Good day. Thank you very much for attending my presentation.

My name is Nomusa Majola and I’ll be presenting my research proposal titled "Real-Time Customer Experience Tracking in Front-Shops using Deep Sentiment Analysis" aiming to harness advanced sentiment analysis models to provide real-time insights into consumer sentiments within retail front-shop settings.

<<<Next Slide>>>

In this presentation, I will outline the significance, research problem, research question, objectives, key literature, methodology and design, ethical considerations, and the proposed timeline for this project.

<<<Next Slide>>>

In the fast-paced realm of modern business, the customer experience, or CX for short, is central to the success and longevity of enterprises, as it fosters customer loyalty and propels revenue growth and market distinctiveness (Lemon and Verhoef, 2016; Verhoef, 2009).

Traditional CX approaches, like surveys and feedback forms, are marred by drawbacks like response bias, delayed feedback, and limited scalability (Verhoef et al., 2015) and are time consuming and expensive to implement.

<<<Next Slide>>>

The rise of social media and online review platforms have accelerated the rate of change in consumer trends (Hajli et al., 2017) and introduced an avenue for customers to express their opinions. Thus creating a trove of textual data rich in Customer experience sentiment (Ott et al., 2011). Sentiment analysis is a subfield of natural language processing (NLP) that involves automatically extracting sentiment from text (Pang & Lee, 2008).

<<<Next Slide>>>

Sentiment analysis using deep learning techniques–or deep sentiment analysis for short–has emerged as a very promising avenue (Zhang et al., 2018).

The effectiveness of deep sentiment analysis has mainly been driven by strides in deep learning methodologies that have significantly bolstered the accuracy of sentiment analysis models. The advent of transformer-based models like Bidirectional Encoder Representations Transformers (BERT) and its variants has marked a paradigm shift in sentiment analysis, allowing models to contextualize words in sentences and effectively capture contextual dependencies (Devlin et al., 2018).

<<<Next Slide>>>

In a business context, deep sentiment analysis holds the promise of capturing larger-scale customer opinions and gauging CX in real-time in a cost effective manner. However, this approach is not tailored to provide more fine-grained insights, such as the CX in specific front-shops of a business where many or even most consumer-agent interactions typically occur.

<<<Next Slide>>>

This research aims to leverage deep sentiment analysis in front-shops to capture and analyse consumer-agent voice interactions and automatically recognise the sentiment in these interactions.

Timely identification of negative sentiments can facilitate swift interventions, enabling businesses to promptly address issues and forestall potential escalations (Verhoef et al., 2009).

Simultaneously, real-time recognition of positive sentiments can facilitate immediate acknowledgment of exemplary customer service, ultimately augmenting customer satisfaction.

<<<Next Slide>>>

I discovered that text-based sentiment analysis is more common than speech-based sentiment analysis, and couldn't find specific literature on front-shop customer sentiment analysis.

Given the time constraints, I’ve decided to describe three key papers on speech-based sentiment analysis as context for this research.

<<<Next Slide>>>

Zhang et al. (2023) presented a novel approach to speech sentiment recognition that makes use of both, context attention, as well as time-frequency features. The proposed method first obtains time-frequency features using a 2D CNN and then utalises a dual global context attention mechanism to rank feature vectors to recognize sentiment.

One of the main strengths of the approach is the dual global context attention mechanism, which capture informative features in speech. However, the limitations of the approach are unclear in the paper, such as its potential performance on datasets significantly different from those used. Finally, a comparison with other recent approaches was also lacking.

<<<Next Slide>>>

Liu et al. (2023) implemented a technique that uses time-frequency features which focus on alleviating sentiment class imbalances, by applying time-domain and frequency-domain features separately. These features were used to recognize sentiments.

It is unclear how the method performs on a larger different dataset than the training set, and the complexity of the approach is also an issue, especially with regards to model interpretation.

<<<Next Slide>>>

Mustaqeem et al. (2023) introduced an innovative model called the Deep Echo-State-Network (DeepESN) characterized by a CNN coupled with a multi-headed attention mechanism. The proposed DeepESN synergizes reservoir computing to reduce model complexity which reduces overfitting and improves interpretability.

The model efficacy was demonstrated on two prominent benchmark datasets, resulting in improved accuracies and reduced computational time, while demonstrating clear subject-independence.

Reduced computational time increases the viability of the approach in deployment. Still, the proposed model’s ability to deal with intricate speech variations like accents and noise is not clear, and the model’s viability is dependent on high quantity and quality data.

<<<Next Slide>>>

To summarise, CNNs are a key technique for feature extraction, coupled with attention-based mechanisms to model temporal dependencies in data. A key downside to these-and in fact all deep learning-models is their complexity and obscurity, as well as dependence on large quantities of high quality data.

All in all, Mustaqeem et al.’s model (Mustaqeem et al., 2023) shows the greatest promise in the proposed research.

<<<Next Slide>>>

Based on the literature review, the research question posed is “How effective is the model proposed by Mustaqeem et al. (2023) in the context of front-shop consumer sentiment recognition?”

<<<Next Slide>>>

The following systematic objectives will guide the process of answering the research question:

Conduct an extensive further literature review to uncover other salient techniques

Data collection and processing in which microphones will be placed in a front-shop and customer-agent interactions will be recorded, standardised and processed to remove personally identifiable information.

Model design and implementation

Model training to produce a converged model

Model evaluation to evaluate the performance and generalization capability of the trained model.

And finally, formulation of an answer to the research question.

