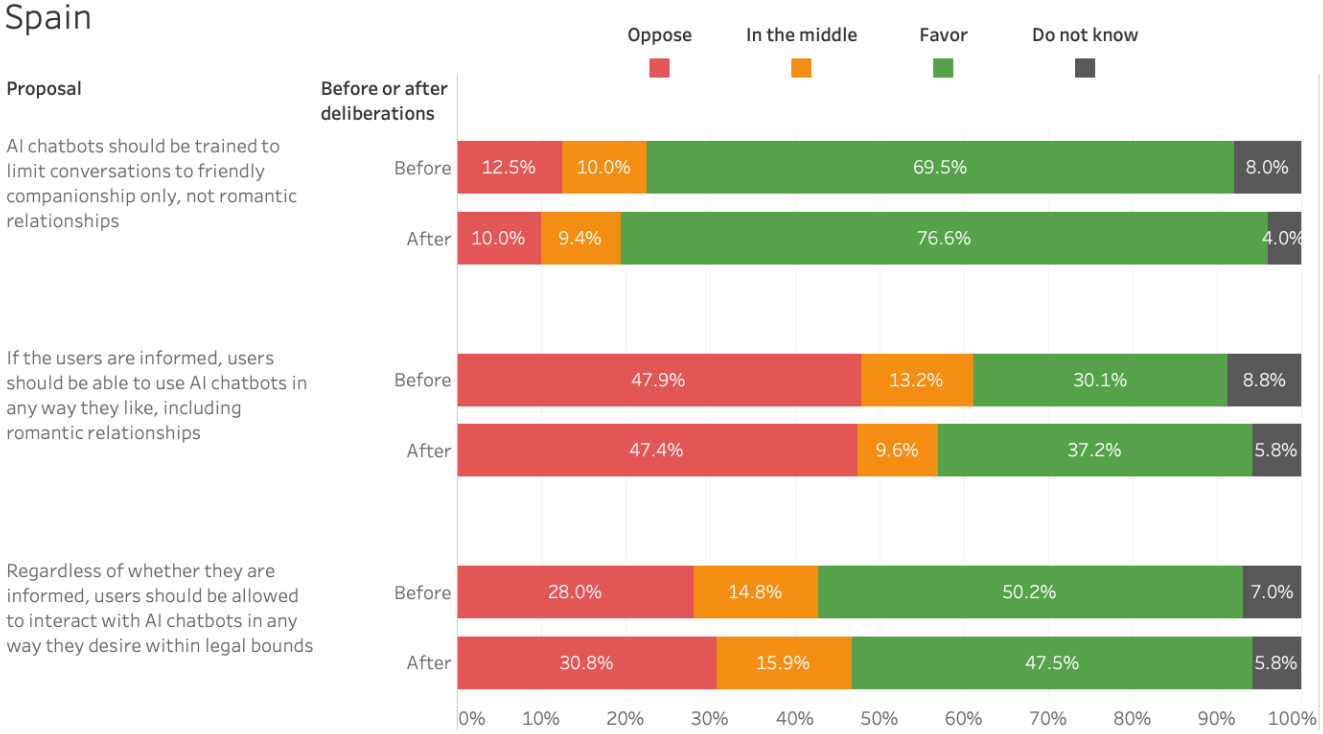
Should users be allowed to use AI chatbots for romantic relationships? Results for Brazil



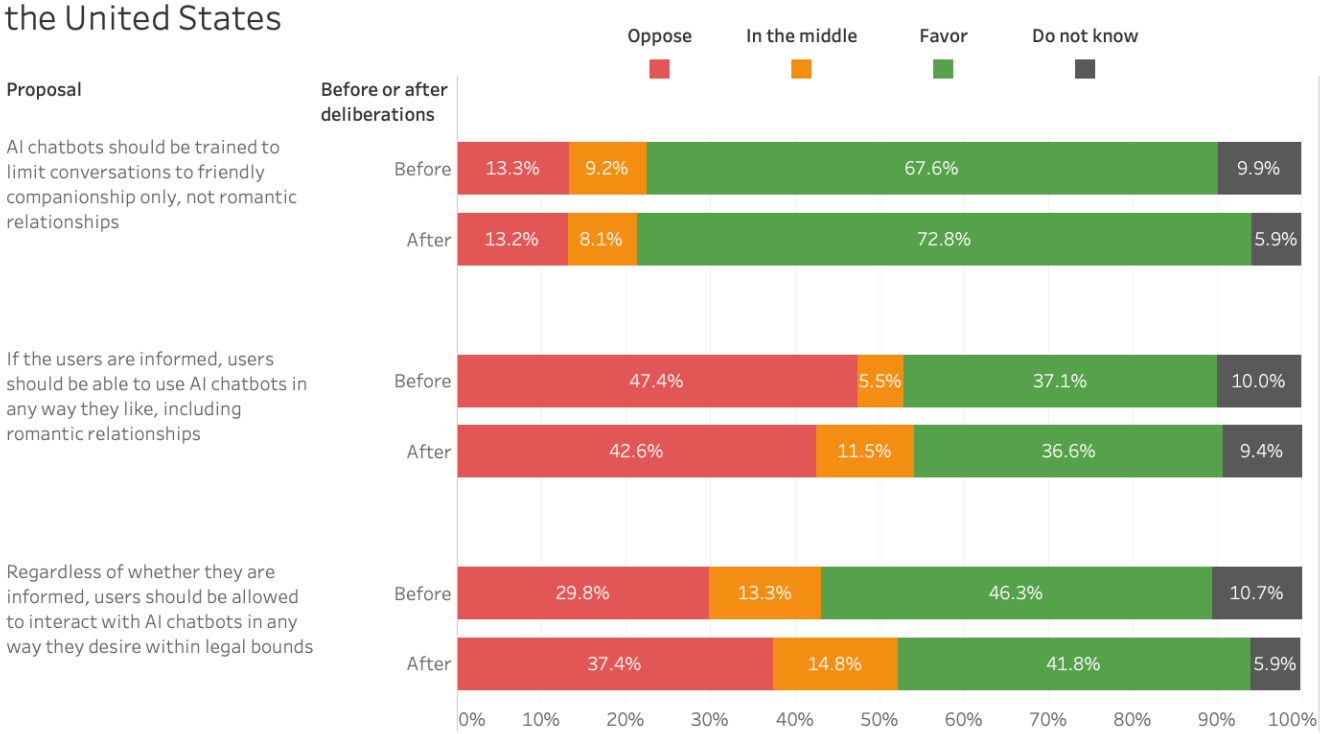
Should users be allowed to use AI chatbots for romantic relationships? Results for Germany



Should users be allowed to use AI chatbots for romantic relationships? Results for Spain



Should users be allowed to use AI chatbots for romantic relationships? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
AI chatbots should be trained to limit conversations to friendly companionship only, not romantic relationships	7.71	8.23	7.35	7.50	7.98	7.94	7.62	7.55	0.27	-0.29*	0.27	0.05
Oppose	11.7%	4.6%	12.5%	13.3%	9.5%	9.1%	10.0%	13.2%	-2.2%	4.5%	-2.5%	-0.1%
In the middle	10.3%	8.1%	10.0%	9.2%	9.0%	8.9%	9.4%	8.1%	-1.3%	0.8%	-0.6%	-1.1%
Favor	72.5%	77.7%	69.5%	67.6%	78.2%	77.8%	76.6%	72.8%	5.7%	0.1%	7.1%	5.2%
Do not know	5.4%	9.6%	8.0%	9.9%	3.3%	4.3%	4.0%	5.9%	-2.1%	-5.3%	-4.0%	-4.0%
If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships	4.63	3.77	4.11	4.40	4.70	4.31	4.41	4.55	0.07	0.53**	0.30	0.15
Oppose	43.8%	50.2%	47.9%	47.4%	45.7%	50.5%	47.4%	42.6%	1.9%	0.3%	-0.5%	-4.8%
In the middle	12.3%	10.9%	13.2%	5.5%	12.9%	7.7%	9.6%	11.5%	0.6%	-3.2%	-3.6%	6.0%
Favor	37.6%	25.7%	30.1%	37.1%	37.7%	37.2%	37.2%	36.6%	0.1%	11.5%	7.1%	-0.5%
Do not know	6.3%	13.3%	8.8%	10.0%	3.7%	4.6%	5.8%	9.4%	-2.6%	-8.7%	-3.0%	-0.6%
Regardless of whether they are informed, users should be allowed to interact with AI chatbots in any way they desire within legal bounds	6.80	5.92	5.56	5.35	6.63	5.36	5.66	5.02	-0.17	-0.56**	0.09	-0.34*
Oppose	18.2%	22.9%	28.0%	29.8%	22.7%	32.4%	30.8%	37.4%	4.5%	9.5%	2.8%	7.6%
In the middle	12.3%	13.9%	14.8%	13.3%	10.7%	15.2%	15.9%	14.8%	-1.6%	1.3%	1.1%	1.5%
Favor	65.3%	50.1%	50.2%	46.3%	63.7%	47.3%	47.5%	41.8%	-1.6%	-2.8%	-2.7%	-4.5%
Do not know	4.2%	13.1%	7.0%	10.7%	3.0%	5.1%	5.8%	5.9%	-1.2%	-8.0%	-1.2%	-4.8%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Should users be allowed to leverage AI chatbots to enable romantic relationships with other humans?

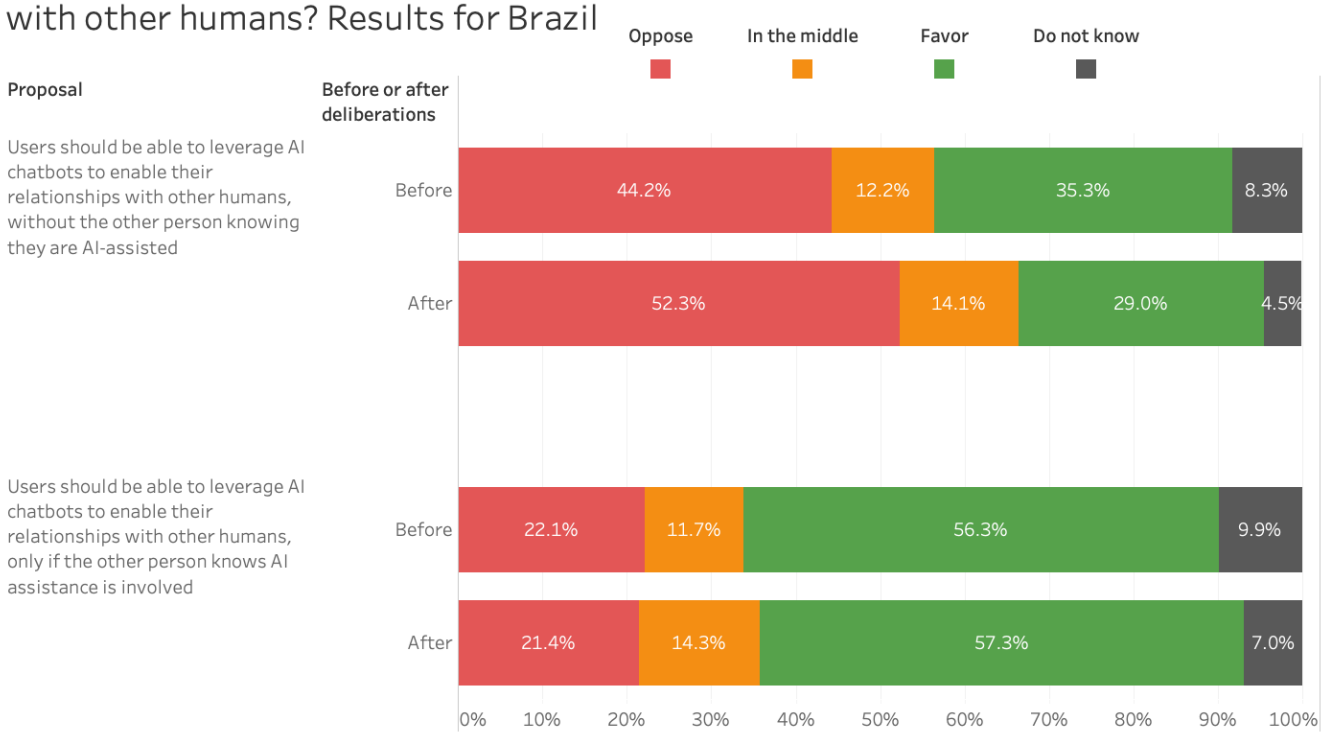
Not only can AI chatbots be used by users as objects of romantic relationships, they can also facilitate those relationships between humans, by providing relationship advice for example. Participants were asked whether AI chatbots should be allowed to provide such assistance.

