A Chatbot to Identify Sentence Ambiguity
(Technical Project)

The Premature Implementation of Emotional AI in Health and Safety Features
(STS Project)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Diksha Jothi

December 9, 2022

On my honor as a University student, I have neither given nor received unauthorized aid
on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

MC Forelle, Department of Engineering and Society

Briana Morrison, Department of Computer Science
Roseanne Vrugtman, Department of Computer Science

Introduction

The progression of making Artificial Intelligence (AI) more human-like in terms of understanding and recreating human behavior has led to the increasing usage of technologies like Affective Computing and Natural Language Processing (NLP). Also referred to as Emotional AI, Affective Computing tracks things like eye movements, vocal inflections, and facial expressions to analyze the subtleties to determine what emotion a person is feeling (Somers, 2019). It is currently used as a means to get feedback on advertisements and customer service as well as health and safety features (Teufl et al., 2010; Dilmegani, 2021). On the other hand, Natural Language Processing is a field that analyzes and replicates human language in particular by parsing written sentences (Teufl et al., 2010). NLP is more commonly used with applications such as chatbots, the autocomplete feature in search engines, and virtual assistants on smartphones (Harkiran, 2020).

Being able to detect human emotion and understanding the language has a huge array of applications, but there are a lot of imperfections that can affect how useful this technology can be. For example, problems in facial recognition have been well documented, and the subtle differences in facial expressions that vary between different cultures show that even basic emotions can be perceived differently (Barrett et al., 2019). In the same way, a written sentence can have different semantics based on the context and the presence of sarcasm or slang that NLP has a hard time distinguishing (Saba, 2021). These inaccuracies are far from being eliminated, yet it is still being pushed to be utilized in many newer applications.

Because of how common the issues of inaccurate interpretations of emotion and language are, the conclusions that these technologies draw can be unreliable, especially when implemented into health and safety features. As more people become reliant on these technologies and the

more it becomes integrated into normal lives, the more consequences there will be for errors that occur in Affective Computing and NLP techniques. I will investigate how prevalent errors can occur in NLP tools in my independent technical project by designing a chatbot that alerts the user of how ambiguous their written input into the machine is. Subsequently, in my STS project, I will explore the occurrence of premature implementation of Emotion AI into technologies involving safety through a literature review.

Technical Project: Chatbot with Built-in Ambiguity Detection

As previously mentioned, Natural Language Processing has already become well implemented into some aspects of our daily lives. Grammar checking programs are commonly utilized among students and can even do things like provide suggestions for better wording or stronger sentence structure (Harkiran, 2020). Chatbots are also commonly used as online customer service to answer frequently asked questions or to redirect the customer elsewhere for better support (Harkiran, 2020). The current state of NLP is that it can perform some tasks really well, namely, those that can be completed by following a fixed set of rules (Levine, 2022). Because grammar rules remain mostly unchanged, grammar tools are reliable for formal uses like academic writing, but they fall short when it comes to analyzing the latest slang and detecting sarcasm.

Based on these examples, it is clear that there are many limitations to NLP, the main one being that the AI lacks enough context to discern ambiguous statements. What is considered ambiguous to a machine is completely different from what humans would consider ambiguous. For example, the term "I saw a person on the beach with my binoculars" could be interpreted as seeing someone through the binoculars or as seeing someone that is holding the binoculars

(Roldós, 2020). With the proper context, it is easier for humans to understand which is the correct meaning of the sentence, but AI requires a lot more context and background information to accurately understand the same thing. The problem that ambiguity causes in NLP is that advances in language modeling have a tendency of generating false statements (Lin et al., 2022). This is especially a problem when the lack of rigorous testing is coupled with uses of NLP that have high standards of accuracy, like in the medical field (Lin et al., 2022). This goes to show that the problem of ambiguity causing inaccuracies in NLP needs to be addressed in order to make more impactful advancements in that field.

