

Text Prediction Using Artificial Intelligence: An Analysis of Two Text Prediction Systems

Hayden Wimmer
hwimmer@georgiasouthern.edu
Georgia Southern University
Statesboro, GA, 30460, USA

Jie Du
dujie@gvsu.edu
Grand Valley State University
Allendale, MI, 49401, USA

Peter Heri
pheri@georgiasouthern.edu
Georgia Southern University
Statesboro, GA, 30460, USA

Abstract

Natural language Processing is a discipline under artificial intelligence that involves interaction between human language and computer systems. It involves analyzing and representation of natural language, the ability to comprehend both text and spoken words. Natural language processing has evolved to the extent of having the ability to give useful responses to human beings. Large language models have been making landmark advances with new more efficient algorithms and improved hardware and processing power. Models like Google's BERT power predictive text in search predictions. Recently, companies have been training models on billions of parameters, a task that was not feasible just a few short years ago. OpenAI is the market leader in this technology; however, competitors have emerged. This research project aims to investigate perceptions of OpenAI versus an emerging competitor, AI21 on the ability to answer questions and predict text. We developed two web applications that allowed users to key in any questions or text in the textbox, the web application will then answer the user query as a response. The web applications were embedded with Jurassic-1 language model API and the GPT-3 language model APIs. Subjects asked the AI systems questions and rated their perceptions of the results. Furthermore, we investigate perceived privacy of AI systems via a post survey.

Keywords: Artificial Intelligence, Natural Language Processing, Text Prediction.

1. INTRODUCTION

Processing and extracting meaning from human text and speech has challenged AI researchers. Both large corpuses of data and mass processing

power are required to construct even the smallest and simplest models and sentient AI remains a distant horizon. Organizations have leveraged the power of the cloud and large computing resources and modern AI techniques to train large language

models using billions of parameters. The most familiar application of large language models is on Google search's type ahead search suggestions and now many chat bots. One of the most famous language models is the Generative Pre-trained Transformer 3 (GPT3) developed by OpenAI. It is the third-generation language prediction model in the GPT-n series. GPT-3 model has 175 billion parameters while its predecessor GPT-2, has 1.5 billion parameters. Recently, more organizations have entered this competitive marketplace of large language models offering their own pre-trained models.

In this work, we examine two of the leading pretrained language models, OpenAI and AI21. We seek to answer which system has better perceived performance as well as the impact of demographics and familiarity with AI on perceptions of privacy of AI systems. We develop two identical systems on the Python/Django framework which submits to its respective web API. We ask survey participants to ask each system a few questions and then answer a brief survey on that system. Which system is presented first is randomized to account for the learning effect. Following testing both Web applications, participants answer a brief post-survey on perceived privacy. We seek to answer a) which system has better performance as indicated by human subjects and 2) do demographics have any influence on participants perceived privacy.

The remainder of this paper is organized as follows, next we provide a brief review of related works, then we discuss our methodology in detail, following methodology we present our results to the aforesaid research questions, and then conclude our work.

2. RELATED WORKS

Language Models in NLP

Natural Language Processing (NLP) has been on the rise in recent years. Several language models perform NLP-related tasks. One of the language models is GPT which can be fine-tuned for a specific task such as sentiment analysis, text summarization. Panchbhai & Pankanti (2021) applied a GPT-2 model for a movie review task. A dataset from the IMDB website was used. 50 reviews of the dataset were used where 46% of the reviews were taken positive and 54% marked negative. Special character prompt and natural language prompt were used to test the GPT-2 model. The results indicated that GPT-2 performed better in the natural language prompt

sentiment analysis compared to the character prompt.

