

Consent form

Research on the quality of Natural Language Generation systems

Dear participant,

Tilburg University investigates the quality of texts written by computers. We would like to know how good these produced texts are and what the influence is of different rating scales. Therefore, we ask you to read the generated texts and to judge them by answering a couple of questions. The answers you provide may be used for scientific publication. Naturally, if this is the case, your answers will be presented anonymously and it will be impossible to connect your answers to your identity. For your participation in this research, you will receive £4.00 (approximately 40 minutes of participation time). You will get 50 conversations along with generated responses. We ask you to rate these responses.

Thank you for your time and participation!

Kind regards,

Anouck Braggaar, Saar Hommes, Peter Blomsma, Emiel Krahmer, Emiel van Miltenburg, Chris van der Lee, Nadine Braun, Frédéric Tomas, Martijn Goudbeek
Tilburg University

Completing this survey will take approximately 40 minutes and is entirely voluntary. This means that you are allowed to stop participation at any moment without having to provide a reason and without any repercussions. No risks are involved in participating. Collection of all data happens in accordance with the new General Data Protection Regulation (GDPR) rules and the Ethical Committee of TSHD has given permission to conduct this research. Data will be processed and saved anonymously. The anonymous data can, when requested, be shared with other researchers.

If you agree to participate, you confirm that:

- you are over 16 years old;
- you agree that your anonymous data will be saved and stored for 10 years;
- you agree that your anonymous data can be used for future research or scientific publications;
- you agree that your anonymous data can be shared with other researchers;
- you understand that you are allowed to stop participation at any moment. Regardless of reason, without having to provide a reason, and without any negative consequences.

If you would like to contact one of the researchers regarding this study, you can send an e-mail to:

A.R.Y.Braggaar@tilburguniversity.edu For comments on—or complaints about—this research it is also possible to contact the Tilburg University School of Humanities and Digital Sciences Research Ethics and Data Management Committee via tshd.redc@tilburguniversity.edu.

Thank you for your time and participation!

Kind regards,

Anouck Braggaar, Saar Hommes, Peter Blomsma, Emiel Krahmer, Emiel van Miltenburg, Chris van der Lee, Nadine Braun, Frédéric Tomas, Martijn Goudbeek
Tilburg University

Do you agree with participating in this study?

Yes

No

Prolific ID

What is your Prolific ID?

Please note that this response should auto-fill with the correct ID

Introduction

Conversation Evaluation Task

Please read the following instructions:

You will be given a short snippet of conversation between two persons containing two sentences. You will be asked to evaluate a third utterance (**response**) on two metrics (readability and coherence).

The response may come from different automated systems. The overall goal of this experiment is to improve chat bots in their ability to converse with humans.

You will be judging the responses on two metrics:

Metric 1: Readability: Is the response easy to understand, fluent and grammatical and does not have any consecutive repeating words.

Metric 2: Coherence: Is the response relevant to the topic of conversation.

Instructions

You will rate the response utterance on metrics of readability and coherence.

Sample conversation 1

Person A: Today I learned that 4 US states have completely banned billboards.

Person B: Living in Los Angeles, I would vote for this so hard. I'm so sick of billboards.

Standard Response: **If this catches on, it would be safe to bet to say that LA would be the last place to do this if they ever even consider it.**

Readability Value of the Standard Response: 100

Coherence Value of the Standard Response: 100

We will now evaluate this response in relation to readability and coherence values of the standard response provided above.

Response: Well I am going going to the baseball game.

Readability Value: 50

Coherence Value: 25

Explanation for Readability: The response scores lower (50) on readability than the readability value (100) of standard response because this response is not fluent and contains repeated words.

