Does Providing a Balanced View Alter Attitudes to Artificial Intelligence?

Chet Gutwein | Kiersten Henderson | Zhaoning Yu

*All data and statistical analysis supporting this project is available at https://github.com/cgutwein/W241_AI_Experimental_Survey

Introduction

The rise of Artificial Intelligence (AI) in increasingly clever software and machines promises unprecedented gains to social and economic welfare. It is hypothesized that AI-based superintelligence will create revolutionary new technologies that could help us eradicate war, disease, and poverty. However, there is the potential for AI systems to cause great harm intentionally or unintentionally [1].

Hollywood frequently portrays AI as “robots taking over the world” and comments from Science and Tech giants Bill Gates, Elon Musk, and Stephen Hawking warn that AI could be the biggest existential threat that humanity faces [2]. Although the news media does often portray a balanced view of AI, we wondered how well-exposed the average person is to both the pros and cons of AI-based technologies and to what extent this shapes their attitudes toward AI.

A recent study found that people who watched sci-fi films that either depicted AI in a positive or negative light actually became more extreme in the view they held of AI prior to watching the film [3] (more supportive or more skeptical of AI) . Thus, when consuming information about AI from a film source, people exhibited confirmation bias. We wondered how presenting people with both the pros and cons of AI-based technologies (a balanced view) in a non-entertainment setting would affect their attitudes toward AI.

Therefore we asked:

Does Providing People with a Balanced View of the Benefits and Risks of Artificial Intelligence Alter their Attitude Towards Artificial Intelligence?

To address this question, we evaluated the opinions of respondents to the following AI topics either with or without an accompanying educational paragraph:

1) How strongly do you feel that widespread adoption of AI technology will be beneficial to society?
2) How strongly do you feel that that widespread adoption of AI technology will have dangerous or undesirable consequences?
3) How concerned are you that widespread adoption of Artificial Intelligence will lead to major job losses across several employment sectors?

4) How strongly do you feel that regulatory oversight should be established prior to the widespread adoption of particular AI technologies?
5) How likely are you to value a movie recommendation from Netflix over a friend or family member?

Our findings are important because how people feel about AI-based technologies will likely affect how technologies are implemented. If opinions change based on learning the pros and cons of AI, we could apply this knowledge to shape policies that are implemented to educate the public so they can make informed decisions.

Experimental Design

Our instrument for delivering treatment to subjects who participated in our study was a survey (implemented through Qualtrics). Subjects were asked the same five (5) questions related to Artificial Intelligence. Each question has two versions, a control version (where there is no supporting text) and a treatment version where the question is preceded by a short paragraph outlining the positive and negative viewpoints related the question. We utilized Amazon's Mechanical Turk (MTurk) service to hire participants. MTurk proved very useful allowing for quick and convenient collection of data, while also allowing us to set the following minimum parameters for subjects:

- All subjects live in the United States
- All subjects are unique and cannot be paid to take the survey more than once

We implemented our treatment utilizing a within-subjects design approach. To do this, respondents received both treatment and control versions of the survey questions.
To avoid potential spillover effects, we presented control questions prior to the treatment questions (i.e. that the respondents learn something in a treatment question that affects their control response).

Respondents were asked a total of five questions (see **Appendix A**). Two of five control questions were presented in a randomized order. We then asked three treatment questions. The three treatment questions were presented in randomized order and respondents received "new" treatment questions (i.e. none of the treatment version of the control question the respondent was already asked).

There could be spillover from one treatment question that affected the responses to subsequent treatment questions. To mitigate these potential effects, we randomized the order in which the treatment questions were presented.

Covariates

AI Knowledge Prior to Survey

We expect *Heterogeneous Treatment Effect* if people are well informed about AI already because this information will add to their knowledge and the treatment may have less or no effect on their opinion. Therefore, prior to the control and treatment questions, we asked:

How well educated are you about the current state of artificial intelligence technology?

Attitude Toward AI Prior to Survey

We also expected *Heterogeneous Treatment Effects* based on respondents attitudes towards AI prior to treatment exposure. We therefore asked:

How do you feel in general about the widespread adoption of artificial intelligence technology?

Covariate Collection

To avoid any potential *priming effects* we requested the covariate information after the control and treatment question responses were collected. Based on a recent Pew Research Center Study of AI-related automation [4], we knew that gender, education and income level affect people's attitudes towards AI and so we collected this information.

