

The Influence of Readers on Emotional Reactions to the Writing
Format of Natural Disaster News:
A Comparison of Human-Written and AI-Written Articles

YI CHEN CHEN

N11564628

Yung Hsin Lin

N11750804

Abstract

With the rapid progress of emerging technology, artificial intelligence (AI) seems to bring more and more influence on human beings' decision-making on information. With the knowledge that AI has been applied in various productive scenario commonly, this report aims to (i) improve that human beings' emotion may be affected by the editor of articles (ii) offering insightful recommendations for upcoming or present technical advancements. We first researched some earlier studies to measure the validity of our hypothesis, and we found six of academic resources could support our hypothesis. Then, we applied a sample dataset that was provided by the teaching team in both condition one and condition two of the experiment, and the analysis result came out with a discussion after we analysed the data carefully. We have analysed the simulated data using the functions of T-test, F-test, and summary statistic. In this step, we examine the simulated data we have been given as though it were participant data from individuals taking part in the study that we have designed. Based on our analysis, we concluded that it is not certain whether AI-generated and human-generated content in news could influence human emotions; however, the title regarding a natural disaster written by humans has shown its impact on positive reactions in readers. Finally, because of the population of artificial intelligence (AI), it is recommended to apply the experiment result data to the current and future development of technology, particularly in the AI area, as a reference to AI identification systems.

Introduction

The report topic investigates the impact of news author (human vs. AI) on readers' positive reaction in regarding news articles about natural disasters. In the context, we set emotion as our dependent variable, while the AI-generated news text is the independent variable. The four stimuli are: (1) title of news: AI-generated declarative news title compared with human-wrote exaggerated title; (2) paragraph one of content: formal language compared with direct language; (3) source of content: ChatGPT comparing with Kim Willsher (human author of the news) (4) paragraph two of content: reported speech written by AI compared with speech written by human. To ensure its authenticity, the news article used throughout the report is carefully selected from a reliable and historical British daily newspaper: The Guardian, founded in 1821.

Since humans are approximately emotional beings, emotions have a certain level of impact on the decisions people make, especially when faced with situations that cause them to feel strongly. A serious natural disaster, such as a volcanic eruption, could be one of the most significant examples. In this report, it is expected that readers' positive reactions will be affected by a volcanic eruption news, and will have varying levels of impact depending on whether the text of news is human-written or AI-written. Hämäläinen et al. (2023) investigate the complex connection between language and emotion. Emotional signs are a natural part of human communication and are expressed through context, tone, and language subtleties. In other words, it holds significant importance in various aspects of human life. However, nowadays, despite emerging technology being developed maturely over the world, it is said that AI tools on which people are highly dependent, such as ChatGPT, QuillBot, and Google Bard, are still lacking that kind of creative expression in human-kind rich emotion (Jakesch et al., 2019), when generating word context, including positive and negative aspects. At the same time, Köbis & Mossink (2021) mentioned that technology's failure to achieve creative expression refers to humans' deep feelings and can be a combination of real-world experience, emotion, and inspiration (Haase & Hanel, 2023). As a result, it's probable that the tools' lack of human experiences makes it difficult for them to represent or duplicate such emotions. Additionally, regarding further potential causes, a technical description has been provided. Although human emotions acknowledged significance, it remains largely overlooked by many engineers and computer scientists, as highlighted in "An Overview of Emotion in Artificial Intelligence" (2022). Moreover, as Kirkpatrick (2023) points out, neural-network-driven systems tend to lack a comprehensive understanding of the real world. If not properly controlled or given the right settings, these systems could produce strange or

incorrect text outputs. Additionally, this neglect may further bring challenges for AI-generated texts, which often struggle to effectively connect emotions due to the lack of emphasis on emotional understanding and expression in AI development. This means, in terms of technology, the challenges in properly expressing emotions and context in texts generated by AI are highlighted by the lack of focus paid to emotional-related technology development.

Our report develops how the contributor of news articles about natural disasters influences positive emotional reactions. Specifically, we compare the articles written by humans with those generated by AI. We hypothesize that readers will demonstrate a lower frequency of positive reactions when exposed to human-written news articles compared to AI-generated content. This hypothesis is grounded in the belief that human authors infuse their writing with emotional depth and empathy, thereby causing a strong emotional relation and less positive emotion among readers for disastrous events. Conversely, we expect that AI-generated articles may lack the rich emotional expression existing in human writing, leading to comparatively fewer reactions from readers. By exploring four stimuli of news' title, two paragraphs, and source, our experiment aims to analysis its methodology with result so that other researchers can replicate and refer to them.