<<<Next Slide>>>

We adopt the design science research approach (Venable et al., 2017) to develop and evaluate a deep neural network-based system for sentiment recognition in the front-shop consumer context. This approach has the following phases:

1. Problem Understanding and Definition

2. Design and Development

2.1 Data Collection and Pre-processing

2.2 Deep Neural Network Architecture

2.3 Training and Validation

3. Evaluation

4. Conclusion

<<<Next Slide>>>

The first is: Problem Understanding and Definition:

The problem space of front-shop consumer sentiment recognition is analysed with a consideration of the potential challenges and a review of the existing literature on sentiment analysis, deep learning techniques, and relevant studies in retail and consumer behaviour to define the scope and requirements of our system.

<<<Next Slide>>>

The second phase is Design and Development which is broken down as follows:

a) Data Collection and Pre-processing: A diverse dataset of customer-agent interactions will be collected, along with customer self-declared experience sentiment of the interaction. A challenge with the collection of such datasets is their naturalness due to the informed consent data collection process. In this regard, negative encounters will be scripted to balance out the data set. A second challenge in this task is to pre-process and clean that data.

<<<Next Slide>>>

b) Deep Neural Network Architecture: The three deep neural networks will be designed and constructed as specified in the relevant papers. The challenge that is expected here is a lack of existing code and incomplete information provided on the implementations which is common with models described in research papers.

<<<Next Slide>>>

c) Training and Validation: The dataset will be split into training, validation, and test sets. The model will be trained and tuned on the training set, with the validation set used for training validation. The testing set is used in the next phase i.e. Evaluation. The main challenge here is to ensure that sufficient data samples are used in training or validation without resulting in too small of a test set that would be too small to inspire confidence.

<<<Next Slide>>>

The third phase is

Evaluation: To evaluate the viability of the proposed model, metrics such as accuracy, precision, recall and F1-score will be used. These are standard metrics that assess the system’s classification performance against the ground truth collected from consumers.

<<<Next Slide>>>

Conclusion: In this phase, the result of the evaluation phase will be used to draw conclusions towards answering the overall research question and providing insights into the implication and application of the research findings

<<<Next Slide>>>

The main challenge that arises as a result of the ethical considerations mentioned is that, as mentioned previously, participants informed of the data collection process will be likely to have more carefully controlled reactions, and more likely, it is expected that negative behaviours resulting in negative sentiment will be less likely under scrutiny.

To counter this, it may be necessary to script some negative encounters to balance out the data set, and it may be necessary to enlist the help of expert narrators in order to increase the naturalness of these recordings.

<<<Next Slide>>>

The research timeline follows the research objectives I spoke about earlier. The first phase spanning five weeks is an extensive literature review. This is followed by data collection which is expected to take a further five weeks. Following this is a one-month phase in which the models will be constructed and trained. Once training is complete, the test set will be used to evaluate and compare the models which is expected to take one week.

<<<Next Slide>>>

In conclusion, this study aims to enhance the accuracy and efficiency of customer sentiment tracking by tapping into cutting-edge technology.

It will address the traditional method limitations by employing advanced techniques that can offer more accurate and immediate insights. It will also focus on techniques that can capture the nuanced interactions between customers and agents.

Among various existing models, the research zeroes in on the innovative approach proposed by Mustaqeem et al. (2023). This model demonstrates substantial promise in speech-based sentiment recognition, making it a focal point of our investigation.

<<<Next Slide>>>

Ethical implications of data collection will be considered , ensuring compliance with regulations. A comprehensive risk assessment is a vital part of the process, addressing potential challenges and ensuring the protection of participants' rights and privacy.

Through this research, we aspire to advance the field of customer experience management by employing advanced sentiment analysis models. The insights obtained from customer-agent interactions can aid businesses in making informed decisions, enhancing customer satisfaction, and maintaining a competitive edge in the modern business landscape.

<<<Next Slide>>>

This is a list of references of all academic sources that have been used so far in this study

<<<Next Slide>>>

This concludes my proposal presentation. I thank you for listening and I would appreciate feedback and questions via email.

**References**

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2019)*, *1*, 4171–4186. https://arxiv.org/abs/1810.04805v2

Hajli, N., Sims, J., Zadeh, A. H., & Richard, M.-O. (2017). A social commerce investigation of the role of trust in a social networking site on purchase intentions. *Journal of Business Research*, *71*, 133–141.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69–96.

Liu, Z.-T., Han, M.-T., Wu, B.-H., & Rehman, A. (2023). Speech emotion recognition based on convolutional neural network with attention-based bidirectional long short-term memory network and multi-task learning. *Applied Acoustics*, *202*, 109178.

Mustaqeem, K., El Saddik, A., Alotaibi, F. S., & Pham, N. T. (2023). AAD-Net: Advanced end-to-end signal processing system for human emotion detection and recognition using attention-based deep echo state network. *Knowledge-Based Systems*, *270*, 110525.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination. *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 309–319.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Venable, J., Pries-Heje, J., & Baskerville, R. (2017). Choosing a design science research methodology. *In Proc. Australasian Conference on Information Systems (ACIS)*, 1–12.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: introduction to the special issue on multi-channel retailing. *Journal of Retailing*, *91*(2), 174–181.

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer experience creation: Determinants, dynamics and management strategies. *Journal of Retailing*, *85*(1), 31–41.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.

Zhang, P., Bai, X., Zhao, J., Liang, Y., Wang, F., & Wu, X. (2023). Speech Emotion Recognition Using Dual Global Context Attention and Time-Frequency Features. *In Proc. 2023 International Joint Conference on Neural Networks (IJCNN)*, 1–7.