Measures of Compliance

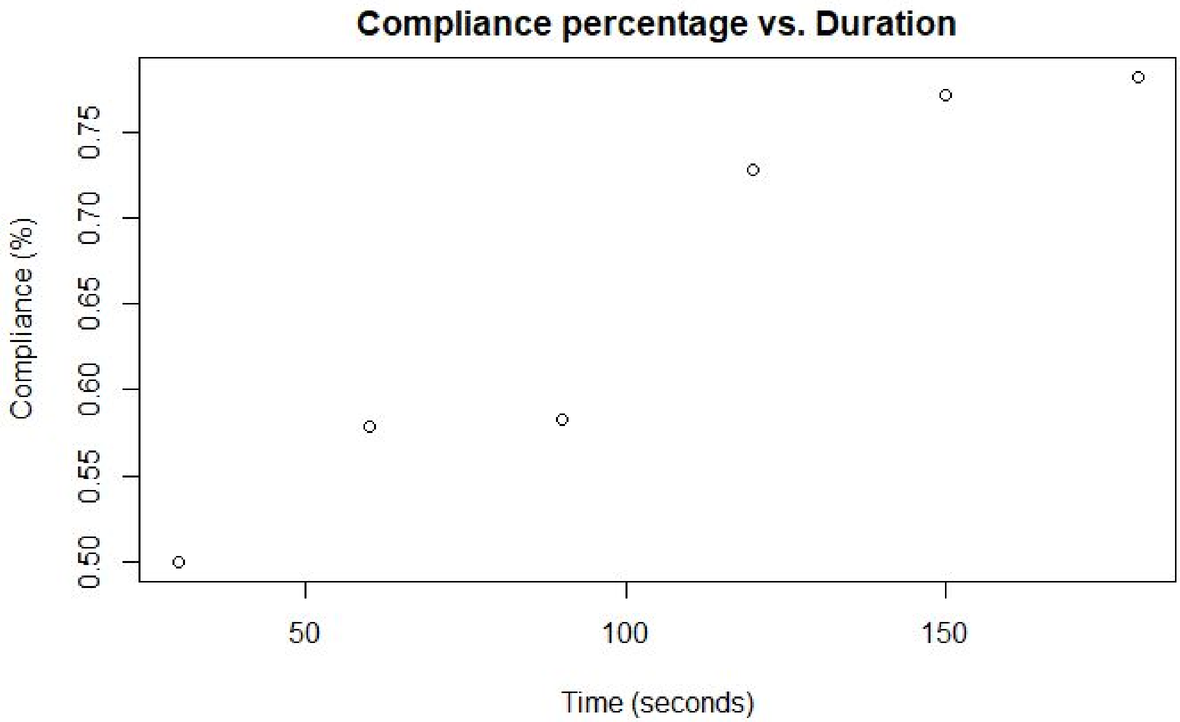
We anticipated that there would be a percentage of participants who would not fully read the treatment message prior to answering the question. This is a form of non-compliance that we addressed by integrating an "attention question" into each survey.

Attention Question Reponses

An "attention question" is a validity check that was used to test whether a participant has fully read through the treatment message by testing his/her knowledge immediately after the treatment message has appeared. In the survey, after the first treatment question the participant was asked the attention question. All subjects who answered the the "attention question" incorrectly were considered Never Takers. We considered these results when calculating the complier average causal effect (CACE).

Survey Completion Time

After collecting the results from Pilot Study 1, we considered alternative measures for Non-Compliance. A comparison of the duration of time each subject took to complete the survey to the result of his/her answer to the attention question showed a correlation. Because the "attention question" was a True/False question and Never Takers would still get it right 50% of the time, we decided it may be appropriate to use the duration as another validity check. We ultimately decided against this, however, because we cannot be sure how fast a subject can read the treatment message in a survey.



Respondent self-evaluated degree of learning from survey

In order to determine if we managed to apply a treatment (ie. teach respondents something they didn't know) we asked respondents the following question:

Did you learn anything new about AI while taking this survey?