Methodology

Participants

Forty participants were get involved in each condition, roughly evenly divided between men and women. Participants are diverse in age bracket, from 18-25 years old to 65+ years old, which are slightly different numbers between them. The largest number of people is between 18-25 years old, and the smallest number of people is over 65 years old. The age bracket and gender details of participants are shown in the bar charts below.

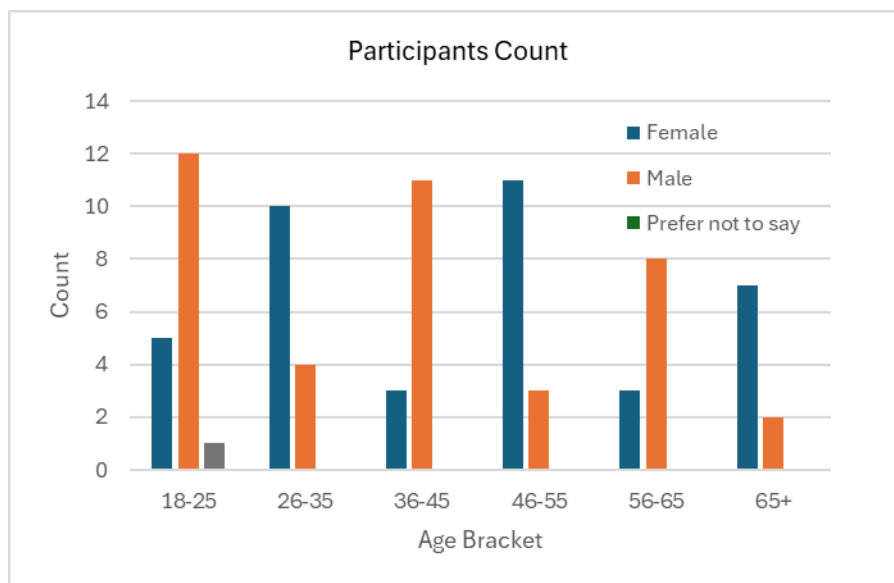


Diagram 1- Participants Count

Materials

The experiment includes four stimuli, which extract different parts from the same natural disaster news. Each stimuli shows two text sections which transfer the same information but presents in human-generated condition and AI-generated condition separately. To make the text more vivid and precise, we use news written by professional journalists as our human-generated text. We found a natural disaster news from The Guardian news website, which describes a volcanic eruption in Grindavík. We selected different parts of texts from this news as different stimuli in human generated text condition. Then, we use Chatgpt to make an AI version of these four text sections. The following is a detailed introduction to each stimulus :

- Stimulus 1 –News Title

Stimuli 1 - Title

Volcanic Eruption
Casts Doubt on the
Future of Grindavík
Town

**AI-
generated**

‘This town might be
over’:
Grindavík residents
face uncertain future
after volcano erupts
again

**Human-
generated**

- Stimulus 2 – Paragraph of News

Stimuli 2 - Paragraph

While there have been no
verified fatalities linked
to the eruptions, concerns
are growing as one
individual is unaccounted
for, presumed to have
plummeted into a
crevasse.

**AI-
generated**

The eruptions have
caused no confirmed
deaths but a man is
missing after
reportedly falling into
a fissure.

**Human-
generated**

- Stimulus 3 – Source

Stimuli 3 - Source

While there have been no
verified fatalities linked
to the eruptions, concerns
are growing as one
individual is unaccounted
for, presumed to have
plummeted into a
crevasse.

Resource: ChatGPT

**AI-
generated**

The eruptions have
caused no confirmed
deaths but a man is
missing after
reportedly falling into
a fissure.

*Resource: Lenore Taylor
Editor, Guardian Australia*

**Human-
generated**

The texts present the same paragraphs as Stimuli 2 but are added resource information below the text. For doing this, we can observe if the surrounding information influences the experiment results.

- Stimulus 4 – Another Paragraph of the News

Stimuli 4 –Paragraph2

"The future is uncertain at this point," remarked Bustion. "This event has cast doubt over the fate of our community. Several homes have been destroyed by fire, and with the water and power infrastructure compromised, we're left without any heat amidst the freezing temperatures."

**AI-
generated**

We don't know what will happen now. It has thrown a question mark over people's lives," Bustion said. "A few houses have burned but the water pipes and electricity have been damaged so there is no heating and it's been really cold."

**Human-
generated**

Based on the dependent variable, which is a positive reaction (Emotion), a same 5-point Likert scale question is applied in each stimulus: I feel less positive when I saw this news. The 5-point Likert scale ranges from the lowest (1) to the highest (5), which are Strongly disagree, Disagree, Neither agree or disagree, Agree, and Strongly agree separately.