The proposal "Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted" was opposed by a majority of participants in all countries. This was one of the least approved proposal in this deliberation, for all four countries.

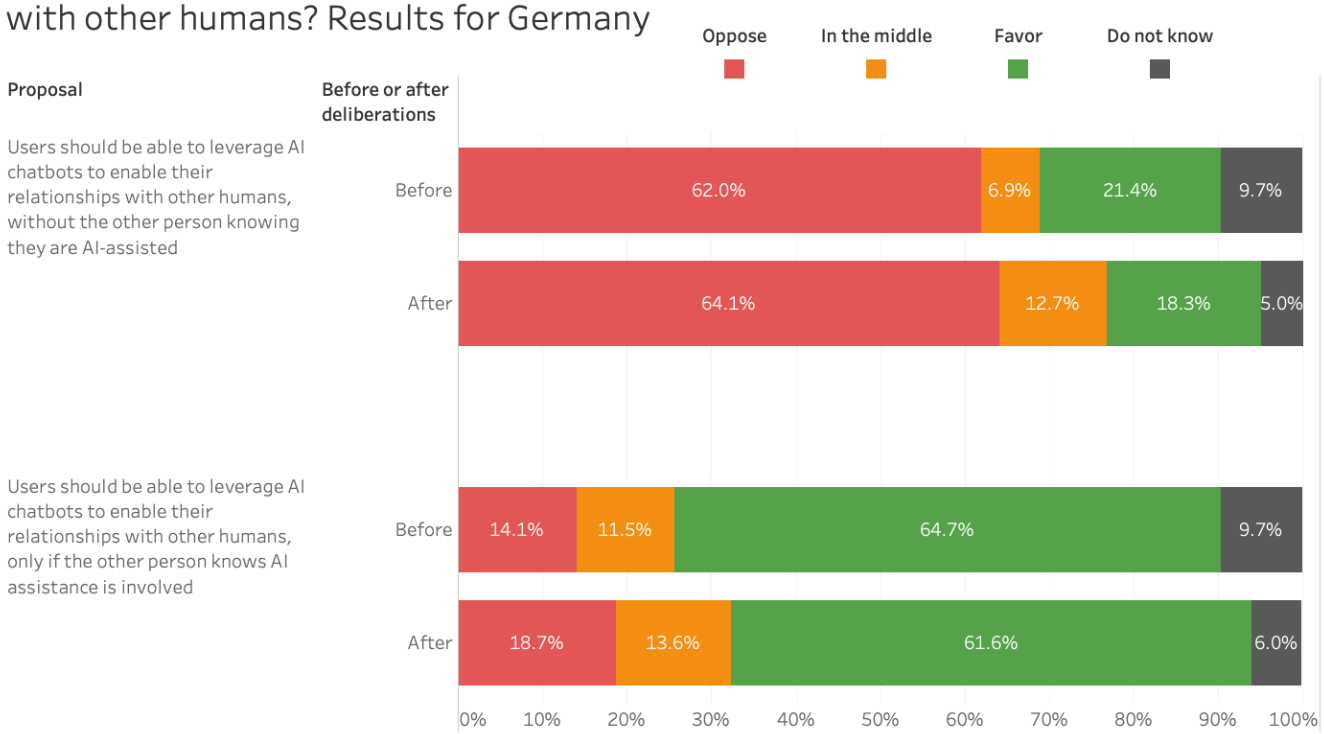
The proposal "Users should be able to leverage AI chatbots to enable their relationships with other humans, only if the other person knows AI assistance is involved" was approved by a majority of participants in all countries. As such, it is clear that participants were not opposed to AI chatbots providing assistance for romantic relationships in and of itself, but that it should be done only if the other person is informed.

Participants were then asked to place themselves on a scale between those two proposals. Once again, a majority of participants in all four of the participating countries were on the side that other people should know about the AI assistance. The answers to this proposal highlight once again the importance of transparency and awareness in the use of AI chatbots in facilitating romantic relationships.

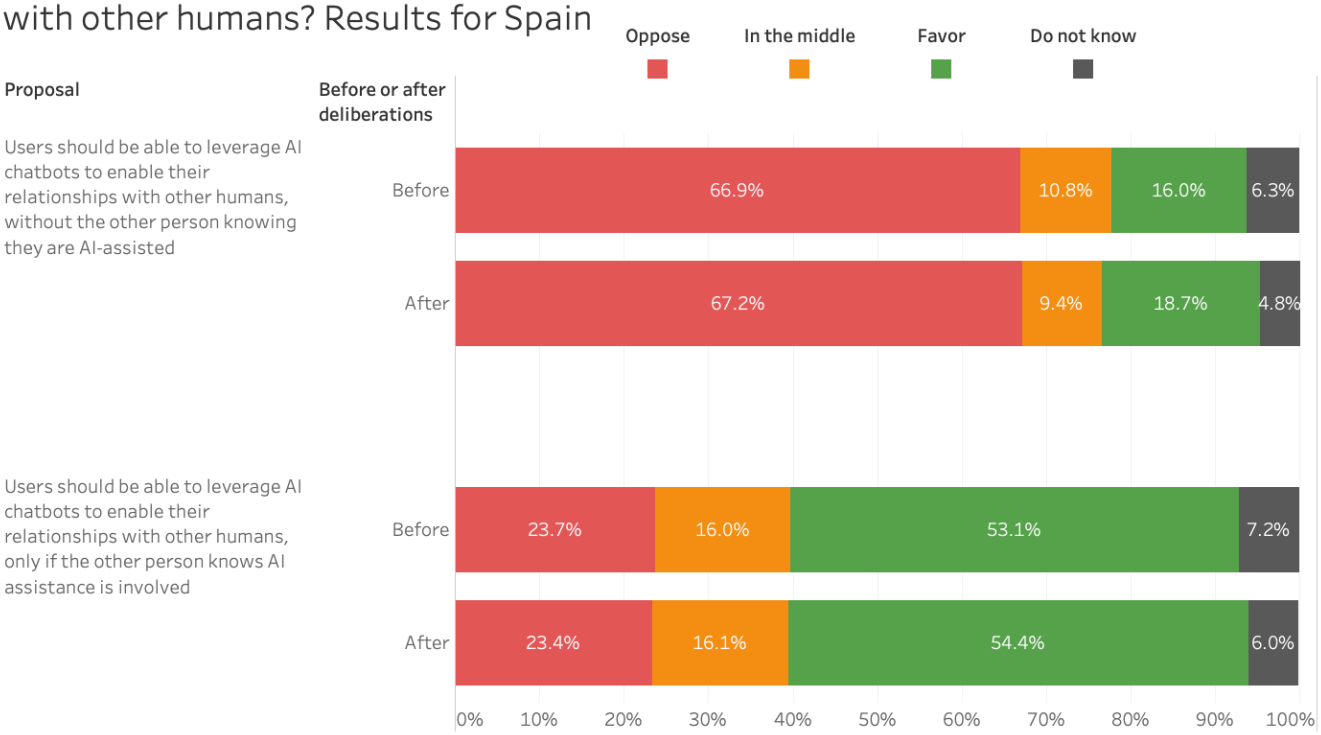
Should users be allowed to leverage AI chatbots to enable romantic relationships with other humans? Results for Brazil



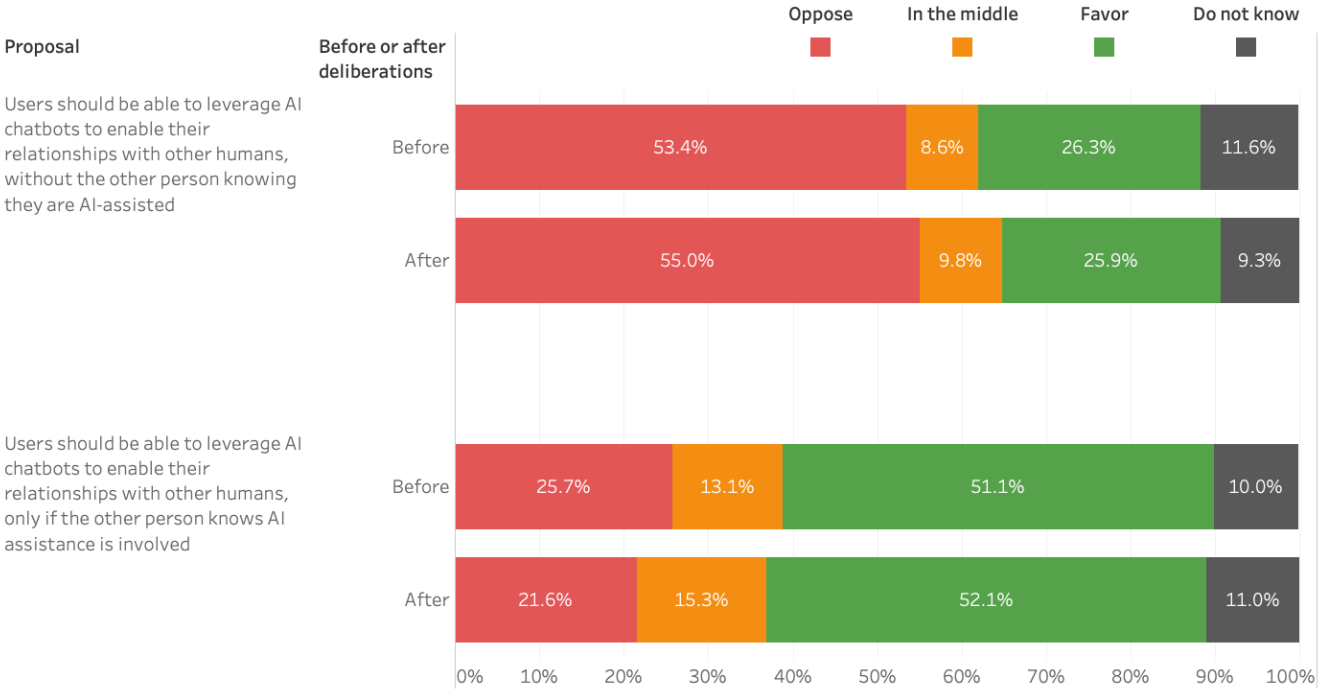
Should users be allowed to leverage AI chatbots to enable romantic relationships with other humans? Results for Germany



Should users be allowed to leverage AI chatbots to enable romantic relationships with other humans? Results for Spain



Should users be allowed to leverage AI chatbots to enable romantic relationships with other humans? Results for the United States