Explanation for Coherence: The response scores lower on coherence (25) than the coherence value (100) of standard response as this response is not consistent with the topic that the speakers involved in the conversation are talking about (Los Angeles, billboards)

You will start your task from next page

Conv1

Here is the conversation:

Person A: **And at least one from a housekeeper.**

Person B: **I really want to see this Maury episode.**

Standard Response: **I would pay money to see someone edit together such a thing.**

Readability Value of the Standard Response: 68

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **It's a great show.**

Please enter readability

Please enter coherence

Response 2: **I don't get it.**

Please enter readability

Please enter coherence

Response 3: **I want to see this episode**

Please enter readability

Please enter coherence

Conv2

Here is the conversation:

Person A: **What is your most successful unethical life hack?**

Person B: **If you have siblings, blame them for everything**

Standard Response: **Unless they are more successful than you, then have a good relationship and take the blame for everything**

Readability Value of the Standard Response: 53

Coherence Value of the Standard Response: 98

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I don't know what you are talking about.**

Please enter readability

Please enter coherence

Response 2: **I don't know if you are being sarcastic or not.**

Please enter readability

Please enter coherence

Response 3: **what if you don't have kids?**

Please enter readability

Please enter coherence

Conv3

Here is the conversation:

Person A: **I am not convinced he 's coming back**

Person B: **not going to happen. that would be crazy if it did though . what would happen?**

Standard Response: **if the president refused to return? the GOP would definitely impeach him and accept pence.**

Readability Value of the Standard Response: 48

Coherence Value of the Standard Response: 94

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I am not sure if that is a thing or not.**

Please enter readability

Please enter coherence

Response 2: **I don't know, but I think he 's going to be a lot worse.**

Please enter readability

Please enter coherence

Response 3: **I'm sure he 'll be fine.**

Please enter readability

Please enter coherence

Conv4

Here is the conversation:

Person A: **what fact are you tired of explaining to people?**

Person B: **no, I am not cheating on my husband, and no, neither of us pressured the other into this. Healthy and consensual nonmonogamy is a thing.**

Standard Response: **sounds more like you are cheating on your husband**

Readability Value of the Standard Response: 96

Coherence Value of the Standard Response: 88

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I'm sorry to hear that.**

Please enter readability

Please enter coherence

Response 3: **I don't believe you.**

Please enter readability

Please enter coherence

Response 4: **I don't think you understand what you are talking about.**

Please enter readability

Please enter coherence

Conv5

Here is the conversation:

Person A: **What supporting character from a tv show would you definitely watch in their own spinoff?**

Person B: **the adventures of Creed Bratton**

Standard Response: **Eh. Creed worked because they made sure not to overuse him. I think his shock value would get pretty old in a show of nothing but him**

Readability Value of the Standard Response: 69

Coherence Value of the Standard Response: 96

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I don't get it.**

Please enter readability

Please enter coherence

Response 2: **I would love to see him as a villain.**

Please enter readability

Please enter coherence

Response 3: **Creed Bratton is the best actor of all time.**

Please enter readability

Please enter coherence

Conv6

Here is the conversation:

Person A: **His tweets are really entertaining tbh. scary and absurd, but entertaining.**

Person B: **I simply can not look away. it just keeps getting weirder.**

Standard Response: **bizarro America is a strange show.**

Readability Value of the Standard Response: 74

Coherence Value of the Standard Response: 92

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I'm not sure if you 're being sarcastic or not.**

Please enter readability

Please enter coherence

Response 2: **I don't get it.**

Please enter readability

Please enter coherence

Response 3: **I think it is a bit of a stretch.**

Please enter readability

Please enter coherence

Conv7

Here is the conversation:

Person A: **I am stealing this thank you very much.**

Person B: **I made this.**

Standard Response: **well then I'm still stealing it, thank you very much.**

Readability Value of the Standard Response: 86

Coherence Value of the Standard Response: 98

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I third this**

Please enter readability

Please enter coherence

Response 2: **I made this.**

Please enter readability

Please enter coherence

Response 3: **you made this?**

Please enter readability

Please enter coherence

Conv8

Here is the conversation:

Person A: **Reminds me of how there are more plastic flamingos than real ones**

Person B: **Plastic flamingos in the front yard is a sign that swingers live there. so they are more swingers than real flamingos?**

Standard Response: **I guess that explains Florida?**

Readability Value of the Standard Response: 49

Coherence Value of the Standard Response: 87

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **plastic flamingos?**

Please enter readability

Please enter coherence

Response 2: **no, it is just a plastic bag.**

Please enter readability

Please enter coherence

Response 3: **I have never heard of that. I have never heard of it.**

Please enter readability

Please enter coherence

Conv9

Here is the conversation:

Person A: **keep him as Erick and keep McAvoy as Charles but for the love of god give everyone else a do-over.**