Bidirectional Encoder Representations from Transformers (BERT) is another language model. BERT was developed by Google and has an encoder that generates contextual embedding from any kind of sequence. This capability is essential for sentiment analysis or question answering. González-Carvajal and Garrido-Merchán (2020) empirically tested a BERT model using different scenarios compared to the traditional TF-IDF (Term Frequency - Inverse Document Frequency) vocabulary fed machine learning algorithms like Ridge Classifier and Voting Classifier. The BERT emerged the best with an accuracy of 0.9387, voting classifier followed with an accuracy of 0.9007, and the Ridge classifier with an accuracy level of 0.8990 (González-Carvajal and Garrido-Merchán (2020).

Wang and Cho (2019) showed that BERT is a Markov random field language model and compared BERT to traditional left-to-right language models. A pre-trained BERT model was taken on a mix of Toronto Book Corpus and Wikipedia with their Pytorch implementation. The quality of the generations was evaluated by computing the BLEU (Bilingual Evaluation Understanding) between the generations and the original data distributions. The BERT generations were found to be of low quality but very diversified compared to GPT.

Yu, Su, and Luo (2019) claimed that BERT lacks task-specific knowledge and domain-related knowledge. They proposed a BERT-based text classification model BERT4TC via constructing auxiliary sentences to address the task-awareness problem. The language model was evaluated on different datasets and the results showed that BERT4TC with suitable auxiliary sentences significantly outperforms both typical feature-based methods and finetuning methods.

Xu, Liu, Shu, and Yu (2019) explored turning customer reviews into a large knowledge base that can be used to answer user questions, which is referred to as Review Reading Comprehension (RRC). A novel post-training approach was used to fine-tune the BERT language model and to increase performance. Restaurant reviews and amazon shopping reviews were used as datasets for the RRC. The experimental results showed that the proposed post-training is highly effective (Xu et al. (2019).

Word Prediction

Trnka and McCoy (2007) evaluated word prediction methods using the keystroke savings and found that a larger amount of out-of-domain language is more beneficial than a smaller amount of in-domain language for training any model and the topic model can be fine-tuned even when there is dissimilarity in text. It was suggested that adaptive language models have the potential to outperform both in-domain and out of domain language models.

Grujić and Milovanović (2019) investigated word prediction based on their associative relations using a neural network. The dataset was collected from publicly available books in the Serbian language, which includes dictionaries, wiki pages, and historical books. The results showed that the number of times a word appears in the training set does not affect the outcome of this particular test. It was also found that in the case of best-performing associations, most clues are not close to the target while the worst-performing associations have all clues together into the target area.

Rojan, Alias, Rajan, Mathew, and Sudarsan (2020) applied the BERT language model into Malayalam. The pre-trained language model task is to predict the original vocabulary id of the masked word based only on its context. The proposed model has the capability of solving general tasks like next sentence prediction and mask language modeling. The model showed outstanding results of accuracy of 83% for next sentence prediction compared to the original BERT with an accuracy of 54% for the same set of data.

Risk and Trust in AI

Crockett, Garratt, Latham, Colyer, and Goltz (2020) investigated whether there was a difference when comparing perceived risk and trust in AI between the general public and college students studying computer science. A questionnaire was developed to collect people's perceptions of trust and risk in AI applications. The participants were asked to first rate the risk of the system on a scale of 0-10 and then answer a set of opinion questions on trust bias, ethics, and whether they support the development of such systems. They found that in specific applications such as medical imaging and diagnosis, there was a significant difference of opinions between the two groups with regards to risk. Both groups scored a medium risk regarding AI taking greater control over cybersecurity and strongly agreed that education in how AI works was significant in building trust. Perception of risk

was found to be greater when the outcome of an error is more personal or serious.

Hengstler, Enkel, Duelli, and Change (2016) investigated trust in applied AI. Semi-structured interviews were conducted with the workers working in the transportation and medical technology industries. Informal follow-ups via emails and short phone calls were also used. The results showed that operational security, data security, and purpose were the eminent factors influencing trust in the technology. A domestic development process for applied AI was proposed which include stakeholder alignment, transparency in development, and early, proactive communication.

