

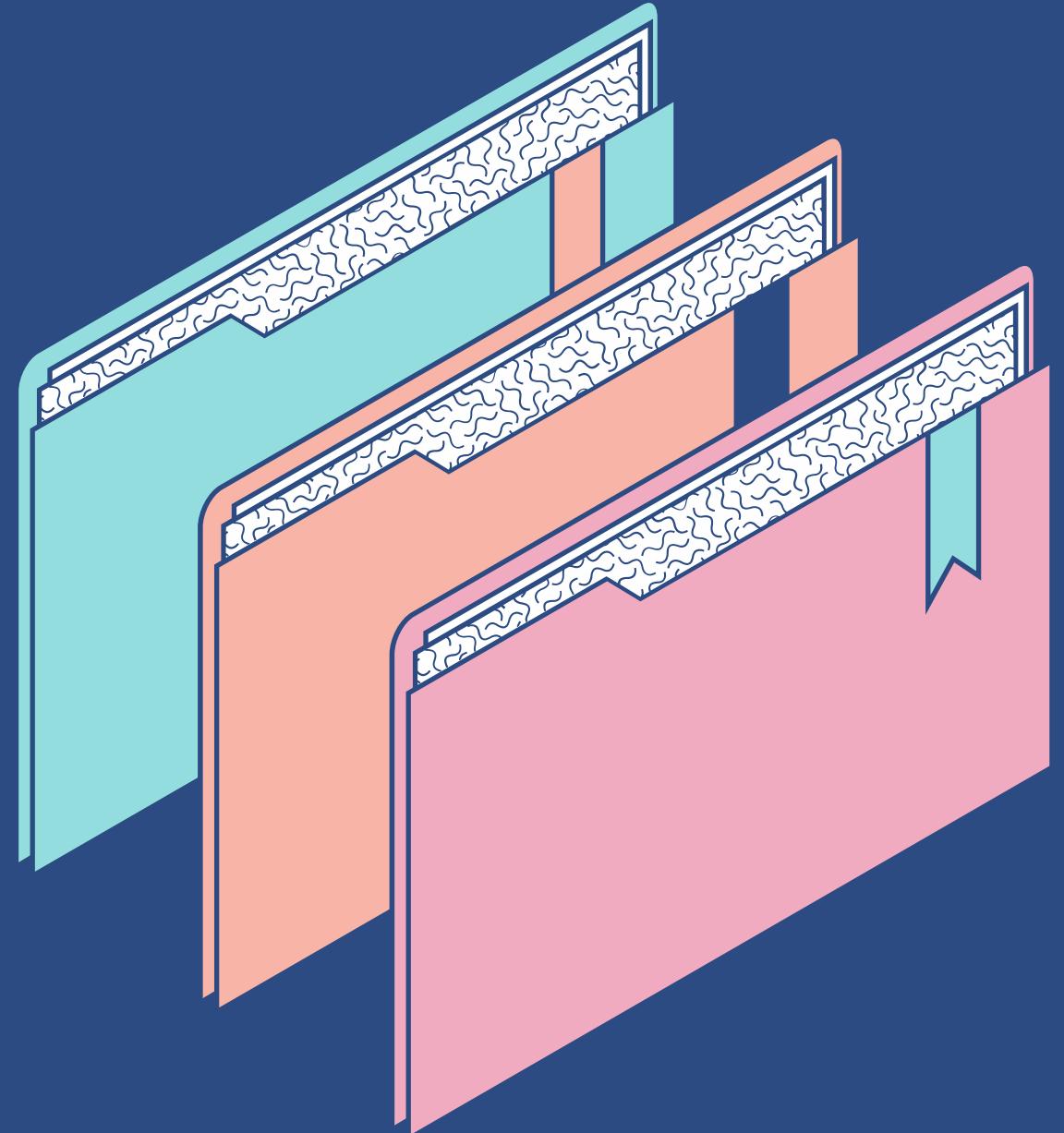


LING 461 EXPERIMENTAL SEMANTICS

Human vs AI irony

Research team: Fariza, Ayaulyym,
Leila, Ayana

Our Research Team



Fariza - ideas, design of experiment,
abstract

Ayaulym - stimuli creation, coding

Leila - abstract, ideas, participants

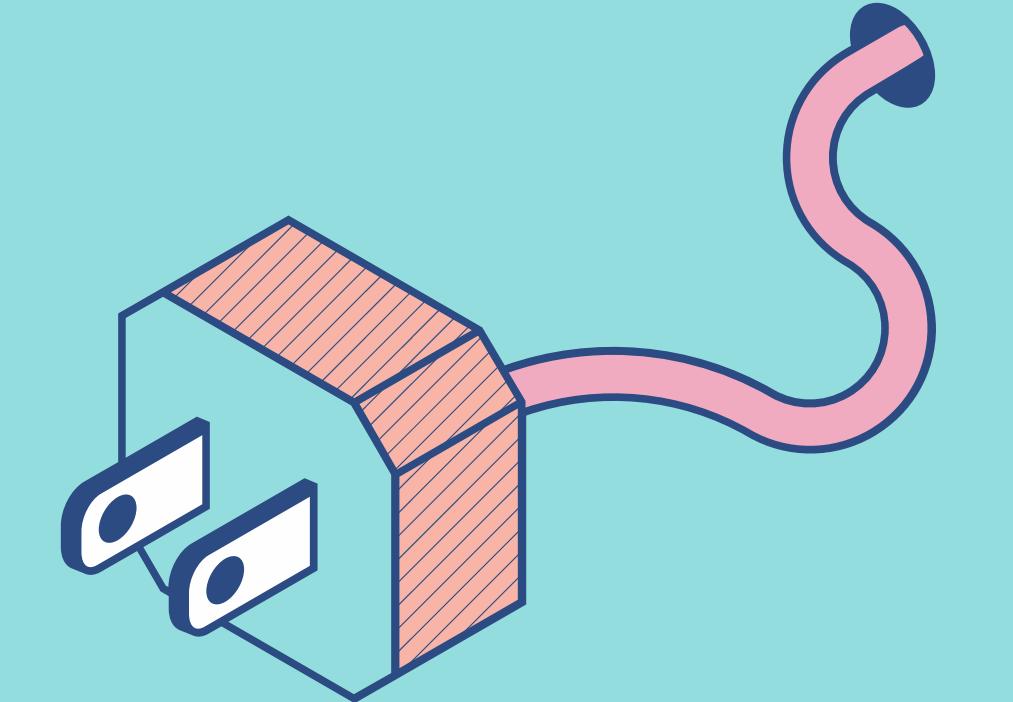
Ayana - collecting data, stimuli
creation

Ai vs Human. Irony!

memes!



sarcasm?



jokes :)