Person B: **would love this, but if I had to pick another to differentiate from the fox, Nikolaj Coster - waldau seems like a great choice**

Standard Response: **isn't that Jamie from got? i could see him killing that roll**

Readability Value of the Standard Response: 101

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I would love to see him as a villain.**

Please enter readability

Please enter coherence

Response 2: **I would love to see him play the role of doom.**

Please enter readability

Please enter coherence

Response 3: **I am not sure what you are trying to say.**

Please enter readability

Please enter coherence

Conv10

Here is the conversation:

Person A: **Yeah he healed his wound with his spray thing, the same one he used to seal the door.**

Person B: **He would have been if doctor strange did not give Thanos the time stone**

Standard Response: **does this happen on Titan?**

Readability Value of the Standard Response: 66

Coherence Value of the Standard Response: 89

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **that's what I thought, but I don't think that's how it works.**

Please enter readability

Please enter coherence

Response 2: **He did not have the power to do that.**

Please enter readability

Please enter coherence

Response 3: **I think it is time to rewatch the gauntlet.**

Please enter readability

Please enter coherence

Conv11

Here is the conversation:

Person A: **The inevitable moment**

Person B: **Isn't that the guy that lost the British election against the idiot that initiated the Brexit referendum?**

Standard Response: **ed miliband**

Readability Value of the Standard Response: 120

Coherence Value of the Standard Response: 71

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **No, it was a joke**

Please enter readability

Please enter coherence

Response 2: **no, that 's the guy who voted for Brexit.**

Please enter readability

Please enter coherence

Response 3: **no, that was the referendum.**

Please enter readability

Please enter coherence

Conv12

Here is the conversation:

Person A: **what do they taste like?**

Person B: **A camel scotch fillet is marbled like good beef, except triple the size. the flavor is very much like beef - slightly stronger but very delicious.**

Standard Response: **The camel is now on my "eat" list. I still have not eaten a venison steak.**

Readability Value of the Standard Response: 81

Coherence Value of the Standard Response: 96

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I've never heard of this before. what's it about?**

Please enter readability

Please enter coherence

Response 2: **I've never heard of that. what's it taste like?**

Please enter readability

Please enter coherence

Response 3: **What is the difference?**

Please enter readability

Please enter coherence

Conv13

Here is the conversation:

Person A: **what will become illegal in the next 50 years?**

Person B: **driving a manual car on a public road.**

Standard Response: **do you mean manual as in "manual transmission" or manual as in "human operated"?**

Readability Value of the Standard Response: 7

Coherence Value of the Standard Response: 96

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I don't think that is illegal.**

Please enter readability

Please enter coherence

Response 2: **I don't get it.**

Please enter readability

Please enter coherence

Response 3: **what if it is already illegal?**

Please enter readability

Please enter coherence

Conv14

Here is the conversation:

Person A: **back when he was the "happy one"**

Person B: **mou has changed so much now. it 's unfortunate that he has.**

Standard Response: **he has not changed, he 's always ahd the same mood swings over the years.**

Readability Value of the Standard Response: 90

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I'm not sure what you 're trying to say.**

Please enter readability

Please enter coherence

Response 2: **yeah, he has been pretty good in the past few years.**

Please enter readability

Please enter coherence

Response 3: **I think he is the best player in the league.**

Please enter readability

Please enter coherence

Conv15

Here is the conversation:

Person A: **what two videogames would you like to see in a crossover game?**

Person B: **total war star wars**

Standard Response: **I think star wars and company of heroes or dawn of war would be a better fit. the cover system, unit groups, vehicles, etc . plus the delicious physics.**

Readability Value of the Standard Response: 44

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I'm not sure what you mean by that.**

Please enter readability

Please enter coherence

Response 2: **that would be awesome.**

Please enter readability

Please enter coherence

Response 3: **star wars?**

Please enter readability

Please enter coherence

Conv16

Here is the conversation:

Person A: **wherever you go, there you are.**

Person B: **what is this a quote from?**

Standard Response: **I hear it in aa. I don't know the original origin.**

Readability Value of the Standard Response: 76

Coherence Value of the Standard Response: 94

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **it is a reference to something like this : URL**