Kim, Ferrin, and Rao (2008) developed a theoretical framework for the trust-based decision-making process. The framework was tested using the structural equation modeling technique on Internet consumer purchasing behavior data collected via a web survey. The results suggested that a consumer's trust has a strong positive impact on purchasing a product and a strong negative effect on consumer's perceived risk. The factors that influence a consumer's trust were also identified, including consumer disposition to trust, reputation, privacy concerns, security concerns, the information quality of the Website, and the company's reputation.

3. METHODOLOGY

During the summer of 2022, we conducted an IRB-approved study of analyzing two text prediction systems. This section gives an overview of the study procedure and describes the text prediction systems development, survey administration, and data analysis procedure.

Study Procedure

The participants were asked to follow a 6-step process to complete the study (see Figure 1). In Step 1, they took a pre-task survey to answer several demographics questions including gender, age, education level, occupation, and familiarity with AI. In Step 2, participants were asked to complete a task of asking a few questions (created by the user) via two web applications developed on Python and the Django framework and hosted on PythonAnywhere. The two systems utilize the OpenAI model and the AI21 model, respectively. The details about the system development are presented later. The interface for the tasks was identical between the OpenAI and AI21 language models and which language model they interacted was randomized



Figure 1: The Study Procedure

by the Qualtrics survey software to mitigate any learning effect. In Step 3, the participants completed a post-task survey to rate the system. In Step 4, the participants performed the same task for the second system. To further emphasize, the systems were identical with the exception of the API with which they communicated, and the tasks were randomized. In Step 5, the participants answered the same questions in the post-task survey for the second system. During the last step, the participants were surveyed on their perceptions of privacy when using the prediction systems.

Development of Two Prediction Systems

To test information usefulness, reliability, sufficiency, satisfaction, and quality of artificial intelligence in text prediction, two web applications were developed which will allow the user to key in any text or a series of words in a text box, the web application will then predict the following words or sentences. The web application was developed using Python Django framework, JavaScript, HTML, and CSS programming languages. Django framework was used to develop the application because of its high-level Python web framework attribute, and its rapid development of secure web applications. JavaScript was used to create div elements in the application, the div elements were used to display the predictions after the user types in any series of words or texts. HTML was used to develop the structure of the web application while CSS was used to modify the design of the application. Among the two web applications, one uses AI21’s Jurassic-1 language model API with the following parameters: maximum tokens of 15, a top K return of 0, and a temperature of 0.0. The second web application is embedded with OpenAI’s GPT-3 language model using the same parameters as the Jurassic-1 language model. The user interface of the two systems is identical (see Figure 2).

To interact with the prediction systems, the user will be making a GET request to the APIs embedded in the web applications by entering any string of text or words. The API will then predict the following strings of texts or words and return them in form of JavaScript Object Notation. To be specific, the user will create a question and type it in the textbox below “Ask a Question” and click the “Search” button when he is done typing. The question typed in by the user

is then displayed first in a blue box and the prediction from the system is displayed in a grey box just below the question box. Two sample questions and the predicted text from the system is shown in Figure 2.

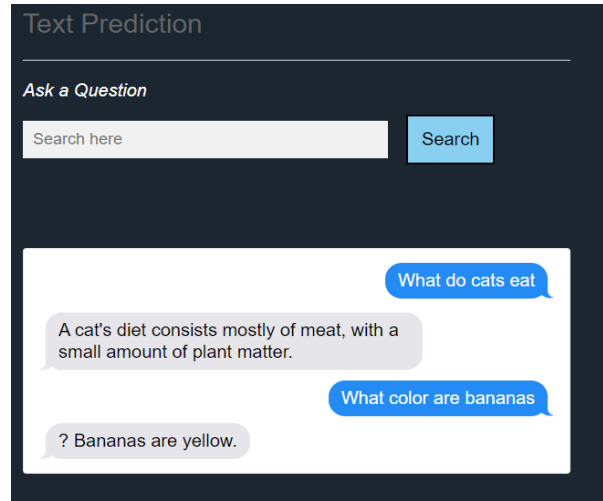


Figure 2: The UI of the Text Prediction System(s)