Experimental procedure

Participants are randomly divided into two groups, which are forty people for either of two conditions. Participants will take the test without knowing the experiment condition, which means they will not know whether the text they see is AI-generated or not. Four different stimuli are shown to participants, they can choose different agree levels with the question statement based on the stimuli they receive.

Results

The analysis technique includes Summary Statistic, F-Test, and independent samples t-Test including equal variances and unequal variances. In F-Test, if $p > .05$, there is no significant difference between the variances of the two conditions, the results pertain to equal variances t-Test; on the contrary, if $p \leq .05$, the results pertain to unequal variances *t-Test*.

F-Test results exhibited $p > .05$ across all stimuli, indicating equal variances. For Stimulus 1, 40 participants in Condition 1 ($M = 3.65$) significantly rated higher than the other 40 participants in Condition 2 ($M = 3.00$), $t(78) = 3.28$, $p < .001$. Stimuli 2, 3, and 4 didn't reveal significant differences between conditions. 40 Condition 1 participants rated slightly higher for Stimulus 2 ($M = 3.075$) than the other 40 Condition 2 participants ($M = 2.825$), $t(78) = 1.13$, $p = 0.13$. 40 participants in each condition in Stimulus 3, assessed similar average scores (Condition 1: $M = 3.275$, Condition 2: $M = 3.15$), $t(78) = 0.75$, $p = 0.23$. As for Stimulus 4, there is also no obvious difference, however, 40 Condition 1 participants ($M = 3.35$) rated higher than Condition 2 ($M = 2.9$), $t(78) = 2.18$, $p = 0.016$.

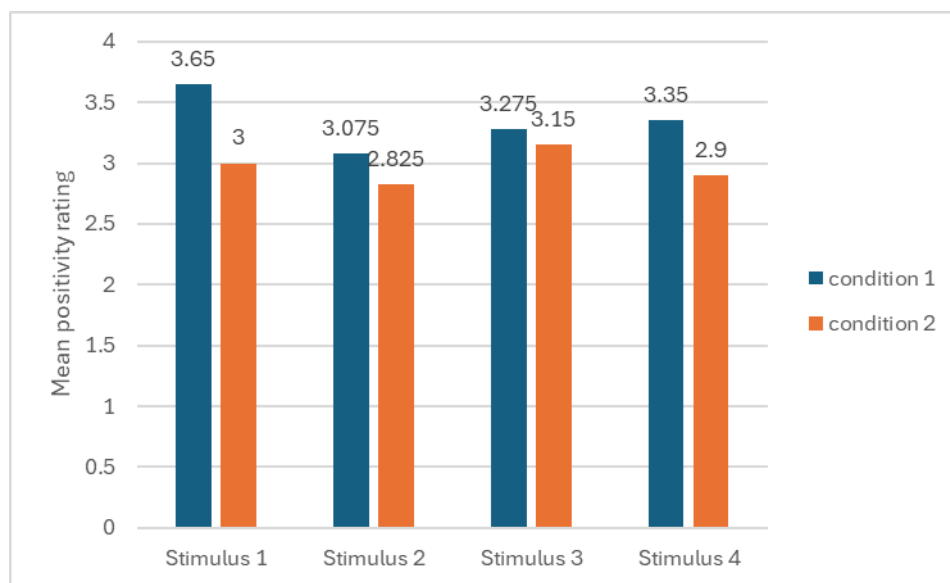


Diagram 2- Mean Positive Rating of each stimulus

Discussion

Among all the Stimuli, only Stimulus 1 shows a significant difference between condition 1 and condition 2, with condition 1 showing notably higher average scores compared to condition 2. It indicates that participants felt more positive when they saw the human-generated news title, comparing to AI-generated news title. This outcome aligns with our hypothesis that decision of positive reaction is lower for the information written by human.

However, for Stimulus 2, Stimulus 3 and Stimulus 4, there is no significant difference between two conditions, which suggests the impact of AI-generated versus human-generated news content on emotions remains uncertain. It represents that the results disconfirm the hypothesis. Additionally, our findings indicate that the addition of resource information, regardless of whether it is humans or AI, does not definitely cause effects in audience responses. The different results in emotional response between Stimulus 1 and the other stimuli may be attributed that the professional journalists know how to compile compelling news titles that enhance click-through rates so that audiences are easily to have a stronger emotional feeling to the headlines.

The experimental results are related to the literature in the introduction on the impact of emotions and AI techniques, particularly in emotionally charged situations such as natural disasters. Hämmäläinen et al (2023) indicate the complex connection between language and emotion in human communication. The emotional reaction can be enhanced by the subtleties in language, and context uniquely written by to human authors. This could be attributed that human writers are emotionally driven; they know how to better use emotional words in communication that can affect other people.