My technical project will be to program a conversational chatbot that recognizes the level of ambiguity that a user-inputted statement has. To accomplish this, I will use Python as the coding language and apply the Transformer Language Model because it is a current popular tool that is good at dynamically understanding the undertones of dialogue (Mondal, 2021). What will make this different from a standard chatbot is that based on the written input provided by the user, it will analyze the number of different interpretations of the inputted sentence. The output will be the chatbot's natural response to the user's input, as well as the number of other likely interpretations and the option to retry the same input in a different way or with more context. In order to roughly identify the number of different interpretations, I will additionally use the Part of Speech Tagging method to identify the different meanings that individual words have by assigning each word a grammatical label (Gorrieri, 2021). I predict that the challenging part of this technical project will be smoothly incorporating Part of Speech Tagging to the rest of the chatbot to be able to identify the different interpretations of a sentence accurately.

Premature Implementation of Emotional AI

4

Artificial Intelligence models are constantly being improved upon and make progress towards the goal of recognizing and replicating human behavior. However, Affective Computing has only recently been introduced into products and business models. For example, some companies are using customer reactions to advertisements to determine how likely they are to purchase a product and others use it to track their employees' moods to encourage productivity (Somers, 2019). However, there are a lot of current technologies that are being developed with Affective Computing for the purpose of promoting health and safety. One current endeavor is the use of Emotional AI to monitor the mood attentiveness of a driver to prevent them from swerving or unintentionally speeding up (Somers, 2019). There have also been implementations such as a wearable device that releases specific scents based on the physiological state of the user to improve their mental health (Amores).

The problem behind Emotional AI is that there are a lot of imperfections in identifying a person's true feelings by using a general standard. For example, facial recognition is unable to understand differences in facial expressions amongst different cultures because it varies so much by region (Purdy et al., 2019). In addition, vocal inflections are also completely distinct for different cultures and languages, and the lack of data on lesser-known languages such as one of the 2,000 languages in Africa means that the chances of accurately recognizing the correct emotion are even lower (Sciforce, 2022).

Given the prevalence of inaccuracies, the current shift of applying Affective Computing to health and safety technologies is an important focus because it is one that is less researched compared to fields like advertising. Minimizing errors that occur is also crucial in health and safety because of how drastically it can affect a person's life (Naik et al., 2022). There is also a severe lack of regulation that is a common inadequacy in emerging technologies, but the

ramifications of errors in terms of health and safety call for more precautions to be taken before being implemented (Naik et al., 2022). Essentially, this problem is important to solve in order to prevent unintentional errors as a result of AI technologies being released too early and to reduce the risk of unaccounted factors causing inaccuracies.

In order to have a better understanding of this problem, I will analyze the connections between Emotional AI technology getting released too early and the users of those products by using the STS framework of Actor Network Theory. Created by Bruno Latour, Michel Callon, and John Law, this framework establishes that everything is both an actor and a network at the same time, and that everything is built around associations (Cressman, 2009). Actor Network Theory is really effective in analyzing the introduction of new technology and how connections may change in the future, which is why it will be useful to establish a network surrounding the developers of these new Emotional AI technologies. In addition, seeing the associations with the other actors such as the product testers, the companies wanting to implement the product, the patients/users of the product, the product itself, etc. will provide insight as to what causes the problem of technology being released too soon.

Research Methodology

In order to explore the established problem, my research will center around the following questions:

- Which implementations of Emotional AI in safety features have the least amount of risk?
- What is the cause of premature implementation in Emotional AI technology?

To answer the first question, I will conduct a literature review of existing studies of Emotional AI to develop an understanding of how accurate its subdivisions are. In other words, I will

compare the individual accuracies of Emotional AI tracking techniques such as vocal inflections, facial expressions, and body language and apply them to see which safety implementations are projected to result in greater inaccuracies. To answer the second question, I will use the framework of Actor Network Theory to create a network to see how the premature implementation of new technologies happens in general health and safety products. I will then use this to translate those discoveries into the context of Emotional AI because of the limited number of studies focusing on Emotional AI.

Anticipated Outcome and Impact

The problem of ambiguity and the lack of proper testing has led to errors in Affective Computing and Natural Language Processing, but with the completion of my technical project, I anticipate that my finished product will show how much more ambiguous commands and questions are than we are inclined to think. This will impact the field of NLP by showing what types of statements lead to more ambiguity and subsequently more inaccuracies. Newer technologies will be able to use this to minimize the ambiguity present in current NLP models. Furthermore, I expect that the conclusion of my STS research will present that a certain type of Emotional AI tracking technique is more accurate than other kinds. I also predict that it will reveal that certain actors have more power over the premature implementation in those technologies. The impact of the STS portion will show which areas of Emotional AI technologies need more testing and consideration before being implemented into society. It will also point out what changes to the connections between major actors need to be made in order to minimize the number of Emotional AI products from being released prematurely and will subsequently minimize the number of inaccurate conclusions from affecting people negatively.