Please enter readability

Please enter coherence

Response 2: **It is a quote from arrested development.**

Please enter readability

Please enter coherence

Response 3: **it is a quote from the office.**

Please enter readability

Please enter coherence

Conv17

Here is the conversation:

Person A: **what are some real facts that sound fake at first?**

Person B: **roughly 70% of the world's oxygen is produced by marine algae.**

Standard Response: **26% of the worlds hot air is produced by politicians .**

Readability Value of the Standard Response: 78

Coherence Value of the Standard Response: 96

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **that is not true at all.**

Please enter readability

Please enter coherence

Response 2: **I'm pretty sure that is a myth.**

Please enter readability

Please enter coherence

Response 3: **I don't believe you.**

Please enter readability

Please enter coherence

Conv18

Here is the conversation:

Person A: **what is the most NSFW thing you have seen at an office gathering or party?**

Person B: **A coworker once stood on a swivel chair**

Standard Response: **officer buckle and Gloria would not be happy**

Readability Value of the Standard Response: 63

Coherence Value of the Standard Response: 89

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **that is a good one.**

Please enter readability

Please enter coherence

Response 2: **How did you find out?**

Please enter readability

Please enter coherence

Response 3: **What was on the chair?**

Please enter readability

Please enter coherence

Conv19

Here is the conversation:

Person A: **homemade my sister's specialty, carrot cake**

Person B: **recipe?**

Standard Response: **remind me ! 2 days**

Readability Value of the Standard Response: 75

Coherence Value of the Standard Response: 92

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **here is the recipe**

Please enter readability

Please enter coherence

Response 2: **here you go**

Please enter readability

Please enter coherence

Response 3: **I'll give it a try!**

Please enter readability

Please enter coherence

Conv20

Here is the conversation:

Person A: **I ran 3 miles today thinking that was 5 km I was so hyped because I did not walk at all then I look at my Nike run achievement and 5 km equal 3.1 miles**

Person B: **this just gave me a weird thought, is there a metric version of the term mileage ? or is that a universal term ? or am i just an idiot?**

Standard Response: **this is a good question**

Readability Value of the Standard Response: 100

Coherence Value of the Standard Response: 81

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **that is what I was going to say.**

Please enter readability

Please enter coherence

Response 2: **I think it is called "measurement"**

Please enter readability

Please enter coherence

Response 3: **it is a unit of measurement.**

Please enter readability

Please enter coherence

Conv21

Here is the conversation:

Person A: **literally the best team possible. I will never forget how good that show is.**

Person B: **I just watched it for the first time a few weeks ago. I wish I could re-experience it over again.**

Standard Response: **my husband and I found it on amazon prime. we binge watched it with the self-control of a weight watcher meeting next to a bakery.**

Readability Value of the Standard Response: 70

Coherence Value of the Standard Response: 97

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I have not watched it yet, but I have never seen it. I'll check it out.**

Please enter readability

Please enter coherence

Response 2: **It's a great show, but it's a great show.**

Please enter readability

Please enter coherence

Response 3: **I watched the first season. it was amazing.**

Please enter readability

Please enter coherence

Conv22

Here is the conversation:

Person A: **what kind of bear is Paddington?**

Person B: **spectacled bear. he does n't really look like one, but he 's from Peru, and that 's the only species of bear that lives there.**

Standard Response: **thank you very much for the response .**

Readability Value of the Standard Response: 90

Coherence Value of the Standard Response: 93

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I don't know what that is, but I'm not sure what you 're trying to say.**

Please enter readability

Please enter coherence

Response 2: **that is not a polar bear.**

Please enter readability

Please enter coherence

Response 3: **I thought he was a bear?**

Please enter readability

Please enter coherence

Conv23

Here is the conversation:

Person A: **are you living in some other timeline? and if so, can you smuggle me over there?**

Person B: **they 're talking about the popular vote . in terms of the popular vote, Hillary had about a million more votes than Donald Trump.**

Standard Response: **over 2 million more you mean?**

Readability Value of the Standard Response: 91

Coherence Value of the Standard Response: 91

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **that is not what I said.**

Please enter readability

Please enter coherence

Response 2: **the popular vote does not matter.**

Please enter readability

Please enter coherence

Response 3: **that is not how elections work.**

Please enter readability

Please enter coherence

Conv24

Here is the conversation:

Person A: **let it cool, it 'll be the only metal colored thing there**

Person B: **actually, it will probably be black and covered with soot. it won't really stand out that much among all the ash.**

Standard Response: **magnet. already posted above this.**

Readability Value of the Standard Response: 33

Coherence Value of the Standard Response: 90

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **that is what I was thinking, but I'm not sure what you mean by that.**

Please enter readability

Please enter coherence

Response 2: **that is what I was thinking, but I'm not sure what you mean by that.**

Please enter readability

Please enter coherence

Response 3: **yeah, but it 's not black and white.**

Please enter readability

Please enter coherence

Conv25

Here is the conversation:

Person A: **tabs or spaces?**

Person B: **the real question is tab - size.**

Standard Response: **is it common to have something else than 4 spaces ?**

Readability Value of the Standard Response: 78

Coherence Value of the Standard Response: 86

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I think you meant spaces.**

Please enter readability

Please enter coherence

Response 2: **I thought it was the same thing.**

Please enter readability

Please enter coherence

Response 3: **what is the difference between the two?**

Please enter readability

Please enter coherence

Conv26

Here is the conversation:

Person A: **but it worked for Turriss! Jesus, I miss that guy he was supposed to be my sweater this Christmas.**

Person B: **I miss him too but he was n't signing with the team so we had to trade him. Hoffman is under contract . does n't make sense to move him.**

Standard Response: **what does not make sense is letting Dorion be in charge of a rebuild**

Readability Value of the Standard Response: 83

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **he was the only one who did not get the joke.**

Please enter readability

Please enter coherence

Response 2: **I think it was a joke, but I don't think it's worth it.**

Please enter readability

Please enter coherence

Response 3: **yeah, but he is still on the bench.**

Please enter readability

Please enter coherence

Conv27

Here is the conversation:

Person A: **who are you and what have you done today?**

Person B: **I'm just a lowly security guard, sitting in a guard shack since 6 am. two more hours. two more hours.**

Standard Response: **that is a cool job, man . don't put yourself down. you are not lowly. that is an awesome job to have.**

Readability Value of the Standard Response: 66

Coherence Value of the Standard Response: 96

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **what kind of job?**

Please enter readability

Please enter coherence

Response 2: **good luck!**

Please enter readability

Please enter coherence

Response 3: **what is your job?**

Please enter readability

Please enter coherence

Conv28

Here is the conversation:

Person A: **he would be fine in the right setting, North Korea, Saudi Arabia, Alabama.**

Person B: **can we vote to send him somewhere else?**

Standard Response: **like Texas?**

Readability Value of the Standard Response: 78

Coherence Value of the Standard Response: 82

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **no, he will be president.**

Please enter readability

Please enter coherence

Response 2: **I don't think so.**

Please enter readability

Please enter coherence

Response 3: **he is already in office.**

Please enter readability

Please enter coherence

Conv29

Here is the conversation:

Person A: **can't run as fast on a bike.**

Person B: **the bike was two - tired.**

Standard Response: **finally, all of the bazooka joe comics, and 1980's joke books paid off! this was brilliant btw.**

Readability Value of the Standard Response: 70

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I thought it was funny.**

Please enter readability

Please enter coherence

Response 2: **that is not how it works.**

Please enter readability

Please enter coherence

Response 3: **that is what I was thinking.**

Please enter readability

Please enter coherence

Conv30

Here is the conversation:

Person A: **a dominos employee stood outside of recently closed papa john in my neighborhood and started selling some pizzas.**

Person B: **wait wtf , since when did the papa john's down in bay ridge close?**

Standard Response: **it's been like a week**

Readability Value of the Standard Response: 116

Coherence Value of the Standard Response: 88

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **it's been a while since I've seen it.**

Please enter readability

Please enter coherence

Response 2: **I don't know, I've never heard of papa johns.**

Please enter readability

Please enter coherence

Response 3: **I'm not sure what you 're trying to say.**

Please enter readability

Please enter coherence

Conv31

Here is the conversation:

Person A: **do it, bro. get stoned and swim with the sharks.**

Person B: **do you think they 'll know I'm high like cats do?**

Standard Response: **ehhh. they might have their suspicions.**

Readability Value of the Standard Response: 56

Coherence Value of the Standard Response: 91

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **don't worry, you'll be fine.**