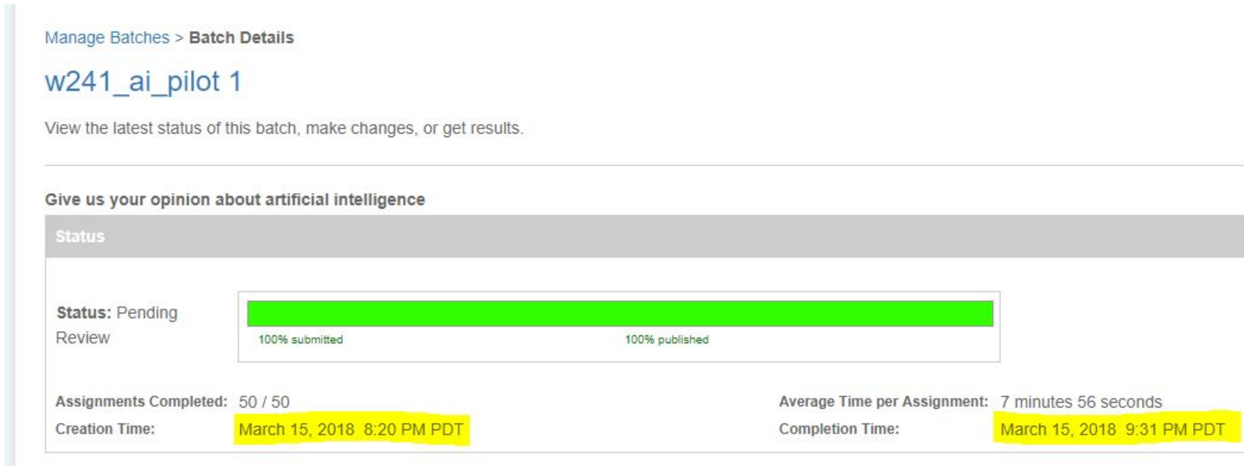
Results of Pilot Studies

We performed two pilot studies in order to evaluate the effectiveness of our survey design and to try to optimize respondent treatment compliance.

Pilot Study 1

Prior to running our experiment, we collected survey data for a small number of participants using a beta version of the survey and used this as a pilot study. The purpose of this was to ensure the functionality of the survey and to identify any significant issues that might arise when we launch the survey for the full experiment. We collected data from 50 MTurk Workers.

Since our team had no experience using MTurk Workers as subjects, our pilot study was mainly focused on how successful we were able to administer the survey through this channel. The most significant takeaway was the speed at which Workers responded and completed the batch. It took just over 1 hour to collect data from 50 participants.



From this, we considered the impact this would have on the experiment. To ensure we collected data over a longer period of time, we decided to break up the data collection for the experiment such that results were not subject to selection bias based on the time of day or day of week that responses were collected.

In addition, we knew that the amount of compensation would factor heavily into how quickly Workers would complete the survey. For the pilot, we paid Workers $0.50 per completed survey. The completion rate we observed from the pilot suggested that we could pay Workers less and still collect a sufficient amount of responses. Thus, we used the results from the survey along with some assumptions and performed power calculations using Randomization Inference to determine how many subjects we would need for the experiment. With 1,000 survey participants, our experiment would detect a treatment effect with a magnitude of 0.2 or greater on our outcome variable 98.5 percent of the time.

A third significant observation was that 37% of Workers had completed the survey in less than 2 minutes, which seemed like it was not enough time to fully read the treatment passages. It is very difficult to ensure that subjects read each passage fully, and we identified the duration which it took each subject to complete the survey as a variable for identifying non-compliance. We ultimately decided to use answers to the attention question for determining compliance. However, this observation motivated a second pilot study, where the Masters criteria of MTurk Workers was used to filter eligible participants with the expectation that more qualified subjects would take more time to fully read the passages in the survey.

Results of Pilot Study 2

To ensure the highest degree of quality responses, we decided to test out MTurk's "Master" designation to filter eligible participants based on their MTurk rating. Per MTurk's website:

"Amazon Mechanical Turk (MTurk) has built technology which analyzes Worker performance, identifies high performing Workers, and monitors their performance over time. Workers who have demonstrated excellence across a wide range of HITs are awarded the Masters Qualification. Masters must continue to pass our statistical monitoring to maintain the MTurk Masters Qualification."

As we collected data from MTurk Masters, it was clear that this was an impractical technique for us to use. Over the course of 7 days, we collected responses from 30 MTurk Master Workers. While the Masters data was not useful in our data analysis, it did give us an additional point of reference from which to compare our final survey compliance and covariate results.

Results of Full Survey

Our results suggest that we did indeed apply a treatment to a majority of respondents; 64 % of respondents report that they learned something during the survey. This suggests that we succeeded in applying a treatment, but because it is a self-reported measure, it is not a good indication of compliance.

Evaluating Compliance

Based on whether respondents answered their attention question correctly, ~24% of respondents were Never-Takers. However, the attention question was a True or False question and so random guessing would produce the correct answer 50% of the time and thus potentially be an underestimate. However, based on our analysis of the time respondents took to answer the survey, we came to a similar estimate that ~26% of respondents were Never-Takers.

Question	Fraction Compliance
Undesirable Consequences	0.52
Jobs	0.74
AI Oversight	0.79
Recommender Systems	0.86
Beneficial to Society	0.86
Average all Questions	0.76

Analysis of Covariates

Consistent with demographics, we found that 52% of survey takers were women. In addition, we found that most people had at least a college level education. Furthermore, most respondents reported a household income of between 0-$50,000 a year.

Analysis at the Individual Question Level

We designed the survey so that we could perform an within-subjects analysis, however, we were also able to compare overall responses from control and treatment questions between subjects.

Our outcome variable in this case was the response that subjects gave regarding their attitude to the question posed and it was a Likert Variable. For the purpose of regression analysis, we assumed these variables were continuous.