Jakesch et al. (2019) emphasize that AI tools are still difficult to produce the emotional expression despite the advance of technology. This limitation is comported with the research of Köbis & Mossink (2021) and further study by Haase & Hanel (2023), who underscore that AI lacks real-world experience and emotion in communication. Therefore, these literatures can be reflected in our results that human-written title transfers stronger expression for natural disaster topic.

Due to the lack of deep understanding of human emotion, AI-generated texts often fail to connect on an emotional level (Kirkpatrick, 2023). This disconnect is primarily due to the less attention to emotional aspects during the process of AI development, as noted in "An Overview of Emotion in Artificial Intelligence" (2022). The studies explain the results in our finding that AI-generated text fails to give an emotional title as humans do.

The limitations of my experimental study can be attributed to many factors. Firstly, the writing style of this news author is neutral, which does not inflame strong emotions. Therefore, except for the news title, the article content does not include many stirring words which may influence people positive reaction. Secondly, feelings can be varied on participants. Individuals can have various reactions toward a same thing due to their difference of backgrounds and personalities. Some people may put a stronger empathy on the victims of natural disasters, but others may not. Furthermore, the age and gender of participants may also influence the results. For example, people may become more sympathetic to natural disasters as they age, and women are generally more empathetic than

men. In addition, this is related to the geographical locations where the participants live. For example, in areas without volcanoes, they are relatively careless of natural disasters such as volcanic eruptions.

The experimentation of the study has shown its significant help ability for future use in technology development. After our analysis, it may help the technology's growth in two ways, including as a standard for search engines' sensitivity reference, building a more reliable and accurate search over the internet, and improving the filter function of the media rating system.

First of all, whether oriented towards mobile or computer devices, it is necessary to provide the users with a satisfying enough experience in web surfing, especially the exact things they expect to answer their desire after pressing "return" on the keyboard. Technically, such as by developing and employing search algorithms capable of understanding user context and language tones, these algorithms should swiftly analyse the query and return results that closely match the user's desired information.

Secondly, as streaming service technology becomes more common in modern society, many family members, from younger to older, are subscribing to this kind of service to spice up their daily lives. Despite the fact that it could effectively narrow the gap between families, the inaccurate film rating system implemented in the streaming application may also cause some issues regarding child abuse. Younger viewers are more susceptible to the impact of certain sensitive content, including violence, sexual content, or substance abuse. Thus, a reliable rating system helps protect these viewers by providing age-appropriate content.

The experimentation can help address these two weaknesses in IT applications by combining the data of our experimentation results and identifying whether the objects are suitable for showing on the screen based on settled standards and data thresholds set by AI engineers.

Reference

- An overview of emotion in artificial intelligence*. (2022, December 1). IEEE Journals & Magazine | IEEE Xplore. <https://ieeexplore.ieee.org/document/9736644>
- Haase, J., & Hanel, P. H. P. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity*, 33(3), 100066. <https://doi.org/10.1016/j.yjoc.2023.100066>
- Hämäläinen, P., Tavast, M., & Kunnari, A. (2023). Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study. *Association for Computing Machinery (ACM)*. <https://doi.org/10.1145/3544548.3580688>
- Jakesch, M., French, M., Ma, X., Hancock, J. T., & Хаамаһ, M. (2019). AI-Mediated Communication. *Association for Computing Machinery (ACM)*. <https://doi.org/10.1145/3290605.3300469>
- Kim Willsher (2024, January 16). ‘This town might be over’: Grindavík residents face uncertain future after volcano erupts again. The Guardian. <https://www.theguardian.com/world/2024/jan/15/grindavik-residents-face-uncertain-future-after-volcano-erupts-again-iceland>
- Kirkpatrick, K. (2023). Can AI demonstrate creativity? *Communications of the ACM*, 66(2), 21–23. <https://doi.org/10.1145/3575665>
- Köbis, N., & Mossink, L. (2021b). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114, 106553. <https://doi.org/10.1016/j.chb.2020.106553>

Statement of Contribution

	YICHEN CHEN	Yung Hsin Lin
Wrote Introduction?	Yes 100(%)	No
Reviewed Introduction?	No	Yes
Wrote Methodology?	Yes 25 (%)	Yes 75 (%)
Reviewed Methodology?	Yes	Yes
Wrote Results?	No	Yes 100(%)
Reviewed Results?	Yes	No
Wrote Discussion?	Yes 50(%)	Yes 50(%)
Reviewed Discussion?	Yes	Yes
Signature	YICHEN CHEN	Yung Hsin Lin