Please enter readability

Please enter coherence

Response 2: **do you have cats?**

Please enter readability

Please enter coherence

Response 3: **if you don't like them, you'll be fine.**

Please enter readability

Please enter coherence

Conv32

Here is the conversation:

Person A: **then he got super hunky towards the end. basically like trying to believe that Channing Tatum could ever possibly play someone who can't get laid.**

Person B: **you got a picture? I can't imagine hunky Urkel.**

Standard Response: **google?**

Readability Value of the Standard Response: 37

Coherence Value of the Standard Response: 74

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I don't think I've ever seen him in a movie.**

Please enter readability

Please enter coherence

Response 2: **here you go**

Please enter readability

Please enter coherence

Response 3: **I don't think I've ever seen a picture of him before.**

Please enter readability

Please enter coherence

Conv33

Here is the conversation:

Person A: **still my favorite release this year**

Person B: **overkills > these. awesome silhouette though**

Standard Response: **honestly I don't agree, the pk looks like a brillo pad**

Readability Value of the Standard Response: 85

Coherence Value of the Standard Response: 92

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **thanks ! I'll have to check it out!**

Please enter readability

Please enter coherence

Response 2: **these are my favorite pair of all time.**

Please enter readability

Please enter coherence

Response 3: **I'm not sure if this is a typo or not.**

Please enter readability

Please enter coherence

Conv34

Here is the conversation:

Person A: **first time watching f1!**

Person B: **enjoy the show! unfortunately, Monaco can be a bit of a parade sometimes, I hope it gets interesting so you have a nice first race!**

Standard Response: **chucked \$5 on Ricciardo top 3 finish for more fun haha**

Readability Value of the Standard Response: 69

Coherence Value of the Standard Response: 91

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I have not seen it yet, but I'll check it out. thanks!**

Please enter readability

Please enter coherence

Response 2: **thanks! I'll check it out!**

Please enter readability

Please enter coherence

Response 3: **thanks!**

Please enter readability

Please enter coherence

Conv35

Here is the conversation:

Person A: **we should collect all of the trash in the world, turn it into a ball, and fling it all into space.**

Person B: **what if it comes back in 1000 years?**

Standard Response: **that is 10 generations away, who cares!**

Readability Value of the Standard Response: 56

Coherence Value of the Standard Response: 94

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **then we can build a wall.**

Please enter readability

Please enter coherence

Response 2: **then you would have to go to the moon.**

Please enter readability

Please enter coherence

Response 3: **what if the earth is flat?**

Please enter readability

Please enter coherence

Conv36

Here is the conversation:

Person A: **I could say the same to you man**

Person B: **you must be fun at parties**

Standard Response: **"just leave it" continues to reply**

Readability Value of the Standard Response: 48

Coherence Value of the Standard Response: 92

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I'm not your buddy, pal.**

Please enter readability

Please enter coherence

Response 2: **you must be fun at parties**

Please enter readability

Please enter coherence

Response 3: **I don't think you're fun at parties.**

Please enter readability

Please enter coherence

Conv37

Here is the conversation:

Person A: **what picture did you use to get the chassis itself? did you draw it or is there a load of blank formula 1 car photos I don't know about haha**

Person B: **I used a launch photo of the rs 17, removed the sponsors and color balanced the layers**

Standard Response: **thanks for the response haha I'm trying to learn how to make concepts like this so thank you!**

Readability Value of the Standard Response: 78

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I don't think I've ever heard of that. I'll have to check it out.**

Please enter readability

Please enter coherence

Response 2: **thanks!**

Please enter readability

Please enter coherence

Response 3: **did you use it?**

Please enter readability

Please enter coherence

Conv38

Here is the conversation:

Person A: **congratulations op! since you already have gone to community college what field are you in / aiming for?**

Person B: **hi there and thank u. I am in the stem field. studying for a degree in it and comp sci.**

Standard Response: **nice ! when things become tough to remember is always rooting for you!**

Readability Value of the Standard Response: 59

Coherence Value of the Standard Response: 94

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I'm sorry to hear that. I hope you have a great day!**

Please enter readability

Please enter coherence

Response 2: **what are you studying?**

Please enter readability

Please enter coherence

Response 3: **what field?**

Please enter readability

Please enter coherence

Conv39

Here is the conversation:

Person A: **DVD released?**

Person B: **not till Oct 17th digital release is tomorrow**

Standard Response: **what is this 'digital release'?**

Readability Value of the Standard Response: 15

Coherence Value of the Standard Response: 92

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **thanks! I'll check it out. thanks!**

Please enter readability

Please enter coherence

Response 2: **that is not how it works.**

Please enter readability

Please enter coherence

Response 3: **I thought it was released?**

Please enter readability

Please enter coherence

Conv40

Here is the conversation:

Person A: **today 25th march is Tolkien reading day**

Person B: **unknownst to me, I finished watching the return of the king today at 1 am.**

Standard Response: **URL wasn't aware you were watching it?**

Readability Value of the Standard Response: 89

Coherence Value of the Standard Response: 90

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I have not read it yet. I'll check it out. thanks!**

Please enter readability

Please enter coherence

Response 2: **same here. I loved it.**

Please enter readability

Please enter coherence

Response 3: **I've been meaning to watch it for a while, but I have not seen it yet.**

Please enter readability

Please enter coherence

Conv41

Here is the conversation:

Person A: **what do you think of bill nye 's new show on Netflix?**

Person B: **it's terrible. I only watched the first episode though, but it was terrible.**

Standard Response: **I tried to like it, but it was so cringe-worthy at parts like he was trying too hard to appeal to a young audience.**

Readability Value of the Standard Response: 64

Coherence Value of the Standard Response: 96

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I'm not sure if you 're being sarcastic or not.**

Please enter readability

Please enter coherence

Response 2: **I liked it better than the first one.**

Please enter readability

Please enter coherence

Response 3: **I thought it was supposed to be the first season.**

Please enter readability

Please enter coherence

Conv42

Here is the conversation:

Person A: **I really like the white shoe with a black sole, I think it 's maybe the Gel Lyte iii "yin yang"**

Person B: **that's a pretty dope colorway. I have a pair of the Gel Lyte iii 's myself, they 're incredibly comfortable and super versatile**

Standard Response: **awesome dude, what CW do you have?**

Readability Value of the Standard Response: 81

Coherence Value of the Standard Response: 90

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I think it's a bit of a stretch, but I don't think I've ever seen a pair of shoes in my life.**

Please enter readability

Please enter coherence

Response 2: **thanks! I'll look into it!**

Please enter readability

Please enter coherence

Response 3: **do you have a link?**

Please enter readability

Please enter coherence

Conv43

Here is the conversation:

Person A: **absolutely, he at first agreed to let Russia question former us ambassadors and others. URL:**

Person B: **he would be impeached soon after he handed over those troops.**

Standard Response: **your faith in republican congress may be misplaced. you did not see the Nunes story?**

Readability Value of the Standard Response: 65

Coherence Value of the Standard Response: 94

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **he would not be impeached.**

Please enter readability

Please enter coherence

Response 2: **that is not how impeachment works.**

Please enter readability

Please enter coherence

Response 3: **that is not what he said.**

Please enter readability

Please enter coherence

Conv44

Here is the conversation:

Person A: **humans built a space rocket with stuff we found in the dirt.**

Person B: **humans built almost everything with things we found in the dirt.**

Standard Response: **I made pasta today with stuff found in the water.**

Readability Value of the Standard Response: 78

Coherence Value of the Standard Response: 95

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I thought we were talking about the dirt.**

Please enter readability

Please enter coherence

Response 2: **we are all humans on this blessed day.**

Please enter readability

Please enter coherence

Response 3: **I'm not sure what you 're trying to say here.**

Please enter readability

Please enter coherence

Conv45

Here is the conversation:

Person A: **I don't remember that scene. was that in part 1/2? I just watched it.**

Person B: **it is the opening sequence of episode 3. long surreal section taking place somewhere between the black lodge and non-existence.**

Standard Response: **episode 3 is already out?**

Readability Value of the Standard Response: 83

Coherence Value of the Standard Response: 88

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **ah, I see. thanks for the heads up!**

Please enter readability

Please enter coherence

Response 2: **I thought it was the black mirror episode.**

Please enter readability

Please enter coherence

Response 3: **ah, I see. thanks.**

Please enter readability

Please enter coherence

Conv46

Here is the conversation:

Person A: **what 's your favorite band?**

Person B: **led zeppelin, Greenday, Coldplay, one republic**

Standard Response: **hello my new friend .**

Readability Value of the Standard Response: 75

Coherence Value of the Standard Response: 89

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I have never heard of this one. what os it about?**

Please enter readability

Please enter coherence

Response 2: **I have never heard of them. i will check them out.**

Please enter readability

Please enter coherence

Response 3: **led zeppelin?**

Please enter readability

Please enter coherence

Conv47

Here is the conversation:

Person A: **still, she should speak English because it is an English country.**

Person B: **wales is, by definition, the part of England & Wales that is not English.**

Standard Response: **Welsh people are certainly English.**

Readability Value of the Standard Response: 33

Coherence Value of the Standard Response: 89

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **no, it is not. it is a British thing.**

Please enter readability

Please enter coherence

Response 2: **Wales isn't a country.**

Please enter readability

Please enter coherence

Response 3: **it is not part of the English language, it is the UK.**

Please enter readability

Please enter coherence

Conv48

Here is the conversation:

Person A: **do these come with the white laces as a spare? I see everyone posting with the white lace swap**

Person B: **nope I bought the laces off of rope lace supply**

Standard Response: **ok dope thanks. I think the white boosted stripes would look really clean on this colorway too.**

Readability Value of the Standard Response: 79

Coherence Value of the Standard Response: 96

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **what laces are you talking about?**

Please enter readability

Please enter coherence

Response 2: **where did you get them?**

Please enter readability

Please enter coherence

Response 3: **thanks! I'll check it out! thanks!**

Please enter readability

Please enter coherence

Conv49

Here is the conversation:

Person A: **tons of stuff including Shawshank. the mist. die hard 2. tales from the crypt: demon knight. he 's a great character actor.**

Person B: **oh yeah, that guy. he looks so different as death.**

Standard Response: **he also plays the British father at the end of a bogus journey (says "my word!" as the wyld stallyns play).**

Readability Value of the Standard Response: 51

Coherence Value of the Standard Response: 97

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **I don't think I've ever seen that movie.**

Please enter readability

Please enter coherence

Response 2: **I think he was the only one who died.**

Please enter readability

Please enter coherence

Response 3: **he looks like he is in the movie.**

Please enter readability

Please enter coherence

Conv50

Here is the conversation:

Person A: **lpt: settle your scores irl and outside. parking lots are a good solution; you 'll get more oxygen and this tangibly improves the taste of victory.**

Person B: **from how far up should I drop my opponent?**

Standard Response: **as high up as you feel comfortable. first time droppers can still get noticeable results from as low as 6 inches**

Readability Value of the Standard Response: 67

Coherence Value of the Standard Response: 94

Your task is to evaluate **three responses provided below** on the metrics of readability and coherence with respect to the values of the **standard response** (100).

Note: A definition of the metrics can be seen if you hover over them in the questions

Response 1: **dependes on where you live.**

Please enter readability

Please enter coherence

Response 2: **depends on what you're doing.**

Please enter readability

Please enter coherence

Response 3: **I don't get it.**

Please enter readability

Please enter coherence

Demographics

Have you taken part in any previous studies apart from this on evaluating conversation?

No

Yes

Do you have any previous experience of taking part in studies that involve interacting with a chatbot?

No

Yes

Thank you!

Thank you for taking part in the study.

debriefing

This is the end of this study. Thanks again for participating! This study looked at the quality of texts that were automatically generated by a neural network. Neural networks need to be “trained” (i.e., learn from examples) in order to be able to produce texts and the training examples that we give them can have a big impact on the texts that these text-generating systems produce. In the experiment you participated in, you saw text generated by three systems and a reference utterance. You were asked to rate these responses. With this, we wanted to investigate if the manner of rating had any influence on the ratings. For example, we compared rating with a Likert scale to best-worst scaling (ranking results from best to worst).

Do you have any questions or comments? Please contact researcher Anouck Braggaar (a.r.y.braggaar@tilburguniversity.edu) For any comments or complaints about this research, you can also contact the Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences via tshd.redc@tilburguniversity.edu.

Click on proceed to be redirected to Prolific.

Powered by Qualtrics