context

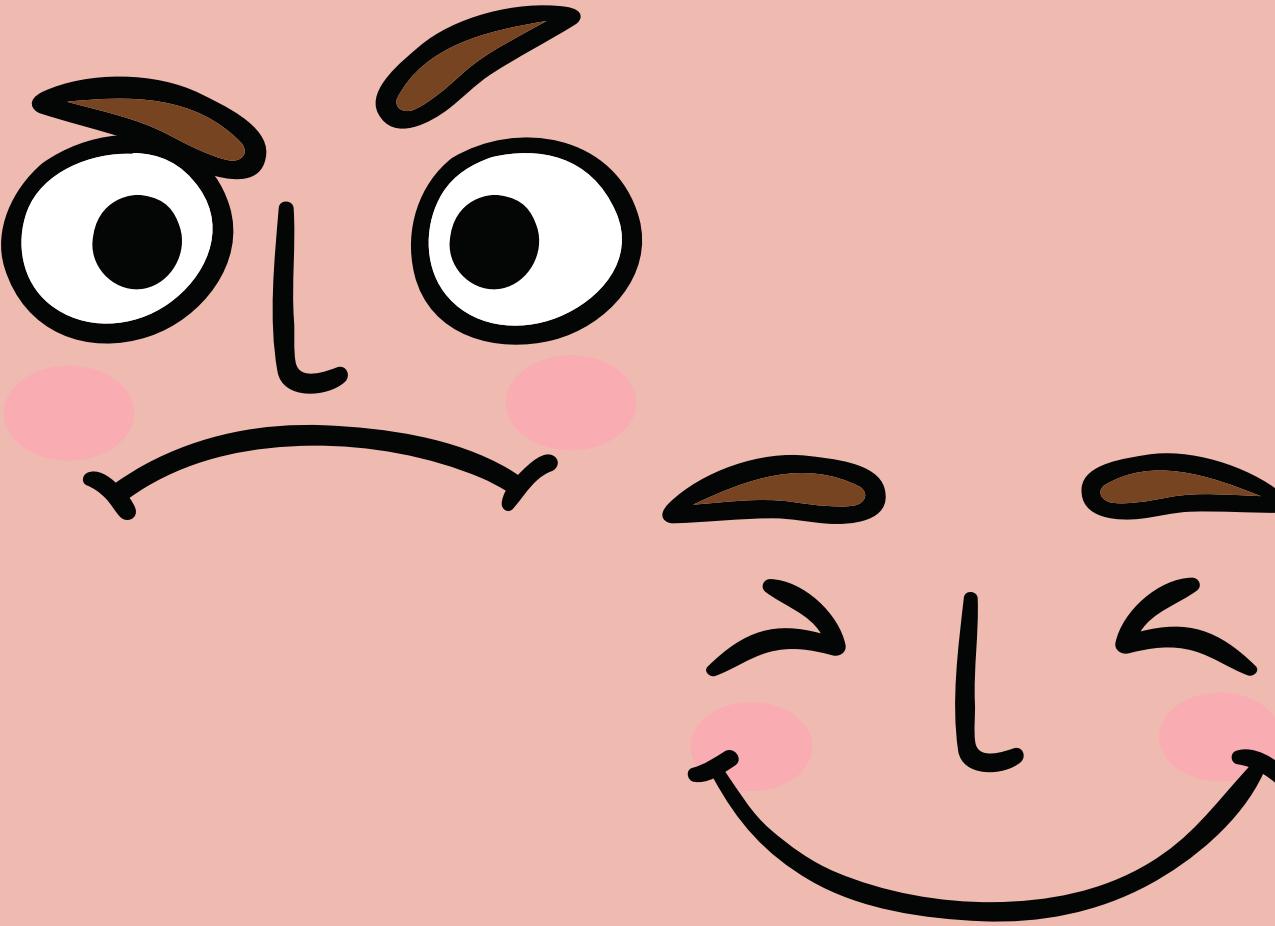


Irony?