We performed OLS regression analyses for each of the five questions posed to calculate the average treatment effect as ITT (**Regression Tables in Appendix B**) for each question. We then considered the average fraction of compliance when calculating the complier average causal effect. We divided the coefficient on the dummy variable for treatment assignment (assignment_jobs, assignment_oversight, assignment_beneficial, assignement_unintended, assignement_recommender) by the overall compliance percentage to calculate the CACE.

We included the covariates we collected (gender, household income, education level), and the variables for respondents baseline attitude to AI and baseline education on AI during regression analysis. We also included an interaction term (treatment assignment **X** AI_educated) to account for the fact that individuals who were already highly educated regarding AI might not learn anything when we applied the treatment (ie. there would be Heterogeneous Treatment Effects) and thus their attitude might not change.

Overall, we found that the treatment altered attitudes towards job losses due to AI-enabled automation (**see Table below**). However we did not detect any change in attitudes to the other questions we asked due to treatment.

The intent to treat for attitudes towards job losses was -0.16 (p=0.019), but considering 76% average compliance, the CACE was -0.21. Therefore, on average, compliars (people who were reading the survey carefully) expected job losses due to AI to be less severe by a fifth of a point on a 5 point scale, which is not practically significant. We therefore found that survey takers became more positive in their attitude towards AI automation, but not by much.

This finding is robust because we find similar effects in multiple model specifications as long as we account for baseline attitude towards AI and prior knowledge of AI.

Question	ITT	p value	CACE
Jobs	-0.16	0.019 *	-0.21
Beneficial to Society	-0.02	0.58	-0.03
AI Oversight	- 0.05	0.49	-0.07

Recommender Systems	-0.04	0.49	-0.05
Undesirable Consequences	0.03	0.67	0.04

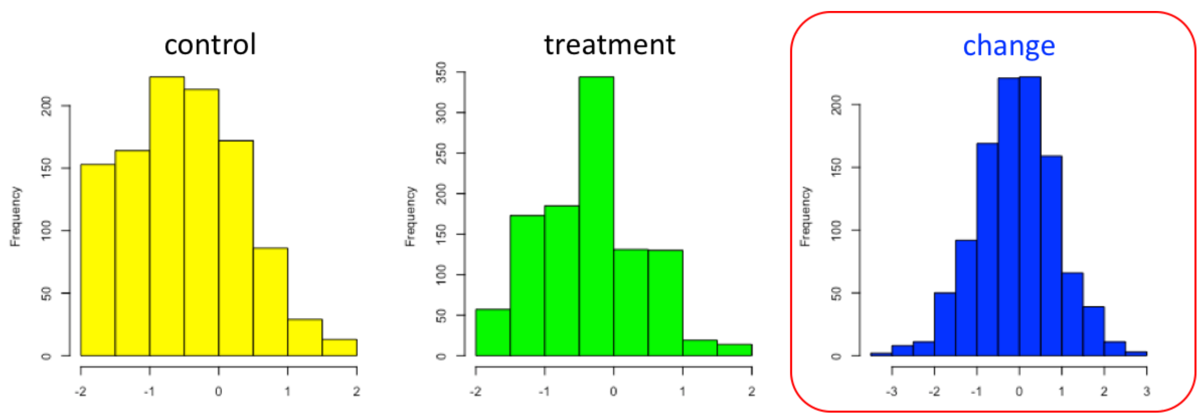
Analysis of Overall Respondent Attitude Change

In our survey, each subject first answers two randomly chosen control questions followed by three treatment questions. We can therefore perform analysis using a within-subjects experimental design. Rather than examining attitude by analyzing each question individually, we can look at how much of a change in attitude is induced by all the treatment questions as a whole. In this case, the treatment is defined as the exposure to three treatment questions that come with informational paragraphs.

To gauge a subject's attitude, we first converted the original Likert-scale answers with a range from 1 to 5 to an "attitude score" with a range from -2 to 2, under the assumption of linear correspondence. A positive score represents a more favorable attitude towards AI and vice versa.

Due to the differences in the phrasing of the questions, different conversions are applied to different questions. For questions 1 ("undesirable consequences"), 2 ("jobs") and 3 ("oversight"), a Likert-scale answer of 5 represents a more negative attitude toward AI and thus is given an attitude score of -2, and an answer of 1 is converted to an attitude score of 2. For questions 4 ("recommender") and 5 ("beneficial"), a Likert-scale answer of 5 represents a more favorable attitude toward AI and thus is given an attitude score of 2, and an answer of 1 is converted to an attitude score of -2.

After obtaining an attitude score for each question, we then define a variable "AI_attitude_control" as the average attitude score for the two control questions and another variable "AI_attitude_treat" as the average attitude score for the three treatment questions. The difference between these two variables is "AI_attitude_change" and is the observed treatment effect for each individual respondent.