References

Amores, J. (n.d.). *Project BioEssence*. MIT Media Lab. Retrieved December 6, 2022, from
https://www.media.mit.edu/projects/bioessence/overview/

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional
expressions reconsidered: Challenges to inferring emotion from human facial movements.
Psychological Science in the Public Interest, *20*(1), 1–68.
https://doi.org/10.1177/1529100619832930

Cressman, D. (2009). A Brief Overview of Actor-Network Theory: Punctualization,
Heterogeneous Engineering & Translation.

Dilmegani, C. (2021, October 5). *Top 10 emotional AI examples/use cases & reasons for
success*. AIMultiple. Retrieved December 6, 2022, from
https://research.aimultiple.com/emotional-ai-examples/

Gorrieri, L. (2021, May 10). *A simple NLP application for ambiguity resolution*. Medium.
Retrieved December 6, 2022, from https://towardsdatascience.com/a-simple-nlp-
application-for-ambiguity-resolution-bb698d19aff7

Harkiran. (2020, December 9). *Top 7 applications of NLP (Natural Language Processing)*.
GeeksforGeeks. Retrieved December 6, 2022, from https://www.geeksforgeeks.org/top-7-
applications-of-natural-language-processing/

Levine, S. (2022, March 1). *It's the Golden Age of Natural Language Processing, so why can't
Chatbots solve more problems?* Medium. Retrieved December 6, 2022, from
https://towardsdatascience.com/its-the-golden-age-of-natural-language-processing-so-why-
can-t-chatbots-solve-more-problems-bada37a427cf

Lin, S., Hilton, J., & Evans, O. (2022). TruthfulQA: Measuring how models mimic human
falsehoods. *Proceedings of the 60th Annual Meeting of the Association for Computational
Linguistics (Volume 1: Long Papers)*. https://doi.org/10.18653/v1/2022.acl-long.229

Mondal, A. (2021, November 9). *AI chatbot: Complete Guide to Build your AI chatbot with NLP
in python*. Analytics Vidhya. Retrieved December 6, 2022, from
https://www.analyticsvidhya.com/blog/2021/10/complete-guide-to-build-your-ai-chatbot-
with-nlp-in-python/

Naik, N., Hameed, B. M., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim,
S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal
and ethical consideration in artificial intelligence in Healthcare: Who takes responsibility?
Frontiers in Surgery, *9*. https://doi.org/10.3389/fsurg.2022.862322

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the Sociology of Technology might benefit each other. *Social Studies of Science*, *14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Purdy, M., Zealley, J., & Maseli, O. (2019, November 18). *The risks of using AI to interpret human emotions*. Harvard Business Review. Retrieved December 6, 2022, from https://hbr.org/2019/11/the-risks-of-using-ai-to-interpret-human-emotions

Roldós, I. (2020, December 22). *Major challenges of Natural Language Processing (NLP)*. MonkeyLearn. Retrieved December 6, 2022, from https://monkeylearn.com/blog/natural-language-processing-challenges/

Saba, W. S. (2021, December 5). *Machine learning won't solve natural language understanding*. The Gradient. Retrieved December 6, 2022, from https://thegradient.pub/machine-learning-wont-solve-the-natural-language-understanding-challenge/

Sciforce. (2022, February 2). *Biggest open problems in natural language processing*. Medium. Retrieved December 6, 2022, from https://medium.com/sciforce/biggest-open-problems-in-natural-language-processing-7eb101ccfc9

Somers, M. (2019, March 8). *Emotion AI, explained*. MIT Sloan. Retrieved December 6, 2022, from https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained

Teufl, P., Payer, U., & Lackner, G. (2010). From NLP (natural language processing) to MLP (machine language processing). *International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security*, 256–269. https://doi.org/10.1007/978-3-642-14706-7_20

Wiggins, W. F., & Tejani, A. S. (2022). On the opportunities and risks of foundation models for Natural Language Processing in Radiology. *Radiology: Artificial Intelligence*, *4*(4). https://doi.org/10.1148/ryai.220119