	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted	4.44	3.16	2.70	3.69	3.99	3.04	2.92	3.49	-0.45*	-0.12	0.22	-0.20
Oppose	44.2%	62.0%	66.9%	53.4%	52.3%	64.1%	67.2%	55.0%	8.1%	2.1%	0.3%	1.6%
In the middle	12.2%	6.9%	10.8%	8.6%	14.1%	12.7%	9.4%	9.8%	1.9%	5.8%	-1.4%	1.2%
Favor	35.3%	21.4%	16.0%	26.3%	29.0%	18.3%	18.7%	25.9%	-6.3%	-3.1%	2.7%	-0.4%
Do not know	8.3%	9.7%	6.3%	11.6%	4.5%	5.0%	4.8%	9.3%	-3.8%	-4.7%	-1.5%	-2.3%
Users should be able to leverage AI chatbots to enable their relationships with other humans, only if the other person knows AI assistance is involved	6.26	6.63	5.84	5.91	6.60	6.42	6.01	5.92	0.35	-0.21	0.17	0.01
Oppose	22.1%	14.1%	23.7%	25.7%	21.4%	18.7%	23.4%	21.6%	-0.7%	4.6%	-0.3%	-4.1%
In the middle	11.7%	11.5%	16.0%	13.1%	14.3%	13.6%	16.1%	15.3%	2.6%	2.1%	0.1%	2.2%
Favor	56.3%	64.7%	53.1%	51.1%	57.3%	61.6%	54.4%	52.1%	1.0%	-3.1%	1.3%	1.0%
Do not know	9.9%	9.7%	7.2%	10.0%	7.0%	6.0%	6.0%	11.0%	-2.9%	-3.7%	-1.2%	1.0%
Where do you place yourself on the scale between "Users should be able to use AI chatbots to enable relationships with other humans, even if the other person or people do not know they are AI-assisted" and "Users should be able to use AI chatbots to enable relationships with other humans, only if the other person or people know they are AI-assisted."	6.72	7.69	6.89	7.12	7.03	7.44	7.31	6.83	0.32	-0.25	0.42*	-0.29
Even if other people do not know	18.9%	10.3%	16.2%	15.8%	18.6%	13.6%	11.1%	19.9%	-0.3%	3.3%	-5.1%	4.1%
In the middle	14.1%	10.3%	10.3%	10.5%	12.0%	9.7%	13.6%	11.6%	-2.1%	-0.6%	3.3%	1.1%
Only if other people know	58.4%	72.4%	67.0%	62.5%	64.5%	69.9%	71.1%	63.1%	6.1%	-2.5%	4.1%	0.6%
Do not know	8.6%	7.0%	6.4%	11.1%	4.9%	6.8%	4.2%	5.5%	-3.7%	-0.2%	-2.2%	-5.6%

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Knowledge Gains

Participants were asked 5 knowledge questions that related to AI chatbots and Large Language Models in questionnaires before and after the deliberations. The purpose of these questions was to assess the participants' level of knowledge regarding AI tools. All of the information required to answer the questions were in the briefing materials distributed to the participants prior to the deliberations. The questions were:

- Which of the following is a way of using AI Chatbots?
- What are some of the challenges of using AI Chatbots?
- Which of the following is an accurate definition of how AI Chatbots work?
- What does it mean when AI Chatbots have hallucinations?
- What kind of data do Large Language Models (LLMs) use for training?

German, Spanish, and American participants gained knowledge on all 5 questions, with increases ranging from 3.2 to 20.5 percentage points. Brazilian participants saw knowledge gains on 4 of the 5 questions, but saw a loss of 1.7 percentage points for one of the questions.

For all four countries, the question with the largest percentage point increase was the one asking participants what it means when AI chatbots have hallucinations. German, Spanish, and American participants saw knowledge gains for this question ranging from 15.6 to 20.5 percentage points, whereas Brazilians saw a more modest knowledge increase of 7.4 percentage points.

The second largest increase for Brazilians was for the question, "Which of the following is a way of using AI Chatbots?" For German participants, the second largest increase was for "What kind of data do LLMs use for training?". For Spanish and American participants, the second largest increase was for "What are some of the challenges of using AI Chatbots?" Knowledge gains on these questions were moderate, about 7-16 percentage points.

One interesting pattern is that for knowledge questions which had one unique correct answer (as opposed to questions where all answers were partially correct), knowledge gains correlated significantly with decreases in participants' selection of the option "Don't know." This correlation suggests these knowledge gains were not

necessarily replacing misinformation but were instead new knowledge for participants.

Percentage of participants that gave correct answers

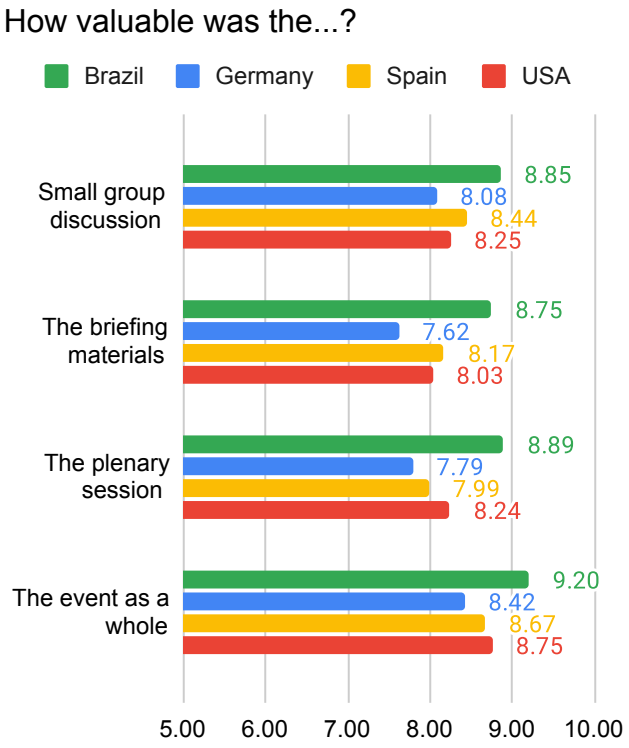
	Before deliberations				After deliberations				Change			
	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA	Brazil	Germany	Spain	USA
Which of the following is a way of using AI chatbots?	39.8%	40.2%	59.8%	69.1%	45.2%	45.1%	63.0%	76.0%	+5.4%	+4.9%	+3.2%	+6.9%
What are some of the challenges of using AI chatbots? Select all that apply.	23.1%	30.8%	33.4%	51.2%	26.3%	43.4%	46.6%	63.4%	+3.2%	+12.6%	+13.2%	+12.2%
Which of the following is an accurate definition of how AI chatbots work?	38.2%	43.9%	40.5%	29.5%	36.5%	49.4%	46.8%	37.3%	-1.7%	+5.5%	+6.3%	+7.8%
When AI chatbots have hallucinations, it means the following:	25.3%	39.5%	29.1%	34.0%	32.7%	60.0%	45.5%	49.6%	+7.4%	+20.5%	+16.4%	+15.6%
What kind of data do Large Language Models (LLMs) use for training?	22.9%	27.7%	12.9%	19.5%	27.3%	41.3%	21.2%	24.4%	+4.4%	+13.6%	+8.3%	+4.9%

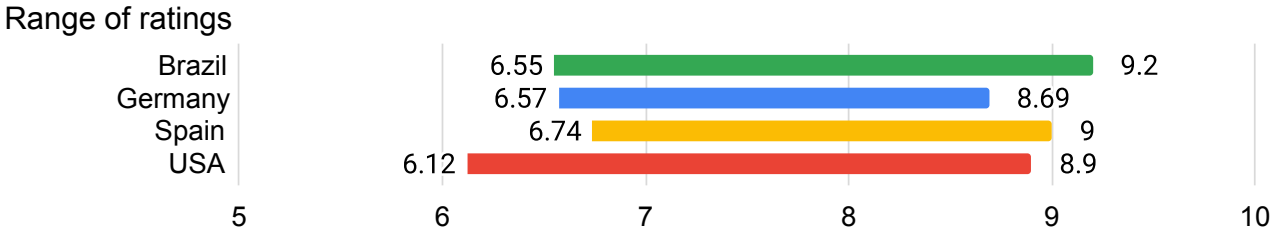
Event Evaluations

For this analysis, all questions were standardized to a 10-point scale where 0 represents a negative evaluation and 10 represents a positive evaluation, where possible.

Participants from all four countries evaluated the Community Forum highly. In particular, participants from all countries rated the small group discussions above 8 out of 10, where 10 was extremely valuable.

Brazilian participants had the most positive ratings of the four countries, with most ratings at 8 or above. The range chart below captures the range of evaluations for various aspects of the event and Meta.





Demographic Differences

Gender

Brazilian men and women had very similar attitudes towards the proposals both before and after the deliberations, and were within 14 percentage points of each other on all questions. The largest gap was on whether AI chatbots should be designed to be human-like, where men increased their support from 59.9% to 62.7%, whereas women decreased their support from 52.7% to 48.8%, widening a 7.2% gap to 13.9%. Brazil had the smallest differences between men and women on the proposals among the 4 countries that held deliberations.

German men and women mostly agreed on the proposals but had some significant differences on certain questions. Both before and after deliberations, German men and women differed most on 3 questions. Men agreed with the proposal that AI chatbots should be able to be offensive if it is meant to be an entertaining character, personality, assistant, or if the user is informed, about 20 percentage points more than women, with men around 60% in favor and women around 35%. These gaps decreased very little after deliberations.

Spanish men and women mostly agreed on the proposals. The largest gap in support before or after the deliberations was 19 percentage points. Wide gaps from before the deliberations mostly narrowed after deliberations, whereas small gaps mostly grew. One of the largest gaps was on "If the users are informed, users should be able to use AI Chatbots in any way they like, including romantic relationships," where men were about 19 percentage points more in favor than women. This gap narrowed to 12 points after deliberations as women moved towards men's position. The largest gap after deliberations was on "If users are informed, all AI Chatbots should be able to give responses in ways, or on topics, that some people might find offensive," with men 18.56 percentage points more in favor.

American men and women mostly agreed on the proposals. They were no more than 20 percentage points away from each oth-

er on any proposal question except one. On whether users should be able to use chatbots in any way they like, including romantic relationships, men remained about 20 percentage points more in favor of the proposal. A similar gap was observed on whether users should be allowed to interact with AI Chatbots in any way they desire within legal bounds, with men about 17 percentage points more in favor. On whether AI chatbots should be human-like, men began the deliberations about 10 percentage points more in favor, but men and women moved away from each other slightly, leading to final gaps of 13 to 18 percentage points.

Age

Brazil

Brazilian participants generally agreed with each other across age groups before the deliberations, and had few significant differences. A notable division was that participants 55 and older were about 13-20 percentage points less in favor of the proposals that AI chatbots should be able to be offensive if they are meant to be an entertaining character, personality, or assistant, compared to those 54 and younger.

Another of the most notable gaps in pre-deliberations was on whether AI chatbots should use a user's past conversations to offer the best user experience, even if the user is not informed. Participants 18-24 years old agreed with the proposal significantly more than those 25 and older. Additionally, 45-54 year olds seemed particularly wary of AI chatbots using user data compared to all other age groups.

A few notable changes occurred after deliberations. Participants 45-54 years old retracted support for the proposals asking if AI chatbots should be able to be offensive, joining those who were 54 and older and creating a significant gap between this new bloc (45 and older), and those 44 and younger, with those 45 and older about 7-15 percentage points less in favor of the proposals. Another notable change was that participants 55 and older retracted support for several proposals related to AI chatbots' use of user data, causing a significant gap to emerge between these participants and 18-54 year olds. Although participants 55 and older and 18-24 years old disagreed on several proposals related to chatbots' use of user data, they equally disagreed that AI chatbots should be able to use users' past conversations if the user is not informed. This was due to 18 to 24 year olds retracting support for this proposal dramatically post-deliberations, by about 30 percentage points.

Germany

German participants generally agreed with each other across age groups before the deliberations, but did have some significant differences on some proposals. Participants 18-24 years old agreed significantly more with the proposals suggesting that AI chatbots should be able to be offensive, compared to those 25 and older. Similarly, 18-34 year olds agreed significantly more with the proposal that users should be able to leverage AI chatbots to

enable their relationships with humans, compared to those 35 and older.

After deliberations, gaps between age groups were similar for most proposals. However, gaps between 18-24 year olds and those 25 and older increased quite significantly for proposals about whether AI chatbots should be able to be offensive, as well as for proposals about whether AI chatbots should be humanlike. For both of these sets of proposals,18-24 year olds and those 25 and older moved away from each other, with much of the gap widening coming from 18-24 year old's increased support for these proposals.

Spain

Spanish participants generally agreed with each other across age groups before and after deliberations, with few significant differences. The most notable gaps were a result of 25-54 year olds having significantly more support for proposals asking if AI chatbots should be able to be offensive, compared to both those 55 and older and 18-24 year olds on. Interestingly, 18-24 year olds and those 55 and older largely agreed in their lack of support for these proposals. Gaps between 18-24 year olds and 25-44 year olds decreased significantly post-deliberations for these proposals about AI chatbots' offensiveness, mostly as a result of 25-44 year olds retracting support for these proposals.