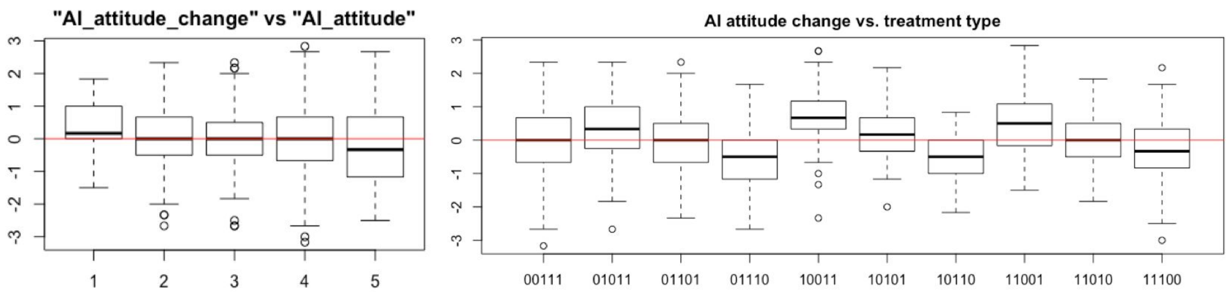


mean = -0.278
p-value < 2.2e-16

mean = -0.269
p-value < 2.2e-16

mean = 0.0087
p-value = 0.7618

Since this is a within-subjects design, we use the mean of “AI_attitude_change” to estimate the ATE. The average of “AI_attitude_control” is -0.278 and the average of “AI_attitude_treat” is -0.269, thus, the estimated ATE is 0.0087, which is not significantly different from zero (with a p-value of 0.7618 for the one-sample t-test). Thus, we conclude that on average, the treatment of exposure to informative messages did not alter people’s overall attitude toward AI.



It is worth noting that although the overall treatment effect is zero, there are individual differences that are dependent on the individual’s self-reported attitude toward AI (“AI_attitude”) and the specific combination of treatment questions (“treatment_type”) an individual was exposed to. Specifically, after controlling for the treatment_type fixed effects and other covariates, our regression analysis of “AI_attitude_change” found a small but statistically significant coefficient of -0.086 (with a p-value < 0.01) on the variable “AI_attitude” (Regression Table in Appendix C). This indicates that for individuals with more extreme views toward AI before taking the survey, the treatment appears to nudge their attitudes slightly to the center. Practically, for every ten people taking the survey, the treatment is sufficient to shift one person’s answers to the treatment questions by about 1 point to the center on the Likert-scale.

Discussion

We found that respondents given a treatment paragraph thought more positively about the severity of job losses that could result from AI-enabled automation. Again, this shift in attitude was not large, but considering that we only delivered a few sentences of treatment in an online survey, it is intriguing.

On average, we did not find a change in *average overall attitude to AI* in response to providing a paragraph of information about AI. We anticipated the absence of an *average change* in *overall AI attitude* because we hypothesized that attitudes to AI might often be driven by emotion. We therefore hypothesized that providing respondents with a balanced view of AI might moderate extreme attitudes - shift those with the most negative and most positive attitudes towards a more neutral attitude. Indeed, this is what we observed. When we controlled for fixed effects, we found that people with very positive baseline attitudes towards AI became slightly more negative in attitude following treatment. Conversely, respondents who initially had a very negative attitude to AI became slightly more positive in attitude following treatment.

Mediation
There are several mechanisms by which the treatment we applied may have affected attitudes to AI. Our hypothesis was that by learning more, individuals behave logically and incorporate this information into their decision making to change their attitude. However, it is also possible that reading a balanced message that delivers both pros and cons cues respondents to behave in a more balanced manner and to respond more rationally than they might otherwise.

Potential for Interference in Study
The “Rejection” feature of MTurk was used during the first batch of our experiment, where we bulk rejected the work of 107 Workers for completing the survey in less than 2 minutes. Of the Workers rejected, nearly half responded with an email expressing their surprise and dismay. This came as a surprise to our team and also a concern as we considered the consequences of potential spillover effects. A few of the Workers’ emails indicated that there was a forum or online community where our Rejections were being discussed, and one Worker even tweeted directly to a member of our research team.

As soon as we realized this we initiated retroactive approvals of the work of all 107 Workers whose work had been rejected to mitigate any further discussion with other potential subjects. In addition, we delayed the second batch of data collection via MTurk for 24 hours to dissipate the spillover effects.

Although this was a misstep in our data collection practices, it is likely inconsequential. Potential survey takers who learned of other Workers’ disappointment regarding our survey would likely be deterred from taking it. Despite this, we were able to collect 1,057 survey responses.

Generalizability
Our conclusions should be interpreted with the caveat that our subjects may not be representative of the general population. MTurk workers are already likely to be comfortable with technology (they use computers for their jobs) and they have a technology-enabled work environment (remote work for Mechanical Turk). MTurk workers might therefore have more positive views towards AI than most people and this could influence the degree to which our treatment would change their attitudes.