Survey Administration

An electronic Likert-scale questionnaire was implemented to survey the participants to gather their ratings on the text prediction systems and their perceived privacy when using the systems. The survey contains three sections: the pre-task section, the post-task section, and the perceived privacy section with a target completion time of less than 15 minutes. The pre-task section contains seven demographics questions including gender, age, education level, occupation, and familiarity with AI. The pos-task section consists of five questions about the overall ratings of the text prediction system. The five questions for the OpenAI are indexed as O1 to O5 and the five questions for the AI21 are indexed as J1 to J5. The perceived privacy section has nine questions about people’s perception of privacy when they use the text prediction system. Survey participants were recruited from the Amazon MTurk service with a restriction to a location in the US. There were 39 total respondents and 8 had to be removed due to non-completion of the survey.

Demographic Category	Percentage
Gender	Male: 24 Female: 14 No binary/third gender: 0 Prefer not to say: 1
Marital Status	Married: 13 Windowed: 1 Divorced: 7 Never married: 18
Age Representation	Under 18: 1 18-24: 1 25-34: 8 35-44: 17 45-54: 6 55-64: 6 65-74: 0 75-84: 0 85 or older: 1
Education Level	High school graduate: 6 Some College: 12 2-year degree: 6 4-year degree: 13 Professional Degree: 1 Doctorate: 1
Occupation	Employed full time: 30 Employed part time: 3 Unemployed looking for work: 2 Unemployed not looking for work: 2 Retired: 1 Student: 0 Disabled: 1
Familiarity with AI	Strongly agree: 8 Agree: 11 Somewhat agree: 16 Neither agree or disagree: 1 Somewhat disagree: 2 Disagree: 0 Strongly disagree: 1
Familiar with text prediction	Strongly agree: 2 Agree: 10 Somewhat agree: 12 Neither agree or disagree: 1 Somewhat disagree: 6 Disagree: 5 Strongly disagree: 3

Table 1: Survey Demographics Profile

All items (see Appendix) except the demographics items are anchored on 7-point Likert scales, with the following weights on each response, Strongly Disagree = 7, Somewhat Disagree = 6, Disagree = 5, Neutral = 4, Agree = 3, Somewhat Agree = 2, and Strongly Agree = 1.

A survey service from Qualtrics was used to administer the survey questions. The data collection yielded 31 usable survey response sets. Table 1 summarizes the demographics of the sample.

Data Analysis

We use SPSS to perform a two-step analysis on the data collected via the survey. First, a paired samples t-test was run to examine whether our participants evaluated the two systems differently. A significant t-value will indicate a significant difference between the ratings on the two systems. Second, we employed ANOVA to test the effects of the demographics items on the participants’ perceived privacy when using the prediction system. In our study, we have seven demographics items and would like to understand whether those demographics items have an effect on people’s perceptions of privacy when using the prediction systems.

4. RESULTS

This section breaks down the results for each step of the data analysis process. The results of a paired samples t-test were reported first to compare the two systems. T-test was chosen because of its ability to adapt to smaller sample sizes (De Winter, 2013). The results of ANOVA between different types of demographics and perceived privacy were reported next.

Comparison of Two Systems

To test the difference between OpenAI and AI21 language models, we employed a paired samples t-test using SPSS to compare the 5 rating questions (J1 – J5 to O1 – O5) between the two systems (see Appendix). Table 2 shows the descriptive statics of each pair. The means from the Open AI model are less than the ones from the AI21 Jurassic model, which means the ratings to the Open AI model are more positive compared to the AI21 model.

The results indicate a significant difference on each rating item between the two systems (see Table 3). We found that in all cases, our participants preferred the responses generated by the OpenAI model over the responses generated by the AI21 model. To be specific,

- For the pair J1-O1, the results indicate a significant difference regarding information usefulness between the two systems (t = 5.668, p = .000). The participants believe that the OpenAI model provides more useful information than the AI21 model.
- For the pair J2-O2, the results indicate a

significant difference regarding information reliability between the two systems ($t = 4.522, p = .000$). The OpenAI model was found to provide more reliable information compared to the AI21 model.