voice tone

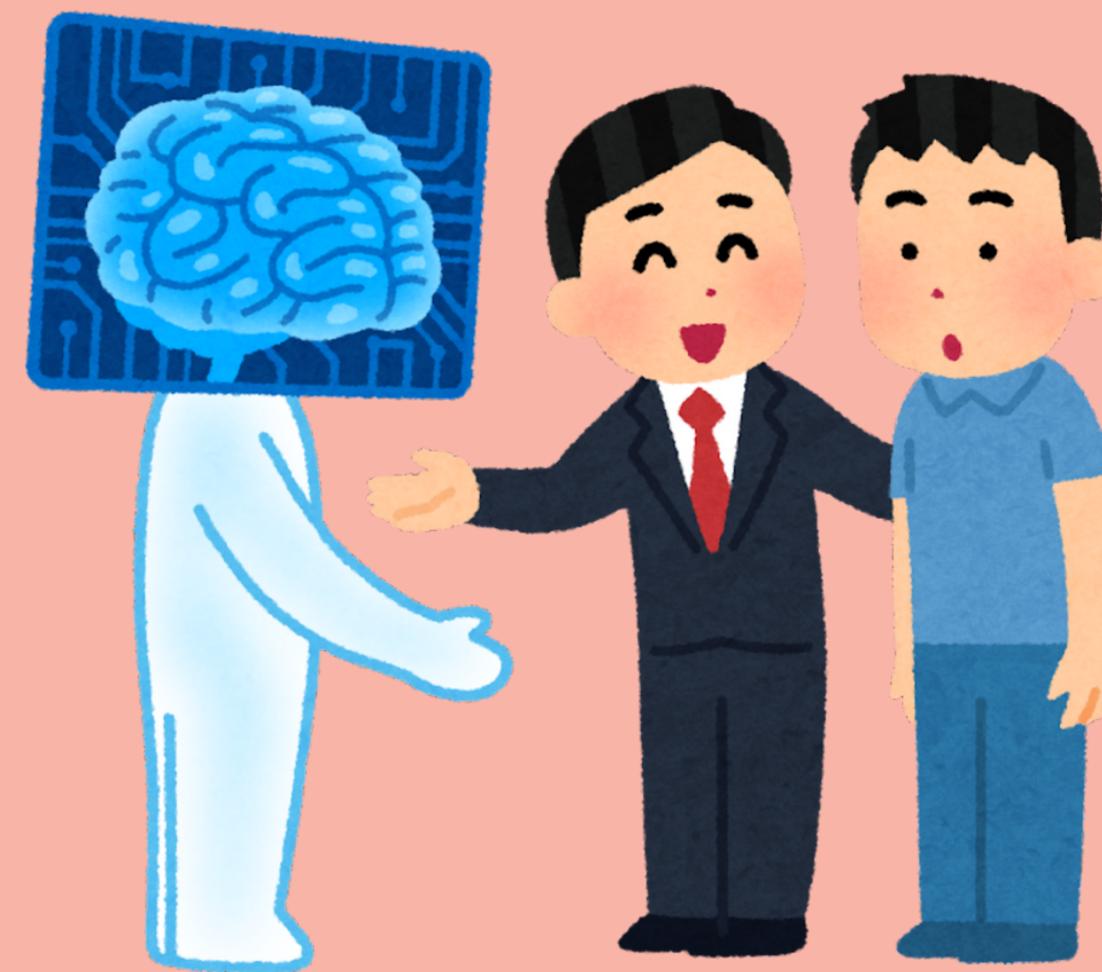


facial expressions



pauses

Ai generated sentences



Ai interpretation

Human perceptions

only listen once



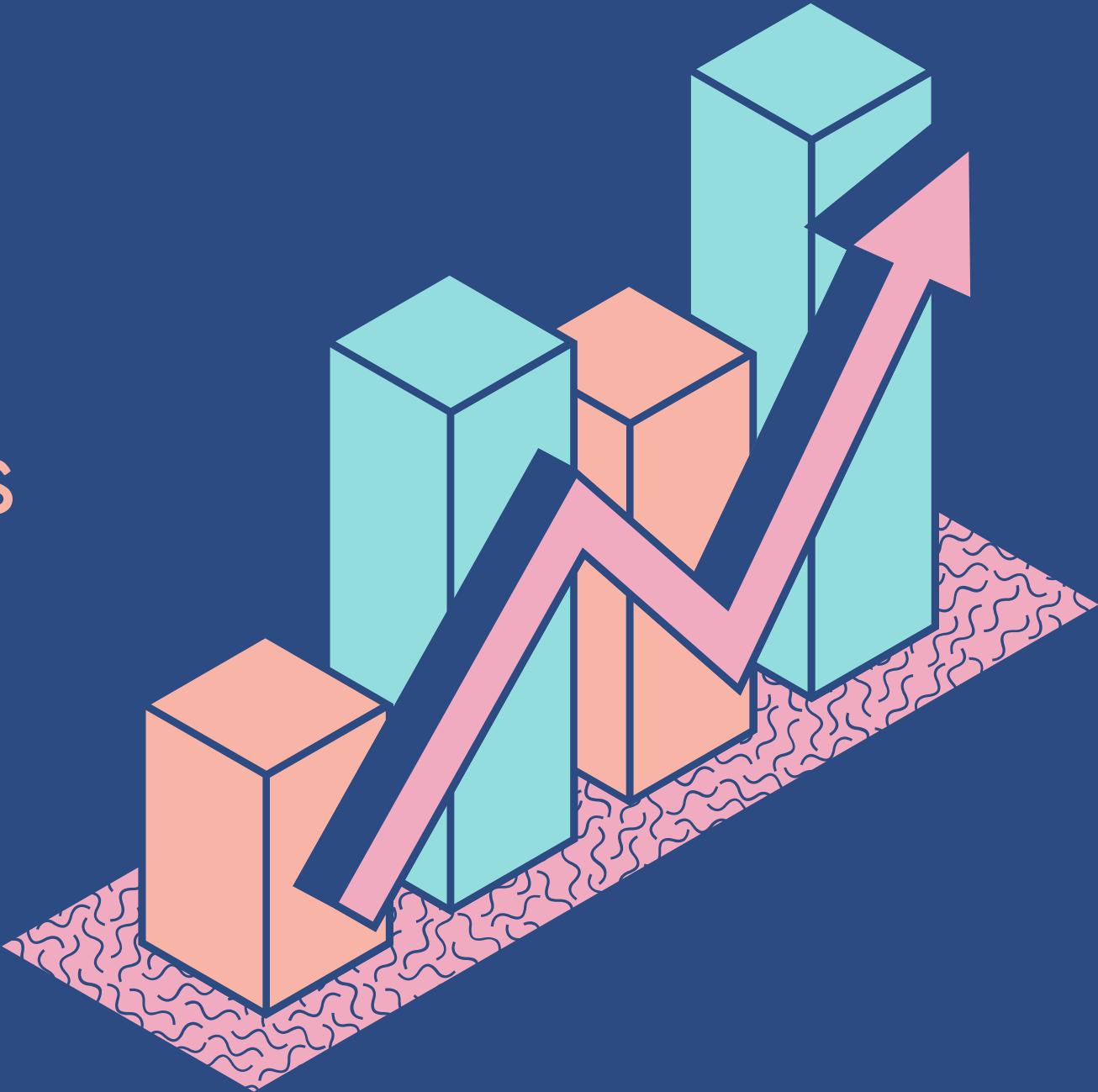


Additional Difficulties and Gaps

- Find English L1 person
- Stimuli and Contextual Understanding
- Pictures and Visual representation
- Ai voice program nuances
- Complexity of Irony
- Answers range