Conclusion
Our findings suggests that campaigns that deliver a balanced view of topics affected by AI technology could prove effective at garnering public support for AI technologies that automate work but lead to job losses. In addition, future studies could build on our observation that people with the most extreme views of AI technology might be the most responsive to balanced messaging by looking for a larger treatment effect if a more resource-intensive form of informative messaging (longer treatment exposures) is provided.

Sources

1) https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/
2) *Clever Computers: The Dawn of Artificial Intelligence.* May 2015. The Economist.
3) Vyacheslav Polonski. *People Don't Trust AI--Here's How We Can Change That.* January 2018. Scientific American.
4) The Pew Research Center. *Automation in Everyday Life*. May, 2017.

APPENDIX A - Qualtrics Survey

How well educated are you about the current state of artificial intelligence technology?

Novice				Expert
1	2	3	4	5

Please give yourself a rating from 1 to 5.

How do you feel in general about the adoption of artificial intelligence technology?

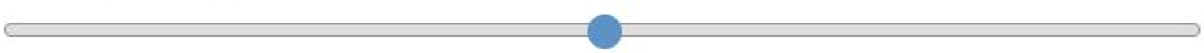
Against		Neutral		the more AI the better
1	2	3	4	5

Please give yourself a rating from 1 to 5.

How strongly do you feel that widespread adoption of AI technology will be beneficial to society?

Not at all				Very strongly
1	2	3	4	5

Use the slider and rate yourself from 1 to 5.



How strongly do you feel that that widespread adoption of AI technology will have dangerous or undesirable consequences?

Not at all				Very Strongly
1	2	3	4	5

Use the slider and rate yourself from 1 to 5.

Please read the following passage and then answer the question below.

Artificial Intelligence (AI) technology has been implemented in self-driving vehicles, used to make parole decisions for prisoners, and utilized to decide who will get bank loans among many controversial use cases. There is much debate surrounding whether or not AI should be regulated.

Some policy experts think that many routine tasks can be automated without the need for regulating how these tasks are accomplished or which tasks can be automated. Some people feel that regulations can be established after the fact - if a particular technology is found to impose on human rights.

Other people are concerned that a digital superintelligence will be developed and that it will pose a significant threat to people. Some call for regulatory oversight before particular technologies are developed so that the potential for human harm could be limited.

How strongly do you feel that regulatory oversight should be established prior to the widespread adoption of particular AI technologies?

Not at all				Very strongly
1	2	3	4	5

Use the slider and rate yourself from 1 to 5.

Based on the information provided in the survey, please answer the following True/False question to the best of your knowledge.

"Guidelines for the regulation of AI technology have been established."

- True
- False

Please read the following passage and then answer the question below.

Historically, the relationship between automation and employment has been complicated. Economists report that technology has increased total employment, rather than reducing it. Estimates of the percentage of total American jobs that are at "high risk" of potential automation due to Artificial Intelligence (AI) technology vary between 9 and 47% of jobs. Jobs categorized as "high risk" include those that are routine, like paralegals or fast food cooks. However, jobs that cannot be automated, like those that are care-related (personal healthcare or the clergy) are likely to increase.

How concerned are you that widespread adoption of Artificial Intelligence will lead to major job losses across several employment sectors?

Not at all				Very concerned
1	2	3	4	5

Use the slider and rate yourself from 1 to 5.

Please read the following passage and then answer the question below.

Machine learning technology has enabled computer programs called Recommender Systems to recommend products and services. These systems work by identifying customers who are similar to you in particular ways. Depending on the company using the Recommender System, the system can sometimes identify tens of thousands of people who are similar to you and make you suggestions based on what these other people liked.

The ways Recommender System can measure your preferences are limited, however the systems have access to data from large numbers of customers. Therefore, Recommender Systems tend to predict customers preference very accurately.

In contrast, people who know you, probably know more about you than a Recommender System does. Many people find the constant trafficking of personal data invasive. Furthermore, having a computer replace a friend's recommendation for a movie or TV show constitutes a threat to humanity for some.

How likely are you to value a movie recommendation from Netflix over a friend or family member?

Not at all				Very likely
1	2	3	4	5

Use the slider and rate yourself from 1 to 5.

Did you learn anything new about AI while taking this survey?

- Yes
- No

What is the highest level of education that you have completed?

- (●) High School
- ( ) Bachelors Degree
- ( ) 0 - 2 years of graduate school
- ( ) > 2 years of graduate school
- ( ) Other

Which of the following best describes your annual household income?

- (●) 0 - $50,000
- ( ) $50,000 - $100,000
- ( ) $100,000 - $150,000
- ( ) $150,000 or greater

What gender do you identify with?