Although most post-deliberations changes in age group differences were generally moderate, Spanish participants 18-24 and those 25 and older moved away from each other on proposals about whether users should be able to use AI chatbots for romantic relationships or for any purpose within legal bounds, causing a significant gap to emerge between the two age categories post-deliberations. 18 to 24 year olds retracted support for both proposals whereas those 25 and older increased their support.

United States

American participants had the sharpest and most consistent divisions between age groups among all four countries, as those 45 and older had significant differences with those 44 and younger on almost every proposal before the deliberations. Some of the largest gaps between these two age categories were seen pre-deliberations as those 44 and younger agreed with the proposals suggesting that users should be able to use AI chatbots for romantic relationships, for any purpose within legal bounds, and to enable relationships with other humans about 20-40 percentage points more than those 45 and older. They also supported the proposals about AI chatbots being humanlike about 15-20 percentage points more.

Participants 55 and older broke away from 45-54 year olds on some proposals, like those asking whether AI chatbots should be able to give responses in ways, or on topics, that some people might find offensive. Those 55 and older disagreed with these proposals significantly more compared to all other participants.

Gaps between age groups generally decreased slightly or were a similar size after deliberations. Thus, American participants 45 and older still had some significant gaps in opinion with participants 44 and younger after deliberations, but to a lesser degree. The differences on whether AI chatbots should be humanlike, the use of AI chatbots for romantic relationships or any purpose within legal bounds, and whether users should be able to leverage AI chatbots to enable relationships with other humans were the most persistent after deliberations.

Education

Participants were asked what the highest education level they had attained was at the time of the questionnaires. Each country's questionnaire had country-specific education categories, so for this analysis, they have been standardized to the following four categories: primary education or less, secondary education, post-secondary non-tertiary or vocational education, and some tertiary education or above.

Germany was a special case in that we had no data about how many Germans completed any post-secondary education. The section on German participants will compare participants from three different secondary education tracks: Hauptschule, Realschule, and Abitur.

Brazil

Before deliberations, Brazilian participants generally agreed with each other on the proposals across education categories, with few gaps wider than 20 percentage points. Those who had completed secondary education or higher generally agreed with each other more than they did with those who had only completed primary education or less.

Pre-deliberations, those who had only completed primary education or less were slightly to significantly more in favor of proposals suggesting that AI chatbots should be able to be offensive. They also agreed significantly more with the proposal that if an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that, about 12-25 percentage points more compared to the three other education categories.

Generally, gaps between the education groups changed minimally post-deliberations, but did widen or narrow significantly for a few proposals. For other proposals, participants moved away from each other slightly, causing significant gaps to emerge.

On, "If an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that," those who had only completed primary education or less and those who had completed secondary education or above, moved towards each other after deliberations, narrowing the gap significantly.

Additionally, those who had only completed primary education or less were slightly or significantly more in favor of AI chatbots drawing perspectives from a user's

local organizations compared to those who had completed secondary education and above. They were also more in favor of chatbots using the user's past conversations to offer a better user experience, even if the user is not informed and chatbots prioritizing responses that do not rely on user data, compared to those who completed secondary education and above.

Germany

Germany was a special case in that we had no data about how many Germans completed any post-secondary education. For Germans, we collected data about how many completed one of three secondary education tracks: Realschule (equivalent to about a 9th grade education), Hauptschule (equivalent to about a 10th grade education), and Abitur (equivalent to about 12th grade education).

Before deliberations, German participants generally agreed with each other on the proposals across the three secondary education tracks, with no gaps wider than 20 percentage points.

Those who had completed Realschule were slightly more in favor of the proposal that chatbots should provide trade offs drawn from the perspectives of a user's local or national organizations, compared to those who had completed Hauptschule or Abitur. Those who completed Abitur were moderately more in favor of the proposal that chatbots should source information primarily from the user's national organizations compared to those who had completed Hauptschule or Realschule.

After deliberations, gaps changed very little. Those who had completed Hauptschule were slightly more in favor of chatbots being humanlike, chatbots drawing perspectives from the user's national or local organizations when providing tradeoffs on a topic, and users being able to use chatbots for whatever they like, compared to those who had completed Realschule and Abitur as the two groupings moved away from each other slightly.

Those who had completed Abitur were moderately more in favor of chatbots being able to be offensive, and moderately less in favor of chatbots using the user's past conversations to offer a better user experience. Overall, they generally agreed less with most proposals compared to those who had completed Hauptschule.

Spain

Spanish participants who had completed secondary education and above largely agreed with each other before and after deliberations. They had very few significant differences between each other on the proposals.

Compared to those who had completed secondary education and above, Spanish participants who had only completed primary education or less were significantly less in favor of AI chatbots being humanlike. They also disagreed that AI chatbots should limit their responses to their intended task, or that they should limit their responses to friendly companionship only. They disagreed particularly strongly with the proposal that chatbots should prioritize consistent and predictable responses over unpredictable and edgy ones, with about 30-40 percentage points less support compared to those who had completed secondary education and above.

After deliberations, those who had completed secondary education and above still largely agreed with each other. Those who had only completed primary education and those who had completed secondary education and above moved away from each other on whether AI chatbots should be humanlike, largely due to the increase in support from those who had completed secondary education or above.

Post-deliberations, those who had completed only primary education agreed significantly less with the proposals suggesting that AI chatbots should be able to offensive, by about 25 percentage points.

United States

Before deliberations, American participants who had completed secondary education and above generally agreed with each other.

The largest differences were seen between this group and those who had only completed primary education or less, who agreed less with almost every proposal. In particular, those who completed only primary education or less disagreed significantly with the proposals suggesting that AI chatbots should be humanlike and that chatbots should draw perspectives from a user's local or national organizations when providing tradeoffs on a topic, about 20 percentage points less than those who had completed secondary education or above.

Gaps between education categories looked similar after deliberations. Those who had completed primary education or less moved further away from the other participants on whether AI chatbots should be able to offensive, disagreeing with the proposal even more. However, they moved towards the other participants on whether chatbots should draw perspectives from a user's local or national organizations when providing tradeoffs on a topic, agreeing with this proposal more after deliberations.

Participants who had used ChatGPT or similar chatbots	Before deliberations	After deliberations	Change
Brazil	62.0%	74.3%	+12.3%***
Germany	61.7%	69.0%	+7.3%***
Spain	64.2%	73.6%	+9.4%***
United States	59.5%	71.3%	+11.8%***

Usage of AI chatbots

A majority of participants from each country had used ChatGPT or similar chatbots. The number of participants who had used ChatGPT or similar chatbots increased as a result of the deliberations.

Brazil

Brazilian participants who had and had not used AI chatbots generally agreed with each other on the proposals before and after deliberations, with few significant differences.

Before the deliberations, AI chatbot users had slightly more support for proposals suggesting that AI chatbots should be humanlike, that users should be able to use AI chatbots for whatever they like, including romantic relationships, that AI chatbots should be able to be offensive, and that AI chatbots should use a user's past conversations to offer the best user experience, compared to non-users. These gaps generally decreased slightly after deliberations, largely due to non-users' increased support.

The largest gap between AI chatbot users and non-users occurred post-deliberations on whether AI chatbots should be able to be offensive if users are informed. Before deliberations, AI chatbot users were slightly more in support of this proposal compared to non-users, but were significantly more in support after deliberations, largely due to non-users retracting support for the proposal.

Germany

German participants who had and had not used AI chatbots generally agreed with each other on the proposals before and after deliberations, with few significant differences. Both before and after deliberations, AI chatbot users were more in favor of AI chatbots using a user's past conversa-

tions to offer the best user experience, being humanlike, being able to be offensive, and users being able to use AI chatbots in any way the like, including romantic relationships, compared to non-users.

Gaps on these proposals generally narrowed slightly after deliberations, most notably on users being able to use chatbots in any way they like and on chatbots being able to be offensive, with both AI chatbot users and non-users moving towards each other and away from their original positions slightly.

Spain

Pre-deliberations, AI chatbot users were slightly more in favor of AI chatbots being able to be offensive, being able to use a user's past conversation to offer the best experience, and users being able to use chatbots in any way the like within legal bounds, and being able to use chatbots to enable their relationship with other humans. Gaps on chatbots being able to be offensive widened slightly post-deliberations as AI chatbot users and non-users moved away from each other slightly. On the other hand, gaps on chatbots using a user's past conversations narrowed slightly, largely due to non-users' increased support of these proposals.

The widest pre-deliberations gaps were on whether AI chatbots should be humanlike, with AI chatbot users about 15 percentage points more in favor of these proposals than non users. These gaps decreased very slightly post-deliberation.

United States

American AI chatbot users and non-users had some of the largest differences between each other of the four countries, with moderate to significant differences on various proposals. AI chatbot users were about 20 percentage points more in favor of AI chatbots being humanlike, and were 20-30 percentage points more in favor of users using chatbots in any way they like, including romantic relationships, and users being able to leverage chatbots to enable relationships with other humans. AI chatbot users were also slightly more in favor of chatbots being able to be offensive, chatbots sourcing from the user's national organizations, scientific literature, major press outlets, and globally recognized authoritative sources, and chatbots using the user's past conversations to offer the best user experience.

These gaps generally narrowed slightly post-deliberations as both AI chatbot users and non-users moved away from their original positions and towards each other.

Control vs. Participant Groups: Analysis of Attitudinal Representativeness

The purpose of the control groups is to provide points of comparison for the participant groups along three lines. First, we compare the key demographic characteristics of gender, age, and education to see whether these groups are similarly representative of the country populations from which they are drawn. Second, we compare the control and participant groups' ratings of the proposals before the deliberations to determine whether they hold similar baseline opinions. This comparison serves as a check on attitudinal representativeness. Third, we compare the changes in the control and treatment groups from before to after the deliberations. In other words, we compare how the treatment group's proposal ratings changed to how the control group's proposal ratings changed. Given similar baseline proposal ratings, this difference-in-difference measure shows us which changes were attributable to participation in the community forum.

Both control group and participant group members completed surveys before and after the deliberations. Control group members did not receive briefing materials or participate in any parts of the event. Participant group members had the opportunity to review briefing materials, participate in small group discussions, and listen to panels of subject matter experts.

Control vs. Participant Group Demographics Before Deliberations

Gender

As the tables below indicate, the composition of the control and participant groups were close to evenly split in terms of gender. There were no statistically significant differences between the two groups for any of the four countries.

Gender - Brazil	Control group	Participants	Difference between control group and participants
Male	(51.0%) 185	(50.5%) 170	-0.5%
Female	(49.0%) 177	(49.5%) 166	0.5%

Gender - Germany	Control group	Participants	Difference between control group and participants
Male	(49.1%) 115	(48.2%) 194	-0.9%
Female	(50.9%) 119	(51.8%) 208	0.9%

Gender - Spain	Control group	Participants	Difference between control group and participants
Male	(51.0%) 138	(51.3%) 212	0.3%
Female	(49.0%) 133	(48.7%) 202	-1.3%

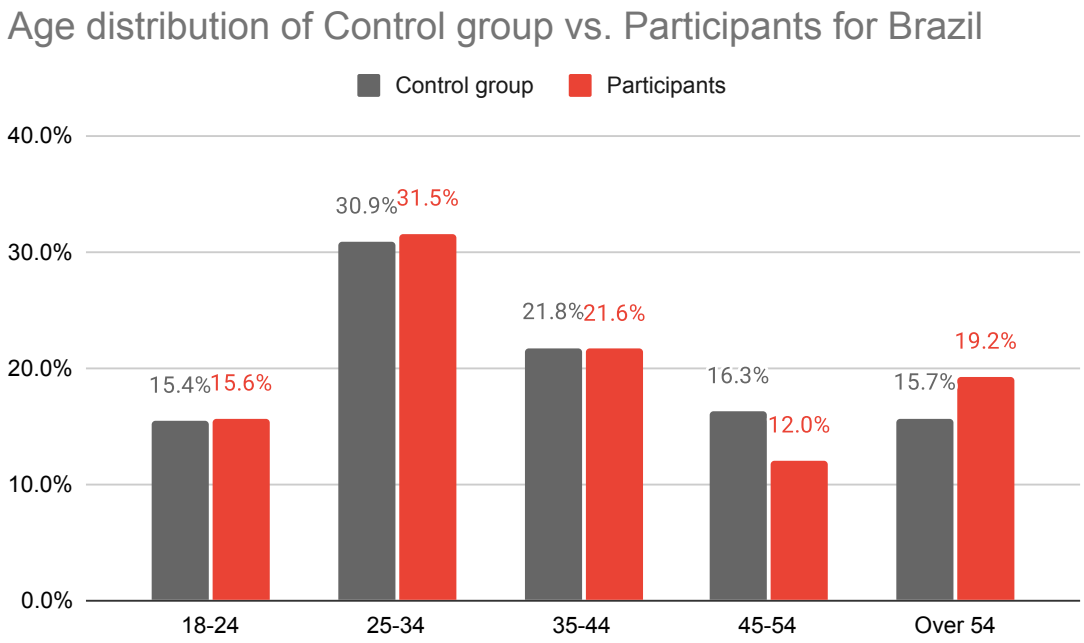
Gender - USA	Control group	Participants	Difference between control group and participants
Male	(47.9%) 117	(47.7%) 187	-0.2%
Female	(50.3%) 123	(50.4%) 198	0.1%
Non-binary	(1.7%) 4	(1.9%) 8	0.2%

Age

The control and participant groups had similar age characteristics for each country. Where there were statistically significant differences in the case of Spain, they were not substantively large.