- For the pair J3-O3, the results show a significant difference regarding information sufficiency between the two systems ($t = 6.722, p = .000$). The OpenAI model was rated higher by our participants regarding information sufficiency.
- For the pair J4-O4, the results show a significant difference on user satisfaction between the two systems ($t = 5.917, p = .000$). The participants were more satisfied with the information provided by the OpenAI model than the one provided by the AI21 model.
- For the pair J5-O5, the results indicate a significant difference on information quality between the two systems ($t = 5.740, p = .000$). The participants rated the information provided by the OpenAI model with a higher quality compared to the one provided by the AI21 model.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	J1	4.55	31	1.690	.304
	O1	2.32	31	1.579	.284
Pair 2	J2	4.71	31	1.865	.335
	O2	2.81	31	1.833	.329
Pair 3	J3	5.26	31	1.751	.314
	O3	2.71	31	1.697	.305
Pair 4	J4	5.03	31	1.703	.306
	O4	2.58	31	1.876	.337
Pair 5	J5	4.94	31	1.731	.311
	O5	2.68	31	1.815	.326

Table 2: Paired Samples Statistics

We, therefore, state that there is a difference regarding overall ratings between the two systems and conclude that the OpenAI model is rated significantly higher than the AI21 model.

Paired Samples Test

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	J1 - O1	2.226	2.186	.393	1.424	3.028	5.668	30	.000
Pair 2	J2 - O2	1.903	2.343	.421	1.044	2.763	4.522	30	.000
Pair 3	J3 - O3	2.548	2.111	.379	1.774	3.323	6.722	30	.000
Pair 4	J4 - O4	2.452	2.307	.414	1.605	3.298	5.917	30	.000
Pair 5	J5 - O5	2.258	2.190	.393	1.455	3.062	5.740	30	.000

Table 3: The Paired Samples t-test Results

Effects of Demographics on Perceived Privacy

We did an ANOVA test between different types of demographics, which include gender, marital status, age, education level, occupation, familiarity in comparison to the nine items of perceived privacy (dependent variables) (see Appendix). To be specific,

- Utilizing gender (D1) as an independent variable, the results showed that there was a statistically significant difference in the responses to four items (P2, P3, P5, and P9) between the gender groups (P2: $F(1, 29) = 4.176, p = .050$; P3: $F(1, 29) = 5.877, p = .022$; P5: $F(1, 29) = 5.100, p = .032$; P9: $F(1, 29) = 8.209, p = .008$).
- Utilizing marital status (D2) as an independent variable, the results revealed that there was a statistically significant difference in the responses to P9 between the marital status groups ($F(3, 27) = 5.717, p = .004$).
- Utilizing age (D3) as an independent variable, there was no significant difference in any perceived privacy items between the age groups.
- Utilizing educational level (D4) as an independent variable, the results revealed that there was a statistically significant difference in the responses to P9 between the education groups ($F(3, 27) = 4.065, p = .017$).
- Utilizing occupation (D5) as an independent variable, the results revealed that there was a statistically significant difference in the responses to three items (P1, P3, and P9) between the occupation groups (P1: $F(3, 27) = 3.370, p = .033$; P3: $F(3, 27) = 3.006, p = .048$; P9: $F(3, 27) = 3.007, p = .048$).
- Utilizing familiarity with AI (D6) as an independent variable, there was no significant difference in any perceived privacy items between these groups.
- Utilizing familiarity with prediction systems (D7) as an independent variable, the results showed that there was a statistically significant difference in the responses to five items (P5, P6, P7, P8, and P9) between the groups (P5: $F(6, 24) = 5.178, p = .002$; P6: $F(6, 24) = 3.592, p = .011$; P7: $F(6, 24) = 3.150, p = .020$; P8: $F(6, 24) = 4.175, p = .005$; P9: $F(6, 24) = 4.067, p = .006$).