- (●) Male
- ( ) Female

Thank you for participating in our survey.
You may now enter the following code into Mechanical Turk to receive your reward:

4724952

APPENDIX B - OLS Regression Analysis of Individual Question Responses

Question 1:

beneficial

	Dependent variable:				
	score_beneficial				
	(1)	(2)	(3)	(4)	(5)
assignment_beneficial	-0.035	-0.025	-0.047	-0.031	0.326*
	(0.064)	(0.045)	(0.063)	(0.044)	(0.144)
AI_attitude		0.694***		0.689***	0.688***
		(0.025)		(0.025)	(0.025)
AI_educated		0.045		0.027	0.103**
		(0.025)		(0.025)	(0.039)
education_level			0.012	0.007	0.010
			(0.031)	(0.022)	(0.022)
house_income			0.024	0.017	0.015
			(0.038)	(0.028)	(0.027)
gender			0.322***	0.145**	0.146**
			(0.061)	(0.045)	(0.045)
assignment_beneficial:AI_educated					-0.130**
					(0.050)
Constant	3.337***	0.966***	3.123***	0.921***	0.714***
	(0.052)	(0.102)	(0.093)	(0.113)	(0.141)
Observations	1,053	1,047	1,052	1,046	1,046
R2	0.0003	0.478	0.028	0.484	0.487
Adjusted R2	-0.001	0.477	0.024	0.481	0.484
Note:			*p<0.05; **p<0.01; ***p<0.001		

Question 2:

jobs

	Dependent variable:				
	score_jobs				
	(1)	(2)	(3)	(4)	(5)
assignment_jobs	-0.104	-0.159*	-0.108	-0.164*	-0.485*
	(0.071)	(0.068)	(0.071)	(0.068)	(0.218)
AI_attitude		-0.377***		-0.377***	-0.376***
		(0.037)		(0.038)	(0.037)
AI_educated		0.135***		0.143***	0.074
		(0.038)		(0.040)	(0.060)
education_level			-0.053	-0.052	-0.047
			(0.034)	(0.032)	(0.032)
house_income			0.015	0.007	0.006
			(0.045)	(0.043)	(0.043)
gender			-0.055	-0.029	-0.026
			(0.070)	(0.069)	(0.069)
assignment_jobs:AI_educated					0.118
					(0.076)
Constant	3.447***	4.325***	3.557***	4.414***	4.590***
	(0.054)	(0.146)	(0.109)	(0.170)	(0.198)
Observations	1,053	1,047	1,052	1,046	1,046
R2	0.002	0.097	0.005	0.099	0.102
Adjusted R2	0.001	0.094	0.001	0.094	0.095
Note:	*p<0.05; **p<0.01; ***p<0.001				

Question 3:

oversight

	Dependent variable:				
	score_oversight				
	(1)	(2)	(3)	(4)	(5)
assignment_oversight	0.068	0.060	0.072	0.063	-0.198
	(0.068)	(0.067)	(0.068)	(0.067)	(0.221)
AI_attitude		-0.154***		-0.151***	-0.155***
		(0.039)		(0.039)	(0.039)
AI_educated		0.021		0.033	-0.028
		(0.036)		(0.038)	(0.063)
education_level			-0.033	-0.032	-0.031
			(0.031)	(0.031)	(0.031)
house_income			0.020	0.014	0.014
			(0.042)	(0.042)	(0.043)
gender			-0.124	-0.095	-0.094
			(0.066)	(0.068)	(0.068)
assignment_oversight:AI_educated					0.095
					(0.076)
Constant	3.842***	4.288***	3.933***	4.330***	4.506***
	(0.054)	(0.150)	(0.101)	(0.170)	(0.225)
Observations	1,053	1,047	1,052	1,046	1,046
R2	0.001	0.020	0.005	0.023	0.024
Adjusted R2	0.00002	0.017	0.002	0.017	0.018
Note:	*p<0.05; **p<0.01; ***p<0.001				

Question 4:

recommender

	Dependent variable:				
	score_recommender				
	(1)	(2)	(3)	(4)	(5)
assignment_recommender	-0.042	-0.045	-0.037	-0.046	-0.160
	(0.068)	(0.065)	(0.068)	(0.065)	(0.206)
AI_attitude		0.280***		0.283***	0.283***
		(0.036)		(0.036)	(0.036)
AI_educated		0.156***		0.166***	0.141*
		(0.037)		(0.038)	(0.056)
education_level			0.006	0.001	0.0003
			(0.035)	(0.033)	(0.033)
house_income			0.045	0.035	0.035
			(0.042)	(0.040)	(0.041)
gender			0.063	-0.073	-0.074
			(0.068)	(0.067)	(0.067)
assignment_recommender:AI_educated					0.041
					(0.072)
Constant	2.755***	1.428***	2.632***	1.370***	1.434***
	(0.053)	(0.137)	(0.103)	(0.160)	(0.194)
Observations	1,053	1,047	1,052	1,046	1,046
R2	0.0003	0.099	0.003	0.101	0.102
Adjusted R2	-0.001	0.096	-0.001	0.096	0.096
Note:	*p<0.05; **p<0.01; ***p<0.001				