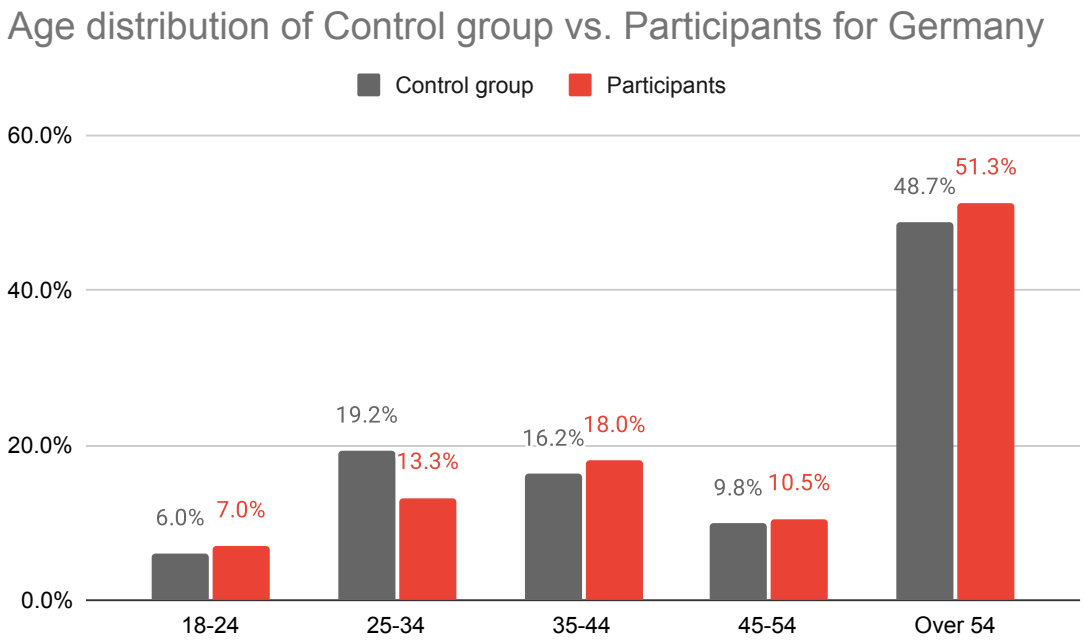
For Brazil, the participant group had about 3.5% more of its members in the over-54 age range, and about 4.2% less in the 45-54 age range. However, the differences were not statistically significant.

	Control group	Participants	Difference between control group and participants
18-24	15.4%	15.6%	0.2%
25-34	30.9%	31.5%	0.7%
35-44	21.8%	21.6%	-0.1%
45-54	16.3%	12.0%	-4.2%
Over 54	15.7%	19.2%	3.5%



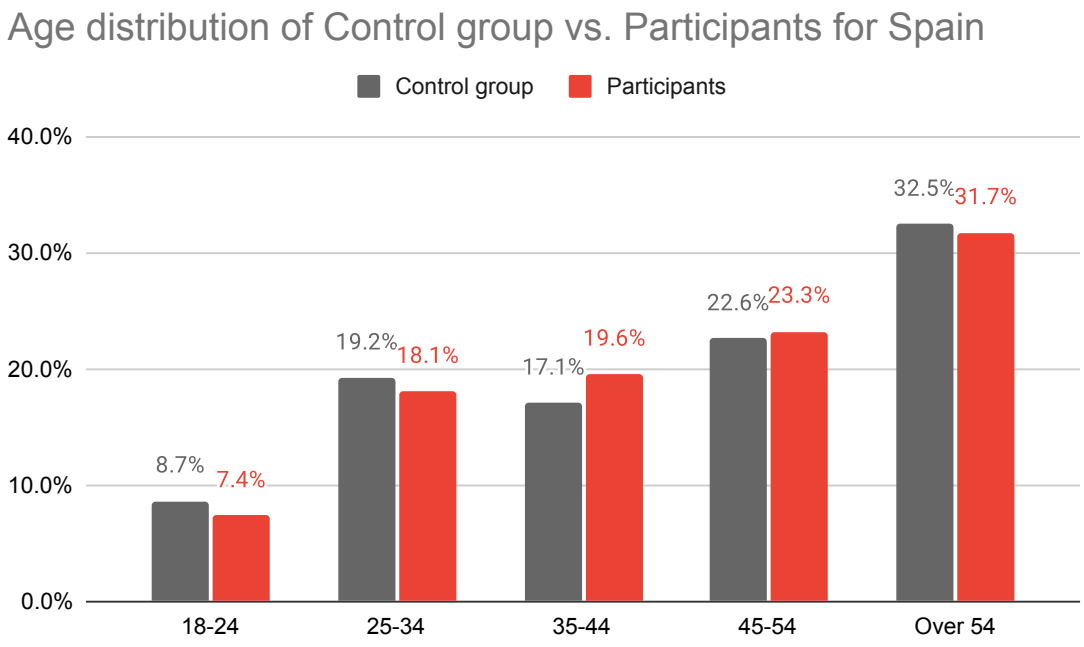
Germany's control group had a greater proportion in the 24-34 age range than the participants group, but the difference was not statistically significant.

	Control group	Participants	Difference between control group and participants
18-24	6.0%	7.0%	1.0%
25-34	19.2%	13.3%	-6.0%
35-44	16.2%	18.0%	1.8%
45-54	9.8%	10.5%	0.7%
Over 54	48.7%	51.3%	2.5%



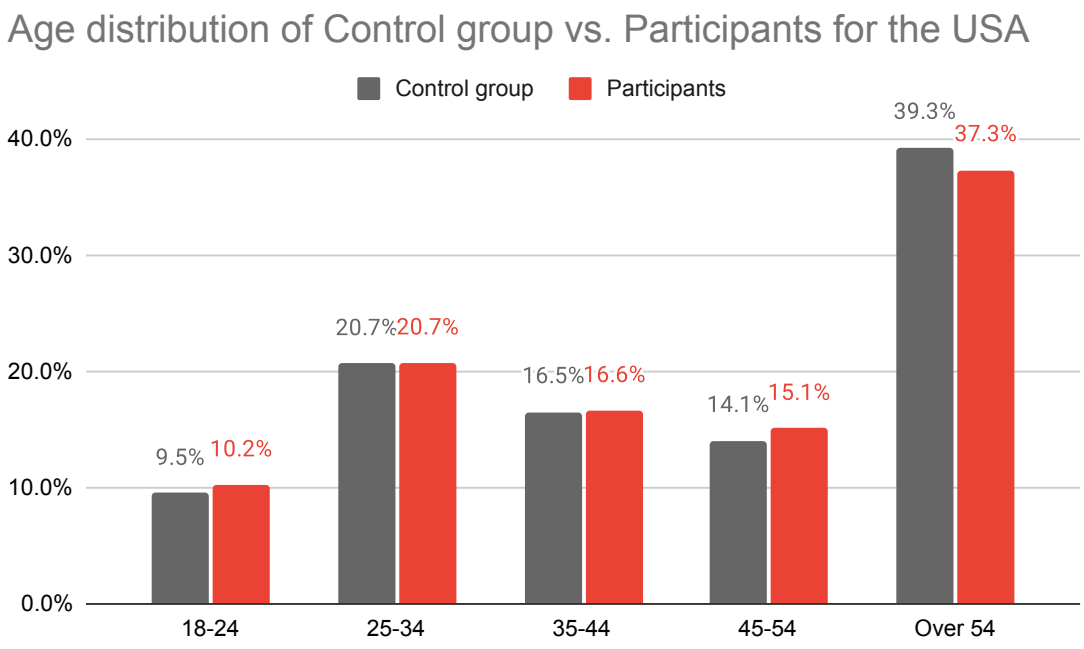
While Spain had some significant differences, they were not substantively large. The participants group was composed of close to 2.5% more 35-44-year-olds, while other age group differences were smaller.

	Control group	Participants	Difference between control group and participants
18-24	8.7%	7.4%	-1.3%
25-34	19.2%	18.1%	-1.2%
35-44	17.1%	19.6%	2.5%
45-54	22.6%	23.3%	0.7%
Over 54	32.5%	31.7%	-0.7%



As shown below, the control and participants groups were close to each other in terms of composition of age groups. For the United States, there were no statistically significant differences in age between the two groups.

	Control group	Participants	Difference between control group and participants
18-24	9.5%	10.2%	0.7%
25-34	20.7%	20.7%	0.1%
35-44	16.5%	16.6%	0.1%
45-54	14.1%	15.1%	1.0%
Over 54	39.3%	37.3%	-1.9%



Education

For Brazil, the participant group had about 8.5% more of its members having completed primary education or less and about 11.8% fewer having completed secondary education, roughly the equivalent of high school, as their highest level of education. However, it also had 3.5% more of its members completing some university education or higher.

Education - Brazil	Control group	Participants	Difference between control group and participants
Primary education or less	13.2%	21.7%	8.5%
Secondary education	54.0%	42.2%	-11.8%
Post-secondary non-tertiary or vocational education	17.8%	17.6%	-0.2%
Some tertiary education or higher	15.0%	18.5%	3.5%

For Germany, there was no data for education beyond the secondary level. The proportion of participant and control group members completing some form of secondary education was roughly the same.

Education - Germany	Control group	Participants	Difference between control group and participants
Primary education or less	N/A	N/A	N/A
Secondary education	98.0%	98.1%	0.1%
Post-secondary non-tertiary or vocational education	N/A	N/A	N/A
Some tertiary education or higher	N/A	N/A	N/A

For Spain, the participant group had 14.4% of its members completing some tertiary education or completing a bachelor's degree or higher. The differences were not statistically significant, however.

Education - Spain	Control group	Participants	Difference between control group and participants
Primary education or less	2.3%	1.5%	-0.8%
Secondary education	32.5%	27.1%	-5.4%
Post-secondary non-tertiary or vocational education	18.6%	22.2%	3.6%
Some tertiary education or higher	46.4%	5880.0%	14.4%

The United States control and participants groups had similar compositions in terms of highest level of education, without statistically significant differences.

Education - United States	Control group	Participants	Difference between control group and participants
Primary education or less	1.1%	4.1%	3.0%
Secondary education	36.8%	33.6%	-3.2%
Post-secondary non-tertiary or vocational education	N/A	N/A	N/A
Some tertiary education or higher	62.2%	62.3%	0.1%

Given these relatively similar distributions of gender, age, education for the control and participants groups for these four countries, it is unlikely that these demographic characteristics are responsible for any differences in changes between the two groups from before to after the community forum.

Attitudinal Representativeness

The control and participant groups were largely similar in their ratings of proposals before the deliberations. This comparison of policy attitudes served as a check on attitudinal representativeness.

As shown below, the control and participant groups for Brazil, Spain and the U.S. had few significant differences in their proposal ratings before the community forum. As noted earlier, the rating scale was 0 = strongly oppose, 5 = in the middle, 10 = strongly support.