We conclude that at least some demographics including gender, marital status, education, occupation, and familiarity with AI have an effect on people’s perceived privacy when using prediction systems.

5. DISCUSSIONS AND LIMITATIONS

The paired samples t-test reveals a significant difference between the two systems on the five overall rating items including usefulness, reliability, sufficiency, user satisfaction, and quality. The OpenAI model was rated significantly higher than the AI21 model by our participants for all the five items. A comparison between the two systems when using the same input is shown in Figure 3. The responses from both systems seem reasonable. When comparing them, the responses from OpenAI are much more detailed. For instance, when a user types in "life is like", the AI21 model will respond with "a box of chocolates", a well-known quote from the movie Forrest Gump which follows by "you never know what you're going to get." It is assumed that people who have watched the movie understand the meaning of this phrase. But for those who never watched the movie, it may be hard for them to comprehend this phrase. It is likely that those people would rate the information provided by the AI21 model not useful. On the other hand, the responses from the OpenAI model are more straightforward. The responses are stand-alone and easy to understand. The comparison between the two systems next to each other using the same input helps explain the findings of our study. Users are more satisfied with the information provided by the OpenAI model which is more useful, sufficient, reliable, and high quality compared to the one provided by the AI21 model.

The OpenAI Model	The AI21 Model
<p>life is like</p> <p>a journey Life is like a journey. It's full of ups and downs, but it's also full of surprises. You never know what's going to happen next.</p> <p>Mackinac Island is like</p> <p>no other place in Michigan. The island is a popular tourist destination for its natural beauty, history, and culture. The island is also home to a number of businesses and organizations, including the Mackinac Bridge Authority, which operates the Mackinac Bridge.</p>	<p>life is like</p> <p>a box of chocolates</p> <p>Mackinac Island is like</p> <p>stepping back in time</p>

Figure 3 A Comparison between the Two Systems Using the Same Inputs

We also examined the effects of demographics on perceived privacy. The results show that there is a significant difference on people's perception of privacy between the demographics groups including gender, marital status, education, occupation, and familiar with AI. Between different gender groups (D1), the means of the responses to four items (P2, P3, P5, and P9) were significantly different. Although ANOVA cannot tell us which groups are different from each other, the results suggest that gender plays a role in

people's perceptions of privacy when using prediction systems. When comparing the effects of marital status (D2) and education (D4) on perceived privacy, we found that there was a significant difference on people's perception whether the AI web application has their best interest in mind (P9) between those demographic groups. Other than that, we did not find any difference on other perception items between those groups. A possible explanation might be that our sample does not have enough representatives for each group in D2 or D4 and thus no difference was detected between the groups. Between different occupation groups (D5), the means of the responses to three items (P1, P3, and P9) were significantly different. The results suggest that different occupation groups have significantly different trust on the system using their personal information properly. When comparing the effect of familiarity with prediction system (D7), we found that there was a significant difference on perceptions of saving time (P6), convenience (P7), and risk (P8 and P9). The results indicate that familiarity with prediction system plays a role in people's perceptions when using such a system. It seems reasonable that user's familiarity of technology impacts their perceived benefits of using technology.

It is interesting to find that age (D3) or familiarity with AI (D6) does not have any effect on the perception items at all. This might be caused by our sample. Our sample did not spread out in the age groups. 90% of the participants indicated that they were familiar with AI, which means only 10% of the participants belong to the group labeled with unfamiliarity with AI. Such unbalanced groups may contribute to the findings that D6 does not influence perceived privacy.

There are several implications. For academics, we developed two text prediction systems using different language models. The systems were tested and evaluated. The findings shed light on how to improve the performance of text prediction systems. There are also implications for practitioners in the field of NLP. The effects of demographics on people's perceived trust, risk, and privacy on a text prediction system can be further explored to better understand how to design text prediction systems.