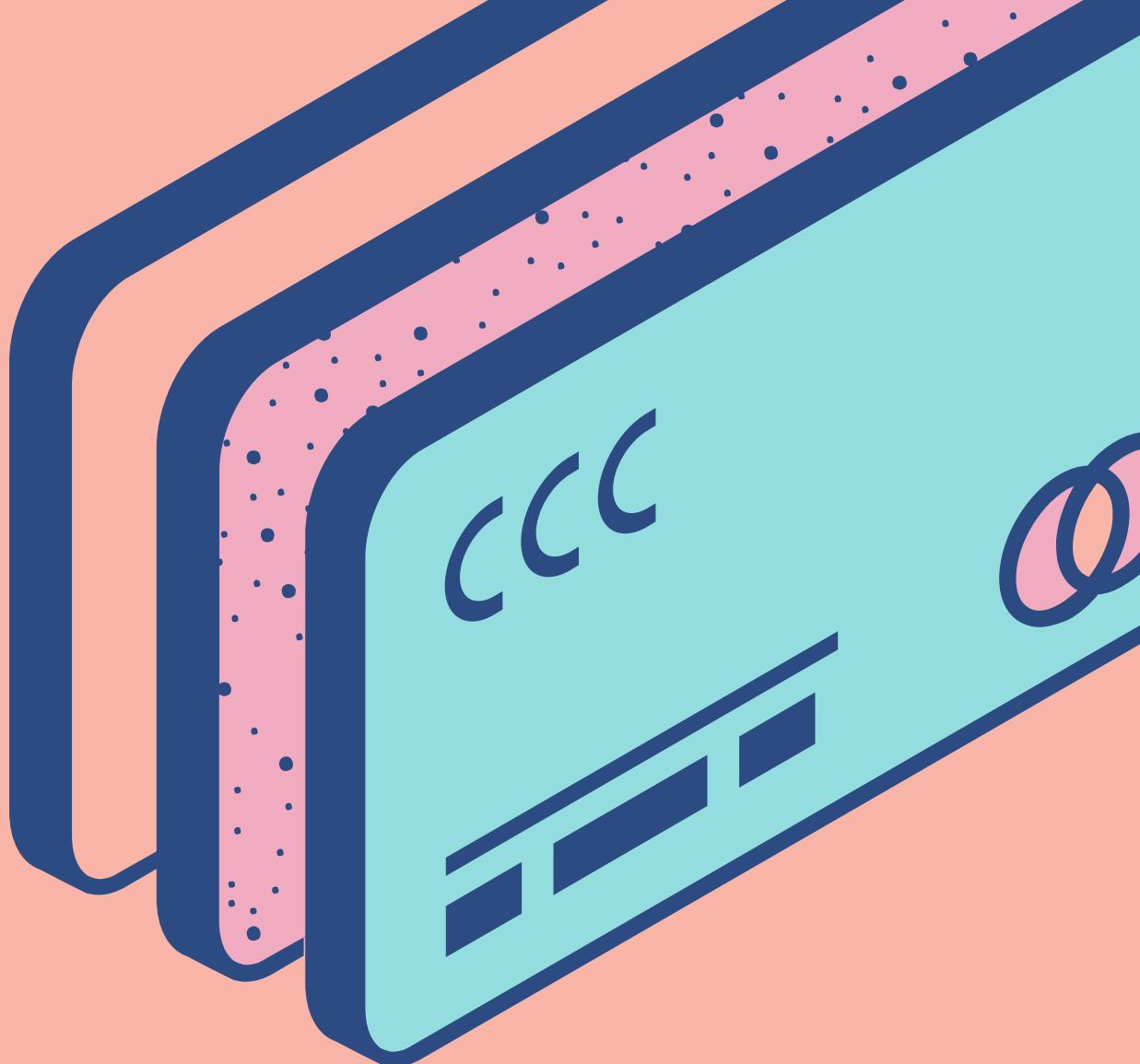
Research question

How much better do participants perceive irony in AI-voiced than in human-voiced utterances when the irony-containing utterances were created by a human?



Hypothesis

People perceive ironic statements differently when they are voiced by AI and by human.



Methods

METHOD

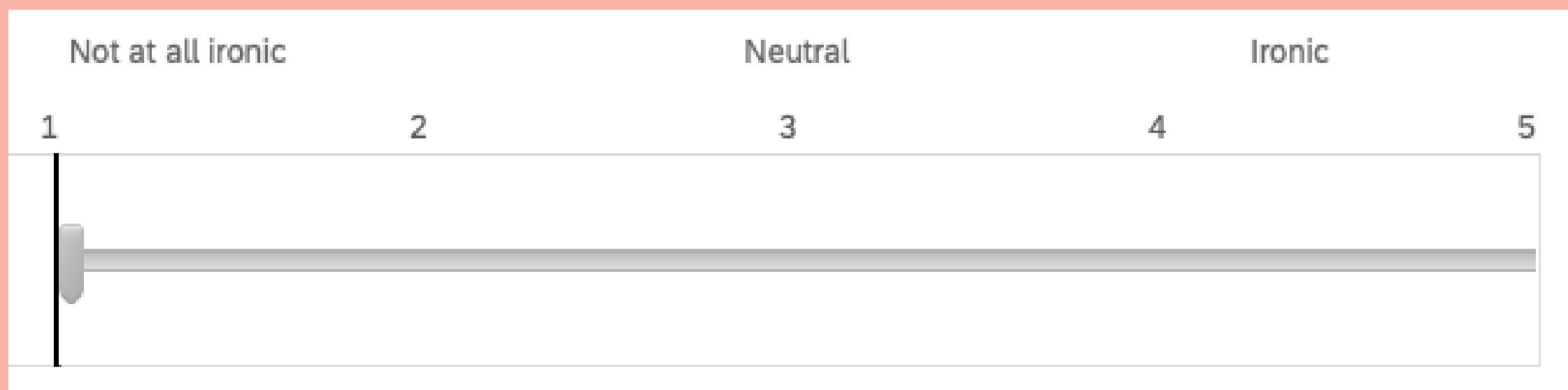
Survey with a 5-point
Slider scale and short
video clips

RECRUITMENT METHOD

Online recruitment via
telegram groups and
instagram stories

COLLECTED SAMPLE

52 online surveys
completed via Qualtrics



Experiment design

Participants received a link to the online (anonymous) survey on Qualtrics and 1 of 2 sets - 7 ironic statements and 7 non-ironic statements (video clips).

Participants were asked to rate ironic statements in video clips on a scale of 1 to 5, with 1 being the least ironic and 5 being the most ironic.

Experiment design

Dependent variable: Level of irony in statements



Experiment design

Dependent variable: Level of irony in statements

Independent variable: Type of voice (human or AI)



Prediction

Human-voiced utterances in irony-containing utterances which were created by a human will be perceived more ironic compared to statements made by the AI voice.