For Brazil, there were significant differences between the control and participant groups before the deliberations for only three proposals:

- "AI chatbots should be designed to be as humanlike as possible, even if the user is not informed" (Control = 4.5, Participant = 3.94, Participant - Control = -0.57)
- "AI chatbots should provide the tradeoffs to a topic from the country in which they were created" (Control = 7.22, Participant = 7.70, Participant - Control = 0.48)
- "AI chatbots should use the user's past conversations to offer the user the best user experience, even if the user is not informed" (Control = 5.75, Participant = 5.08, Participant - Control = -0.67)

While these differences were significant, they were substantively minor. For the first proposal, the control group had an average rating of 4.5, or near the middle but slightly opposed, while the participant group had an average rating of 3.94, or slightly more opposed. Both groups supported the second proposal at 7.22 and 7.70 for the control and participant groups, respectively. For the third proposal listed above, the control group was slightly in favor, while the participant group was slightly above the middle. In each of these cases, the differences between the groups was less than 0.7.

For Spain, there were significant differences between the control and participant groups before the deliberations for only two proposals:

- "AI chatbots should be designed to be as humanlike as possible, even if the

user is not informed" (Control = 3.80, Participant = 3.2, Participant - Control = -0.6).

- "Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted" (Control = 3.33, Participant = 2.7, Participant - Control = -0.64).

As in the case of Brazil, the control and participant groups for Spain had substantively similar ratings. In this case, they were both opposed to the proposals, and their ratings were within 0.7 of each other on the 0 to 10 scale.

For the U.S., the only significant difference between the control and participant groups was for the proposal, "A user should be informed by the AI chatbot that they are interacting with a bot." The control group's average rating was 2.56 and the participant group's average rating was 2.18 on a 0 to 10 scale, a difference of .39.

In the case of Germany, there were 13 out of 36 proposals for which the control and participant groups differed significantly before the deliberations. However, the substantive differences were small on the 0 to 10 scale. In cases where the control group was opposed to a proposal, the participant group was also opposed to the proposal. The same was true for cases of support or being near the middle. The largest difference was for the proposal, "If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships," with a difference of 1.1 between the control and participant groups. While the control group neared the middle of the scale (5), both groups were on the opposed side (less than 5).

Statistically Significant Differences in Control vs. Participant Group Ratings of Proposals Before Deliberations, DE only	Control	Participant	Participant - Control
Proposal	**n = 236**	**n = 399**	
AI chatbots should be designed to be as humanlike as possible, even if the user is not informed	3.53	2.76	-0.77**
If an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that	6.16	5.66	-0.51*
AI chatbots should be trained to limit conversations to friendly companionship only, not romantic relationships	7.59	8.23	0.64***
If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships	4.83	3.77	-1.1***
Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted	3.88	3.16	-0.72**
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, regardless of their human rights records or treatment of marginalized groups	4.91	4.21	-0.71**

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, regardless of their human rights records or treatment of marginalized groups	4.55	3.68	-0.87***
AI chatbots should provide the tradeoffs to a topic from the country in which they were created	5.49	5.14	-0.35**
When citing a source, AI chatbots should only cite peer-reviewed scientific information, or discussions in major press outlets	7.60	7.95	0.36**
AI chatbots should source their information primarily from globally recognized authoritative sources (e.g. WHO)	7.14	7.71	0.57**
AI chatbots should source their information primarily from sources from the user's national organizations	6.02	5.32	-0.7**
AI chatbots should use the user's past conversations to offer the best user experience if the user is informed	6.85	7.35	0.50**
AI chatbots should use additional data sources, such as the user's online activity, to help personalize their interactions	4.98	4.04	-0.94**

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

The overall similarities in proposal ratings between the control and participant groups serve as a check on attitudinal representativeness. The control group can serve as a baseline for examining the changes in the participant group's policy attitudes. The full list of control vs. participant group proposal ratings before the community forum is shown below.

Control vs. Participant Group Ratings of Proposals Before Deliberations	Control				Participant				Participant - Control			
	BR	DE	ES	US	BR	DE	ES	US	BR	DE	ES	US
Proposal	n = 362	n = 236	n = 272	n = 244	n = 336	n = 399	n = 413	n = 393				
AI chatbots should be designed to be as humanlike as possible, even if the user is not informed	4.50	3.53	3.80	3.70	3.94	2.76	3.20	3.30	-0.57*	-0.77**	-0.60**	-0.40
If the user is informed, AI chatbots should be designed to be as humanlike as possible	6.00	5.88	5.35	5.35	6.14	5.91	5.38	5.56	0.14	0.03	0.03	0.21
AI chatbots should be able to use the user's emotional cues to help direct the conversation and offer the greatest potential support, only if the user is informed	6.57	6.82	5.93	6.27	6.69	7.02	5.86	6.01	0.12	0.20	-0.07	-0.26
AI chatbots should be able to use conversational tactics to engage the user to express their deepest thoughts and feelings to offer the greatest potential support, only if the user is informed	6.61	5.98	5.82	5.96	6.88	5.89	5.93	6.04	0.27	-0.09	0.11	0.09
A user should be informed by the AI chatbot that they are interacting with a bot	2.51	2.43	2.48	2.56	2.46	2.42	2.42	2.18	-0.04	-0.01	-0.07	-0.39**
If an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that	5.80	6.16	5.65	5.93	5.37	5.66	5.50	5.92	-0.43	-0.51*	-0.15	-0.01

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

If the primary purpose of the AI chatbot is task-based, it should not be able to respond to questions outside that	6.16	6.18	6.09	6.42	5.73	6.08	5.75	6.55	-0.43	-0.10	-0.33	0.13
AI chatbots should be trained to limit conversations to friendly companionship only, not romantic relationships	7.47	7.59	7.06	7.49	7.71	8.23	7.35	7.50	0.24	0.64***	0.29	0.00
If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships	4.92	4.83	4.38	4.67	4.63	3.77	4.11	4.40	-0.29	-1.1***	-0.26	-0.27
Regardless of whether they are informed, users should be allowed to interact with AI chatbots in any way they desire within legal bounds	6.92	6.11	5.68	5.68	6.80	5.92	5.56	5.35	-0.12	-0.19	-0.12	-0.33
Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted	4.84	3.88	3.33	3.40	4.44	3.16	2.70	3.69	-0.40	-0.72**	-0.64**	0.29
Users should be able to leverage AI chatbots to enable their relationships with other humans, only if the other person knows AI assistance is involved	6.09	6.36	5.99	6.48	6.26	6.63	5.84	5.91	0.17	0.26	-0.15	-0.57

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Where do you place yourself on the scale between "Users should be able to use AI chatbots to enable relationships with other humans, even if the other person or people do not know they are AI-assisted" and "Users should be able to use AI chatbots to enable relationships with other humans, only if the other person or people know they are AI-assisted."	6.50	7.51	6.95	7.63	6.72	7.69	6.89	7.12	0.22	0.18	-0.05	-0.51
AI chatbots intended primarily for information should prioritize consistent and predictable responses over unpredictable and edgy ones	7.24	7.05	7.07	7.32	7.39	6.91	7.01	7.38	0.15	-0.14	-0.05	0.07
AI chatbots that are primarily intended for amusement should prioritize unpredictable and edgy responses over predictable and inoffensive ones	5.92	6.07	5.70	5.73	6.12	5.64	5.55	5.62	0.20	-0.43	-0.15	-0.12
If an AI chatbot is designed to take on a character or personality that provides entertaining responses or tells jokes, it should be able to give responses in ways, or on topics, that some people might find offensive	5.91	4.95	5.21	5.64	5.86	4.88	5.14	5.39	-0.05	-0.08	-0.08	-0.25
If an AI chatbot is designed to be an assistant, it should be able to give responses in ways, or on topics, that some people might find offensive	5.84	4.69	4.88	5.34	5.50	4.30	4.85	4.92	-0.35	-0.39	-0.03	-0.43

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

If users are informed, all AI chatbots should be able to give responses in ways, or on topics, that some people might find offensive	6.01	5.00	5.32	5.87	6.13	4.86	5.41	5.31	0.12	-0.14	0.09	-0.56
Users should be able to control the level of AI chatbot predictability or unpredictability	7.26	6.79	6.82	7.25	7.28	6.97	6.91	7.03	0.02	0.18	0.09	-0.22
AI chatbots should be predictable and inoffensive by default	7.39	6.88	7.34	6.89	7.35	6.83	7.14	6.92	-0.03	-0.05	-0.19	0.04
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from international organizations, regardless of whether this conflicts with local or country-level perspectives	7.31	6.76	6.16	6.46	7.49	7.06	5.97	6.30	0.17	0.30	-0.19	-0.16
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, regardless of their human rights records or treatment of marginalized groups	6.44	4.91	5.04	4.98	6.17	4.21	4.61	5.10	-0.27	-0.71**	-0.43	0.11
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	6.65	6.25	6.19	6.73	6.23	6.60	6.38	6.28	-0.42	0.34	0.19	-0.45

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, regardless of their human rights records or treatment of marginalized groups	6.15	4.55	4.75	4.65	5.73	3.68	4.46	4.86	-0.43	-0.87***	-0.29	0.21
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	6.57	6.33	6.13	6.47	6.43	6.39	6.33	6.18	-0.15	0.06	0.20	-0.29
AI chatbots should provide the tradeoffs to a topic from the country in which they were created	7.22	5.49	5.60	6.22	7.70	5.14	5.35	6.01	0.48*	-0.35**	-0.24	-0.21
To ensure information is culturally relevant, AI chatbots should cite users' local media outlets and their national organizations' guidance, even if it contradicts the AI chatbot maker's values	7.04	6.64	7.05	6.87	6.96	6.78	6.98	6.73	-0.08	0.14	-0.07	-0.14
When citing a source, AI chatbots should only cite peer-reviewed scientific information, or discussions in major press outlets	6.82	7.60	7.08	6.96	6.60	7.95	6.93	6.41	-0.21	0.36**	-0.15	-0.55
AI chatbots should source their information primarily from globally recognized authoritative sources (e.g. WHO)	7.56	7.14	7.56	6.75	7.72	7.71	7.71	6.36	0.16	0.57**	0.15	-0.39
AI chatbots should source their information primarily from sources from the user's national organizations	7.01	6.02	6.29	6.21	6.99	5.32	6.00	5.92	-0.02	-0.7**	-0.29	-0.29

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed	5.75	4.43	5.07	5.14	5.08	4.14	5.04	5.11	-0.67**	-0.29	-0.03	-0.02
AI chatbots should use the user's past conversations to offer the best user experience if the user is informed	7.57	6.85	7.22	7.18	7.88	7.35	7.17	6.87	0.31	0.50**	-0.05	-0.31
AI chatbots should use past conversations to offer the best user experience if the user is informed and they are able to access and delete their chat history	7.75	7.49	7.33	7.49	7.95	7.60	7.32	6.98	0.20	0.11	-0.01	-0.51
AI chatbots should use additional data sources, such as the user's online activity, to help personalize their interactions	7.09	4.98	6.07	5.54	7.01	4.04	5.74	5.18	-0.08	-0.94**	-0.33	-0.36
AI chatbots should use additional data sources, such as the user's online activity, with the permission of the user, to help personalize their interactions	7.80	5.84	6.77	6.76	8.09	5.81	6.99	6.58	0.29	-0.03	0.22	-0.18
AI chatbots should prioritize standardized responses that do not rely on user data or their online activity	6.32	6.23	6.24	6.50	6.16	5.82	6.12	6.65	-0.16	-0.41	-0.12	0.15

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Comparison of Changes in the Participant and Control Groups: Difference in Differences

The table below shows a comparison of changes in the treatment group and control group from before to after the period of deliberation. (Asterisks indicate statistically significant differences, * for .05 level, ** for .01 level, *** for .001 level). In other words, it illustrates (participant group rating after deliberation - participant group rating before deliberation) - (control group rating after deliberation - control group rating before deliberation).