There are some limitations in this study. While the sample size is acceptable, a much larger sample size would give more reliable statistical results. The ANOVA results show that there is a significant difference between demographics groups on people's perceived privacy. But one of the

limitations of ANOVA is that it won't tell you which statistical groups were different from each other. For instance, we found a significant difference on perceived risk between groups based on familiarity with prediction system. But it is unclear which groups had a different mean unless additional test is conducted. We will address such limitation by running additional test to better understand the effects of demographics on people's perceived privacy when using prediction systems. Additionally, testing the antecedents and factors and assigning weights to these factors to study why participants favored on set of responses is a direction for future research.

6. CONCLUSION

Natural language processing is a large domain, text prediction is one of the major tasks natural language processing presents. This research project has presented two language models; Jurassic-1 and the GPT-3 language model, tested their efficiency, accuracy, reliability, and trust in artificial intelligence. Participants were able to interact with the two language models and later on asked a few questions about how their experience was in using the application. Paired samples t-test and ANOVA were then done to test which language model is better and if demographics influence perceived privacy in text prediction.

7. REFERENCES

- Crockett, K., Garratt, M., Latham, A., Colyer, E., & Goltz, S. (2020, 19-24 July 2020). *Risk and Trust Perceptions of the Public of Artificial Intelligence Applications*. Paper presented at the 2020 International Joint Conference on Neural Networks (IJCNN).
- De Winter, J. C. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research, and Evaluation, 18*(1), 10.
- González-Carvajal, S., & Garrido-Merchán, E. C. J. a. p. a. (2020). Comparing BERT against traditional machine learning text classification.
- Grujić, N. D., & Milovanović, V. M. (2019). *Natural Language Processing for Associative Word Predictions*. Paper presented at the IEEE EUROCON 2019-18th International Conference on Smart Technologies.
- Hengstler, M., Enkel, E., Duelli, S. J. T. F., & Change, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *105*, 105-120.
- Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems, 44*(2), 544-564. doi:<https://doi.org/10.1016/j.dss.2007.07.001>
- Rojan, A., Alias, E., Rajan, G. M., Mathew, J., & Sudarsan, D. (2020). *Natural Language Processing based Text Imputation for Malayalam Corpora*. Paper presented at the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC).
- Trnka, K., & McCoy, K. F. (2007). *Corpus studies in word prediction*. Paper presented at the Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility.
- Wang, A., & Cho, K. J. a. p. a. (2019). Bert has a mouth, and it must speak: Bert as a markov random field language model.
- Xu, H., Liu, B., Shu, L., & Yu, P. S. J. a. p. a. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis.
- Yu, S., Su, J., & Luo, D. (2019). Improving BERT-Based Text Classification With Auxiliary Sentence and Domain Knowledge. *IEEE Access, 7*, 176600-176612. doi:[10.1109/ACCESS.2019.2953990](https://doi.org/10.1109/ACCESS.2019.2953990)

APPENDIX A
Constructs and Measurement Items

Constructs	Measurement items
Demographics	Age (D1) Gender (D2) Occupation (D3) Marital Status(D4) Education level(D5) Overall, I am familiar with Artificial intelligence(D6) I am familiar with text prediction using artificial intelligence(D7)
Overall Rating	Overall, I think the website provides useful information (J1/O1) This website provides reliable information (J2/O2) This website provides sufficient information when I try to type any text (J3/O3) I am satisfied with the information this website provides (J4/O4) Overall, the information that website provides is of high quality (J5/O5)
Perceived Privacy	I am concerned that: The web vendor will use my personal information for other purposes without authorization (P1) This website is collecting too much information from me (P2) The web vendor will sell my personal information to others without my permission(P3) This website vendor gives the impression that it predicts texts accurately (P4) This text prediction site is trustworthy (P5) I can save time using this website(P6) I think using this web application for text prediction is convenient (P7) How would you rate your overall perception of risk from this site (P8) I believe the artificial intelligence web application has my best interest in mind (P9)