Question 5:

undesirable

	Dependent variable:				
	score_undesirable				
	(1)	(2)	(3)	(4)	(5)
assignment_undesirable	-0.140*	-0.026	-0.145*	-0.028	0.039
	(0.068)	(0.060)	(0.068)	(0.061)	(0.198)
AI_attitude		-0.535***		-0.534***	-0.534***
		(0.033)		(0.033)	(0.033)
AI_educated		0.219***		0.220***	0.235***
		(0.035)		(0.035)	(0.052)
education_level			0.004	0.002	0.003
			(0.034)	(0.031)	(0.031)
house_income			0.036	0.023	0.024
			(0.045)	(0.040)	(0.040)
gender			-0.058	-0.019	-0.019
			(0.068)	(0.062)	(0.062)
assignment_undesirable:AI_educated					-0.025
					(0.069)
Constant	3.220***	4.277***	3.181***	4.239***	4.197***
	(0.051)	(0.134)	(0.107)	(0.152)	(0.192)
Observations	1,053	1,047	1,052	1,046	1,046
R2	0.004	0.210	0.005	0.211	0.211
Adjusted R2	0.003	0.208	0.002	0.206	0.205
Note:					*p<0.05; **p<0.01; ***p<0.001

APPENDIX C - Regression Analysis on Individual's AI Attitude Change

AI attitude change

	Dependent variable: AI_attitude_change					
	(1)	(2)	(3)	(4)	(5)	(6)
AI_attitude	-0.100***	-0.088**	-0.086**	-0.085**	-0.085**	-0.086**
	(0.029)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
AI_educated		-0.045	-0.045	-0.045	-0.045	-0.045
		(0.029)	(0.030)	(0.030)	(0.030)	(0.031)
education_level			0.046	0.046	0.045	0.046
			(0.027)	(0.027)	(0.027)	(0.027)
house_income			-0.039	-0.038	-0.038	-0.038
			(0.033)	(0.033)	(0.033)	(0.033)
gender			-0.034	-0.038	-0.037	-0.038
			(0.055)	(0.055)	(0.055)	(0.055)
learn_anything				-0.032	-0.032	-0.033
				(0.055)	(0.055)	(0.055)
duration_bin					0.005	0.003
					(0.061)	(0.063)
attention_correct						0.021
						(0.064)
treatment_type01011	0.368**	0.365**	0.364**	0.360**	0.360**	0.361**
	(0.132)	(0.133)	(0.134)	(0.134)	(0.134)	(0.134)
treatment_type01101	-0.067	-0.066	-0.051	-0.050	-0.050	-0.047
	(0.119)	(0.120)	(0.121)	(0.120)	(0.120)	(0.121)
treatment_type01110	-0.576***	-0.582***	-0.577***	-0.580***	-0.580***	-0.578***
	(0.121)	(0.122)	(0.122)	(0.122)	(0.122)	(0.122)
treatment_type10011	0.812***	0.808***	0.807***	0.808***	0.808***	0.810***
	(0.122)	(0.123)	(0.123)	(0.123)	(0.123)	(0.124)
treatment_type10101	0.236*	0.231	0.238	0.239*	0.239*	0.241*
	(0.120)	(0.122)	(0.121)	(0.121)	(0.122)	(0.122)
treatment_type10110	-0.438***	-0.440***	-0.429***	-0.430***	-0.430***	-0.426***
	(0.111)	(0.112)	(0.112)	(0.113)	(0.113)	(0.113)
treatment_type11001	0.505***	0.507***	0.512***	0.508***	0.509***	0.512***
	(0.136)	(0.136)	(0.135)	(0.135)	(0.136)	(0.136)
treatment_type11010	0.040	0.044	0.058	0.055	0.055	0.061
	(0.117)	(0.117)	(0.117)	(0.117)	(0.117)	(0.119)
treatment_type11100	-0.309*	-0.311*	-0.303*	-0.306*	-0.306*	-0.307*
	(0.135)	(0.135)	(0.135)	(0.135)	(0.135)	(0.136)
Constant	0.288*	0.374*	0.349*	0.372*	0.367*	0.353
	(0.130)	(0.148)	(0.161)	(0.167)	(0.175)	(0.182)
Observations	1,047	1,047	1,046	1,046	1,046	1,044
R2	0.199	0.201	0.204	0.205	0.205	0.205
Adjusted R2	0.191	0.193	0.194	0.193	0.192	0.192

Note: *p<0.05; **p<0.01; ***p<0.001