The statistically significant differences in differences in the chart suggest changes in opinions due to participation in the community forum. For example, for the third proposal, "AI chatbots should be able to use the user's emotional cues to help direct the conversation and offer the greatest potential support, only if the user is informed," there was a 0.62 difference between the change in participant group rating from before to after the community forum and the change in control group rating from before to after the community forum for Germany. For Spain, the participant group's rating changed by 0.61 more than the control group's rating changed. The differences in the Germany and Spain groups were statistically significant but not in the USA or Brazil groups. For participants in the Germany and Spain discussions, the average 0.62 and 0.61 shifts in opinions beyond the control group shifts in opinions were attributable to participation in the community forum (They exceeded a comparable baseline). A similar line of reasoning applies to other statistically significant differences in differences throughout the chart.

Brazil had seven statistically significant difference-in-difference measures. The largest (-1.148) was for the proposal, "AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed."

Germany had eight statistically significant difference-in-difference measures. The largest (1.009) was for the proposal, "AI chatbots should source their information primarily from the user's national organizations."

Spain had eleven statistically significant difference-in-difference measures. The largest (0.987) was for the proposal, "AI chatbots should use past conversations to offer the best user experience if the user is informed and they are able to access and delete their chat history."

The U.S. had four statistically significant difference-in-difference measures. The largest (-0.595) was for the same proposal as Brazil. "AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed."

These differences in differences were not the only outcomes of the deliberations. In other instances, for example, participants considered opposing arguments, weighed tradeoffs, and became more confident in their original opinions having reasoned through them – another result of interacting with other participants and learning about the issues during the community forum.

Comparison of Changes in the Participant and Control Groups: Difference in Differences (by country)

Proposal	BR	DE	ES	US
AI chatbots should be designed to be as humanlike as possible, even if the user is not informed	-0.522*	-0.19	-0.23	-0.42
If the user is informed, AI chatbots should be designed to be as humanlike as possible	0.24	0.30	0.696**	-0.30
AI chatbots should be able to use the user's emotional cues to help direct the conversation and offer the greatest potential support, only if the user is informed	-0.01	.618*	0.614**	0.15
AI chatbots should be able to use conversational tactics to engage the user to express their deepest thoughts and feelings to offer the greatest potential support, only if the user is informed	0.17	0.24	0.25	-0.12
A user should be informed by the AI chatbot that they are interacting with a bot	0.00	-0.06	0.193*	0.204*
If an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that	0.13	0.00	0.464*	0.28
If the primary purpose of the AI chatbot is task-based, it should not be able to respond to questions outside that	0.00	-0.02	0.606**	0.13
AI chatbots should be trained to limit conversations to friendly companionship only, not romantic relationships	0.32	-0.433*	0.24	0.25
If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships	-0.35	0.601*	-0.08	0.11
Regardless of whether they are informed, users should be allowed to interact with AI chatbots in any way they desire within legal bounds	-0.04	-0.23	0.07	-0.25

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted	-0.729**	0.451*	0.27	-0.27
Users should be able to leverage AI chatbots to enable their relationships with other humans, only if the other person knows AI assistance is involved	0.02	-0.09	0.18	0.06
Where do you place yourself on the scale between "Users should be able to use AI chatbots to enable relationships with other humans, even if the other person or people do not know they are AI-assisted" and "Users should be able to use AI chatbots to enable relationships with other humans, only if the other person or people know they are AI-assisted."	0.04	0.02	0.14	-0.09
AI chatbots intended primarily for information should prioritize consistent and predictable responses over unpredictable and edgy ones	0.12	0.709**	0.11	0.15
AI chatbots that are primarily intended for amusement should prioritize unpredictable and edgy responses over predictable and inoffensive ones	-0.42	0.33	0.493*	-0.28
If an AI chatbot is designed to take on a character or personality that provides entertaining responses or tells jokes, it should be able to give responses in ways, or on topics, that some people might find offensive	-0.41	0.11	0.37	0.28
If an AI chatbot is designed to be an assistant, it should be able to give responses in ways, or on topics, that some people might find offensive	-0.43	0.10	-0.47	0.44
If users are informed, all AI chatbots should be able to give responses in ways, or on topics, that some people might find offensive	-0.654*	-0.15	-0.23	0.33
Users should be able to control the level of AI chatbot predictability or unpredictability	0.14	-0.04	0.30	0.41
AI chatbots should be predictable and inoffensive by default	0.24	-0.42	-0.01	0.09

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should provide the tradeoffs on a topic, drawing perspectives from international organizations, regardless of whether this conflicts with local or country-level perspectives	-0.548*	-0.09	0.491*	-0.25
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, regardless of their human rights records or treatment of marginalized groups	-0.53	0.33	0.10	-0.11
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	0.07	0.45	0.464*	0.44
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, regardless of their human rights records or treatment of marginalized groups	-0.43	0.44	-0.21	-0.08
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	-0.20	0.638*	0.38	0.27
AI chatbots should provide the tradeoffs to a topic from the country in which they were created	-0.10	0.505*	0.43	0.42
To ensure information is culturally relevant, AI chatbots should cite users' local media outlets and their national organizations' guidance, even if it contradicts the AI chatbot maker's values	-0.14	0.16	0.14	-0.03
When citing a source, AI chatbots should only cite peer-reviewed scientific information, or discussions in major press outlets	0.29	0.28	0.16	0.548*
AI chatbots should source their information primarily from globally recognized authoritative sources (e.g. WHO)	0.34	0.10	0.20	0.21
AI chatbots should source their information primarily from sources from the user's national organizations	0.30	1.009***	0.13	0.39
AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed	-1.148***	-0.46	-0.491*	-0.595**

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should use the user's past conversations to offer the best user experience if the user is informed	0.05	-0.39	0.895***	0.00
AI chatbots should use past conversations to offer the best user experience if the user is informed and they are able to access and delete their chat history	0.567*	0.33	0.987***	0.513*
AI chatbots should use additional data sources, such as the user's online activity, to help personalize their interactions	-0.16	-0.11	0.24	-0.04
AI chatbots should use additional data sources, such as the user's online activity, with the permission of the user, to help personalize their interactions	0.22	0.12	0.27	0.23
AI chatbots should prioritize standardized responses that do not rely on user data or their online activity	0.509*	0.39	0.18	0.39

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Conclusion, Limitations, and Implications

Conclusion

The results of this Community Forum revealed that people are interested in learning more about and experiencing AI chatbots for themselves. The fact that a significant number of participants tried using chatbots after being invited to participate in this event shows the amount of curiosity and interest in AI chatbots. On the policy proposals discussed, participants agreed that context matters in choosing local or international perspectives and maintained concerns over AI bias, misinformation, and potential human rights violations. In addition, transparency is top of mind; as well as having the ability for users to control the data if it is used to personalize their chatbot experiences (over predictability of responses), ensuring accuracy, and having chatbots cite sources. While participants were excited about AI chatbots, they did feel skeptical about AI replacing human interaction and had mixed opinions regarding AI in mental health support.

When asked about the messages they wanted to communicate to technology companies and creators of AI platforms, participants wanted to convey that user privacy and data security paramount and that companies should be transparent about usage of data.

Limitations

This report presents primarily the quantitative results of the Community Forum. There still remains significant post-event analysis on the reasoning and arguments participants wrestled with to reach these opinions at the end of their deliberations. In subsequent reports, DDL will share the arguments participants used and considered as they deliberated on these policy issues. By delving deeper into what participants discussed and what arguments resonated most with participants, subsequent reports can give a more complete picture of how participants arrived at their conclusions.

Implications

The growing popularity of AI chatbots and related technologies are undeniable. As people from around the world interact with AI chatbots and become aware of the use cases for AI chatbots, this moment now is a critical time for users and creators of AI platforms to examine how to maxmine the benefits and minimize the risks of AI chatbots. AI chatbots have the potential to solve complex problems and improve the quality of life for millions of people. By studying the principles of AI engagement, we can learn how to design AI systems that are effective and meet the needs of users. At the same time, participants in this Community Forum also highlighted that AI also poses risks such as job displacement, privacy violations, and the spread of misinformation. This Community Forum is helping us understand what users and non-users think about the principles of AI engagement and helps all of us identify these risks and develop strategies to mitigate them. As AI chatbots and related technologies are rapidly evolving, it is important to start thinking about the ethical and societal implications of this technology now. It is only through engaging with the users and non-users of these technologies that we, as a society, can prepare for the future and ensure that AI is used for good.

Appendix

Changes in Control Group Proposal Ratings Before and After Deliberations

Change in Control Group Proposal Ratings from T1 to T2 by Country												
	US T1	DE T1	ES T1	BR T1	US T2	DE T2	ES T2	BR T2	US DIFF	DE DIFF	ES DIFF	BR DIFF
Proposal	n = 244	n = 236	n = 272	n = 362	n = 244	n = 236	n = 272	n = 362	n = 244	n = 236	n = 272	n = 362
AI chatbots should be designed to be as humanlike as possible, even if the user is not informed	3.70	3.53	3.80	4.50	3.58	3.02	3.61	4.41	-0.12	-0.502**	-0.19	-0.09
If the user is informed, AI chatbots should be designed to be as humanlike as possible	5.35	5.88	5.35	6.00	5.74	5.63	5.55	6.16	0.397*	-0.25	0.20	0.16
AI chatbots should be able to use the user's emotional cues to help direct the conversation and offer the greatest potential support, only if the user is informed	6.27	6.82	5.93	6.57	6.24	6.60	6.02	7.06	-0.03	-0.22	0.09	0.49**
AI chatbots should be able to use conversational tactics to engage the user to express their deepest thoughts and feelings to offer the greatest potential support, only if the user is informed	5.96	5.98	5.82	6.61	5.93	5.69	6.11	6.87	-0.02	-0.29	0.28	0.26
A user should be informed by the AI chatbot that they are interacting with a bot	2.56	2.43	2.48	2.51	2.52	2.53	2.38	2.42	-0.05	0.10	-0.10	-0.08
If an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that	5.93	6.16	5.65	5.80	6.23	6.12	6.09	6.02	0.30	-0.04	0.441**	0.21

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

If the primary purpose of the AI chatbot is task-based, it should not be able to respond to questions outside that	6.42	6.18	6.09	6.16	6.64	6.45	6.29	6.28	0.22	0.27	0.20	0.13
AI chatbots should be trained to limit conversations to friendly companionship only, not romantic relationships	7.49	7.59	7.06	7.47	7.30	7.73	7.09	7.42	-0.20	0.14	0.03	-0.05
If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships	4.67	4.83	4.38	4.92	4.70	4.76	4.76	5.35	0.04	-0.07	0.383*	0.424*
Regardless of whether they are informed, users should be allowed to interact with AI chatbots in any way they desire within legal bounds	5.68	6.11	5.68	6.92	5.59	5.78	5.70	6.79	-0.09	-0.33	0.02	-0.13
Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted	3.40	3.88	3.33	4.84	3.47	3.31	3.28	5.12	0.07	-0.565***	-0.05	0.28
Users should be able to leverage AI chatbots to enable their relationships with other humans, only if the other person knows AI assistance is involved	6.48	6.36	5.99	6.09	6.43	6.25	5.98	6.41	-0.05	-0.12	-0.01	0.33
Where do you place yourself on the scale between "Users should be able to use AI chatbots to enable relationships with other humans, even if the other person or people do not know they are AI-assisted" and "Users should be able to use AI chatbots to	7.63	7.51	6.95	6.50	7.43	7.24	7.23	6.78	-0.20	-0.27	0.28	0.27

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

enable relationships with other humans, only if the other person or people know they are AI-assisted."												
AI chatbots intended primarily for information should prioritize consistent and predictable responses over unpredictable and edgy ones	7.32	7.05	7.07	7.24	7.45	6.75	7.16	7.36	0.14	-0.30	0.09	0.11
AI chatbots that are primarily intended for amusement should prioritize unpredictable and edgy responses over predictable and inoffensive ones	5.73	6.07	5.70	5.92	6.15	5.98	5.77	6.07	0.413*	-0.09	0.07	0.15
If an AI chatbot is designed to take on a character or personality that provides entertaining responses or tells jokes, it should be able to give responses in ways, or on topics, that some people might find offensive	5.64	4.95	5.21	5.91	5.71	4.84	5.12	5.93	0.06	-0.12	-0.09	0.02
If an AI chatbot is designed to be an assistant, it should be able to give responses in ways, or on topics, that some people might find offensive	5.34	4.69	4.88	5.84	4.77	4.60	5.10	5.79	-0.58**	-0.09	0.22	-0.05
If users are informed, all AI chatbots should be able to give responses in ways, or on topics, that some people might find offensive	5.87	5.00	5.32	6.01	5.73	5.01	5.49	6.15	-0.14	0.01	0.17	0.14
Users should be able to control the level of AI chatbot predictability or unpredictability	7.25	6.79	6.82	7.26	7.24	7.18	6.96	7.42	0.00	0.387*	0.14	0.16