Materials

Videos with context pictures and background voices generated by AI or humans.

The person who voiced these statements was specifically chosen as a **native English speaker** to fully convey the irony presented in the video.

The AI's voice and tone were chosen to be **as similar as possible** to a person and his voice.

The utterances, in other words stimuli, and fillers were **created by humans** and are similar for both groups.

Sample stimuli



ELECTION 선거 2014

Vote by filling a square with (✓). 채우기 정사각형

KIM JONG-UN 김정은
 KIM JONG-UN 김정은


Filler: “You look like a bomb?”

Stimuli: “We got so many candidates”



Participants

Supposedly, out of 52 participants, we assume that prevailing majority were of the ages 18-23 as the survey was held among the peers (friends, groupmates).

The gender distribution would be approximately 1:1.

Native language would be also considered to be Kazakh and Russian as for significant majority with the level of English language being not lower than B1.

Data: human



4.6



3.3

Data: AI



3.9



3.1



search

Statistical analysis

```
? Coefficients:  
            Estimate Std. Error z value Pr(>|z|)  
is_AI     0.6332    0.2185   2.898  0.00375 **  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
[ ] exp(0.6332)
```

```
1.88362855674662
```

Ordinal regression for Likert scales
Package: Ordinal
[R notebook.ipynb - Colab - Google](#)

Verdict: Hypothesis and prediction were disproved

	Human	AI
1	3.4	4.2
2	3.4	3.9
3	4.6	3.9
4	3.5	3.1
5	4	3.5
6	4	3.8
7	3.8	3.9
8	3.3	3.8

3.75

3.76

Possible explanation

- Misunderstanding of participants
- Other factors that influence the irony



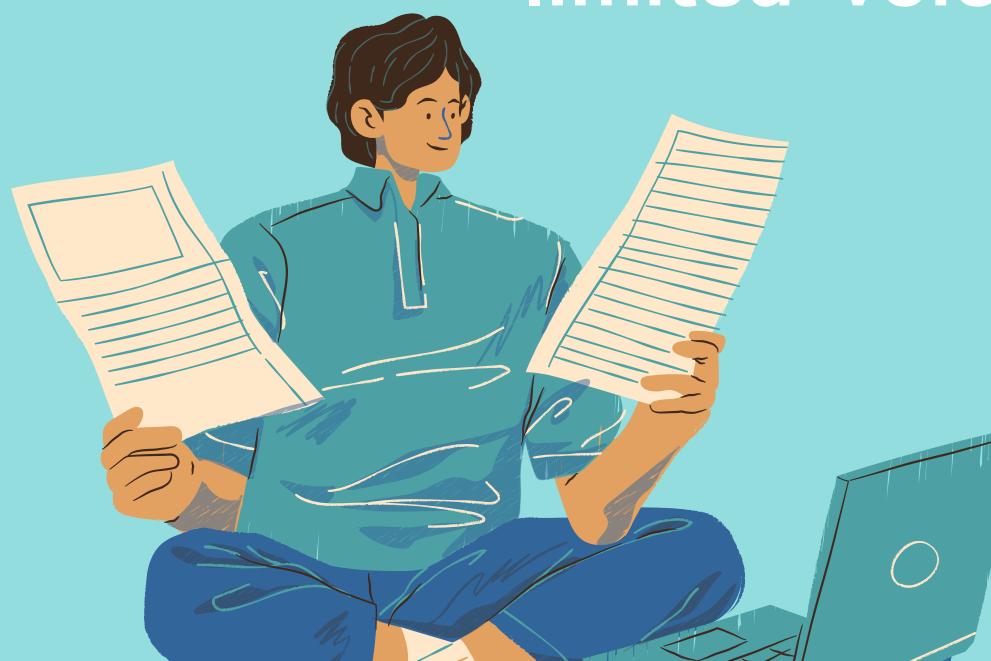
**“Should we evaluate the
ridiculousness of
statements?”**

Limitations

- small number of participants
- small group of participants



- limited voice acting for both humans and AI
- limited voice acting for both humans and AI



Future Research

- consider different voices: human and AI
- consider correlation between the voices
- increase the target population: language, age group
- new methods of assessment