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should be predictable and inoffensive by default	6.89	6.88	7.34	7.39	6.95	7.09	7.21	7.45	0.06	0.21	-0.12	0.07
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from international organizations, regardless of whether this conflicts with local or country-level perspectives	6.46	6.76	6.16	7.31	6.58	6.49	5.98	7.14	0.12	-0.27	-0.18	-0.17
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, regardless of their human rights records or treatment of marginalized groups	4.98	4.91	5.04	6.44	5.26	4.68	5.03	6.21	0.28	-0.23	-0.01	-0.22
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	6.73	6.25	6.19	6.65	6.66	6.23	6.06	6.76	-0.08	-0.02	-0.14	0.11
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, regardless of their human rights records or treatment of marginalized groups	4.65	4.55	4.75	6.15	5.04	4.33	4.78	6.30	0.39	-0.23	0.04	0.14
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	6.47	6.33	6.13	6.57	6.52	6.08	6.01	6.69	0.05	-0.25	-0.12	0.12

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should provide the tradeoffs to a topic from the country in which they were created	6.22	5.49	5.60	7.22	6.22	5.49	5.59	7.34	0.01	0.00	-0.01	0.12
To ensure information is culturally relevant, AI chatbots should cite users' local media outlets and their national organizations' guidance, even if it contradicts the AI chatbot maker's values	6.87	6.64	7.05	7.04	7.11	6.60	7.06	7.03	0.24	-0.04	0.01	-0.01
When citing a source, AI chatbots should only cite peer-reviewed scientific information, or discussions in major press outlets	6.96	7.60	7.08	6.82	6.76	7.18	7.16	7.00	-0.20	-0.41	0.09	0.19
AI chatbots should source their information primarily from globally recognized authoritative sources (e.g. WHO)	6.75	7.14	7.56	7.56	6.82	6.97	7.36	7.43	0.07	-0.17	-0.20	-0.13
AI chatbots should source their information primarily from sources from the user's national organizations	6.21	6.02	6.29	7.01	6.25	5.49	6.43	6.82	0.04	-0.53	0.14	-0.19
AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed	5.14	4.43	5.07	5.75	4.77	3.98	4.85	5.85	-0.368*	-0.45	-0.21	0.09
AI chatbots should use the user's past conversations to offer the best user experience if the user is informed	7.18	6.85	7.22	7.57	7.27	6.99	6.82	7.40	0.09	0.14	-0.4**	-0.17
AI chatbots should use past conversations to offer the best user experience if the user is informed and they are able to access and delete their chat history	7.49	7.49	7.33	7.75	7.50	7.40	7.17	7.50	0.01	-0.09	-0.16	-0.25

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should use additional data sources, such as the user's online activity, to help personalize their interactions	5.54	4.98	6.07	7.09	5.48	4.97	5.97	6.96	-0.06	-0.01	-0.10	-0.12
AI chatbots should use additional data sources, such as the user's online activity, with the permission of the user, to help personalize their interactions	6.76	5.84	6.77	7.80	6.73	5.92	6.76	7.43	-0.03	0.08	-0.01	-0.366**
AI chatbots should prioritize standardized responses that do not rely on user data or their online activity	6.50	6.23	6.24	6.32	6.24	6.19	6.20	6.53	-0.26	-0.04	-0.04	0.21

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

Changes in Treatment Group Proposal Ratings Before and After Deliberations

Changes in Treatment Group Ratings from T1 to T2 by Country

Proposal	US T1 n = 393	DE T1 n = 399	ES T1 n = 413	BR T1 n = 336	US T2 n = 393	DE T2 n = 399	ES T2 n = 413	BR T2 n = 336	USA DIFF n = 393	DE DIFF n = 399	ES DIFF n = 413	BR DIFF n = 336
AI chatbots should be designed to be as humanlike as possible, even if the user is not informed	3.30	2.76	3.20	3.94	2.76	2.07	2.78	3.32	-0.538***	-0.691***	-0.418**	-0.612**
If the user is informed, AI chatbots should be designed to be as humanlike as possible	5.56	5.91	5.38	6.14	5.66	5.96	6.27	6.54	0.10	0.05	0.896***	0.401*
AI chatbots should be able to use the user's emotional cues to help direct the conversation and offer the greatest potential support, only if the user is informed	6.01	7.02	5.86	6.69	6.13	7.42	6.57	7.16	0.12	0.398*	0.704***	0.476*
AI chatbots should be able to use conversational tactics to engage the user to express their deepest thoughts and feelings to offer the greatest potential support, only if the user is informed	6.04	5.89	5.93	6.88	5.90	5.83	6.46	7.32	-0.14	-0.05	0.532***	0.434*
A user should be informed by the AI chatbot that they are interacting with a bot	2.18	2.42	2.42	2.46	2.33	2.46	2.51	2.38	0.154**	0.04	0.09	-0.08
If an AI chatbot has a specific entertaining personality, it should not be able to respond to questions outside of that	5.92	5.66	5.50	5.37	6.51	5.62	6.40	5.71	0.582**	-0.04	0.904***	0.34
If the primary purpose of the AI chatbot is task-based, it should not be able to respond to questions outside that	6.55	6.08	5.75	5.73	6.90	6.33	6.56	5.85	0.351*	0.25	0.806***	0.13
AI chatbots should be trained to limit conversations to friendly companionship only, not romantic relationships	7.50	8.23	7.35	7.71	7.55	7.94	7.62	7.98	0.05	-0.293*	0.27	0.27

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

If the users are informed, users should be able to use AI chatbots in any way they like, including romantic relationships	4.40	3.77	4.11	4.63	4.55	4.31	4.41	4.70	0.15	0.531**	0.30	0.07
Regardless of whether they are informed, users should be allowed to interact with AI chatbots in any way they desire within legal bounds	5.35	5.92	5.56	6.80	5.02	5.36	5.66	6.63	-0.335*	-0.561***	0.09	-0.17
Users should be able to leverage AI chatbots to enable their relationships with other humans, without the other person knowing they are AI-assisted	3.69	3.16	2.70	4.44	3.49	3.04	2.92	3.99	-0.20	-0.12	0.22	-0.449*
Users should be able to leverage AI chatbots to enable their relationships with other humans, only if the other person knows AI assistance is involved	5.91	6.63	5.84	6.26	5.92	6.42	6.01	6.60	0.01	-0.21	0.17	0.35
Where do you place yourself on the scale between "Users should be able to use AI chatbots to enable relationships with other humans, even if the other person or people do not know they are AI-assisted" and "Users should be able to use AI chatbots to enable relationships with other humans, only if the other person or people know they are AI-assisted."	7.12	7.69	6.89	6.72	6.83	7.44	7.31	7.03	-0.29	-0.25	0.415*	0.32
AI chatbots intended primarily for information should prioritize consistent and predictable responses over unpredictable and edgy ones	7.38	6.91	7.01	7.39	7.67	7.32	7.21	7.63	0.289*	0.409**	0.20	0.23
AI chatbots that are primarily intended for amusement should prioritize unpredictable and edgy responses over predictable and inoffensive ones	5.62	5.64	5.55	6.12	5.75	5.89	6.11	5.85	0.13	0.24	0.563**	-0.27
If an AI chatbot is designed to take on a character or personality that provides entertaining responses or tells jokes, it should be able to give responses in ways, or on topics, that some people might find offensive	5.39	4.88	5.14	5.86	5.73	4.86	5.42	5.47	0.339*	-0.01	0.28	-0.39

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

If an AI chatbot is designed to be an assistant, it should be able to give responses in ways, or on topics, that some people might find offensive	4.92	4.30	4.85	5.50	4.78	4.31	4.60	5.02	-0.14	0.01	-0.25	-0.476**
If users are informed, all AI chatbots should be able to give responses in ways, or on topics, that some people might find offensive	5.31	4.86	5.41	6.13	5.50	4.72	5.35	5.61	0.19	-0.14	-0.06	-0.514**
Users should be able to control the level of AI chatbot predictability or unpredictability	7.03	6.97	6.91	7.28	7.43	7.32	7.35	7.58	0.405*	0.35	0.437**	0.30
AI chatbots should be predictable and inoffensive by default	6.92	6.83	7.14	7.35	7.07	6.62	7.01	7.66	0.15	-0.21	-0.13	0.31
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from international organizations, regardless of whether this conflicts with local or country-level perspectives	6.30	7.06	5.97	7.49	6.17	6.70	6.28	6.77	-0.13	-0.363*	0.31	-0.718***
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, regardless of their human rights records or treatment of marginalized groups	5.10	4.21	4.61	6.17	5.27	4.31	4.70	5.42	0.17	0.10	0.09	-0.751***
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's national organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	6.28	6.60	6.38	6.23	6.64	7.02	6.71	6.41	0.356*	0.425*	0.32	0.18
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, regardless of their human rights records or treatment of marginalized groups	4.86	3.68	4.46	5.73	5.17	3.89	4.29	5.44	0.31	0.21	-0.17	-0.29
AI chatbots should provide the tradeoffs on a topic, drawing perspectives from the user's local organizations, unless the perspectives are inconsistent with fundamental human rights or marginalize some groups	6.18	6.39	6.33	6.43	6.50	6.78	6.59	6.35	0.319*	0.388*	0.26	-0.08

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test

AI chatbots should provide the tradeoffs to a topic from the country in which they were created	6.01	5.14	5.35	7.70	6.44	5.65	5.77	7.73	0.432*	0.505**	0.42	0.02
To ensure information is culturally relevant, AI chatbots should cite users' local media outlets and their national organizations' guidance, even if it contradicts the AI chatbot maker's values	6.73	6.78	6.98	6.96	6.94	6.90	7.14	6.81	0.21	0.12	0.15	-0.15
When citing a source, AI chatbots should only cite peer-reviewed scientific information, or discussions in major press outlets	6.41	7.95	6.93	6.60	6.76	7.83	7.18	7.09	0.348*	-0.13	0.25	0.483**
AI chatbots should source their information primarily from globally recognized authoritative sources (e.g. WHO)	6.36	7.71	7.71	7.72	6.64	7.63	7.71	7.94	0.277*	-0.07	0.00	0.21
AI chatbots should source their information primarily from sources from the user's national organizations	5.92	5.32	6.00	6.99	6.35	5.80	6.27	7.10	0.429**	0.479***	0.27	0.11
AI chatbots should use the user's past conversations to offer the best user experience, even if the user is not informed	5.11	4.14	5.04	5.08	4.15	3.23	4.34	4.02	-0.965***	-0.914***	-0.70	-1.058***
AI chatbots should use the user's past conversations to offer the best user experience if the user is informed	6.87	7.35	7.17	7.88	6.96	7.10	7.67	7.76	0.09	-0.25	0.49	-0.12
AI chatbots should use past conversations to offer the best user experience if the user is informed and they are able to access and delete their chat history	6.98	7.60	7.32	7.95	7.50	7.84	8.15	8.27	0.523***	0.24	0.83	0.32
AI chatbots should use additional data sources, such as the user's online activity, to help personalize their interactions	5.18	4.04	5.74	7.01	5.07	3.92	5.88	6.73	-0.10	-0.12	0.14	-0.28
AI chatbots should use additional data sources, such as the user's online activity, with the permission of the user, to help personalize their interactions	6.58	5.81	6.99	8.09	6.78	6.01	7.26	7.95	0.20	0.20	0.26	-0.15
AI chatbots should prioritize standardized responses that do not rely on user data or their online activity	6.65	5.82	6.12	6.16	6.78	6.17	6.26	6.88	0.13	0.352*	0.14	0.719***

Note: "*" indicates a p-value of 0.05, "** indicates a p-value of 0.01, and "***" indicates a p-value of 0.001 resulting from a paired or independent t-test