B.Comp. Dissertation

# Development of Artificial Intelligence/Machine Learning Pipeline for Pre-training a Sales Foundation Model

By

Loo Wee Sing, Caspian Michael (A0214109R)

Department of Computer Science
School of Computing
National University of Singapore

2024/2025

B.Comp. Dissertation

# Development of Artificial Intelligence/Machine Learning Pipeline for Pre-training a Sales Foundation Model

By

Loo Wee Sing, Caspian Michael (A0214109R)

Department of Computer Science
School of Computing
National University of Singapore

2024/2025

Project No: H201570
Advisor: Senior Lecturer Cristina Carbunaru

Deliverables:
- Report: 1 Volume
- Project Repository: 1 Volume

## Abstract

The use of Artificial Intelligence/Machine Learning (**AI/ML**) for sales activities have gained significant interest over the past year spearheaded by an industry push for the development of AI products — as seen by the rise of commercialised AI (*e.g.* ChatGPT, Gemini, MetaAI *et cetera*.). At the heart of this growing interest are the significant improvements in the Large Language Models (**LLMs**) that drive these innovations. Yet, one of the greatest challenges in the field remains — the costly training (i.e. *learning*) process of LLMs. These costs typically lead to commercial AI/ML efforts being arrested in smaller, resource-strapped companies like StaffAny. In my work, I have implemented a model training pipeline based on Deep Bayesian Active Learning (**DBAL**) (Gal et. al, (2017)) to train a model ensemble that forms the foundation model for StaffAny's sales unit. I have also developed a Python-based application that highlights how to employ the predictive power of the foundation model in a User Interface (**UI**). Due to resource constraints, the model's application was only evaluated under simulated conditions.

Subject Descriptors:

I.2.7 Natural Language Processing
I.2.6 Learning
H.4.2 Information Systems Applications
H.3.3 Information Search and Retrieval
J.1 Administrative Data Processing

Keywords:

Software Engineering, Artificial Intelligence, Machine Learning, Deep Bayesian Active Learning, Ensembling, Natural Language Processing, AI in Sales

Implementation Hardware:

Apple M1 Chip, Macbook Air (M1 2020) Silicon

Implementation Software:

macOS Sequoia 15.0.1, Python 3.12, JupyterLab, pandas, seaborn, SciPy, NumPy, Matplotlib, TensorFlow, scikit-learn, PyTorch, Whisper, XLM-RoBERTa(**XLM-R**), OpenSMILE, Text-to-Text-Transfer-Transformer(**T5**), Streamlit, Flask, HubSpot API, Aircall API, StaffAny HubSpot Dataset, StaffAny Aircall Recordings, StaffAny Twilio Recordings

## Acknowledgements

Firstly, I am incredibly grateful to StaffAny for giving me this opportunity to take on such an ambitious project.

To my ex-boss, Eugene, thank you for having the trust and confidence in me to work with sensitive company data. I hope that you will find the results of this FYP useful.

To my company supervisor, Jeremy, your advice and direction were instrumental in navigating the complex requirements and ensuring I stayed focused on the end goal. It was the consistent check-ins early on in the project which led me to researching and implementing DBAL in the first place.

To Nicholas Malz, your academic review, *Deep Bayesian Active Learning with High-Dimensional Data – A Critical Review*, was an inspiration. It helped shape my approach to many challenges throughout this project. I would even say that it was pivotal. As Gen-Z would say, you came in clutch.

To my academic supervisor, Professor Cristina, thank you for your patience in dealing with my deafening radio silence throughout this entire journey. I would not have had the space to complete this project if it were not for the long leash of understanding accorded to me throughout the past year.

To Serena from Fuzzy Sequence, your inputs at the beginning of the project helped to steer my efforts and direction.

Finally, to my family, friends & Maddy, watching me hack out this project over the past year and losing sleep over the project completion, thank you for your support. I could not have done this without all of you. Your belief in my abilities and differing insights were invaluable. Thank you all for making this project possible.

## List of Figures

Table of Contents

StaffAny

## 1. Introduction

My Final Year Project (FYP) aims to enhance the sales process of StaffAny, a Human Resource Management System(HRMS) Company, through the development of an AI tool for the sales team. The key problem faced is a complex sales workflow that distracts the sales team from essential sales activities, ultimately contributing to a reduced ability to close more clients and generate more revenue.

To address this, an AI-driven AI/ML pipeline is built by leveraging numerical data from StaffAny's Custom Relationship Management (**CRM**) tool[1] and call recordings on their client communication toolchains[2] as training data. The resultant models trained from this data are in an effort to improve the outcome of a sales event[3] — to set an appointment with a sales prospect.

### 1.1 Using AI/ML Pipelines for Pre-training

AI/ML pipelines are integral in managing the process of training ML models to eventually serve AI tools. This training process is often complex, especially when dealing with large, diverse data sets. The training process is also typically iterative — involving a series of finetuning before the model has been trained. This is particularly so in the context of DBAL.

In the context of StaffAny, an AI/ML pipeline will allow for the flow of numerical & textual data from StaffAny's CRM Tool and audio data from StaffAny's client communication toolchains directly into the model training process. It is not simply a data pipeline as apart from automating data retrieval, other aspects of the training process are also part of this AI/ML pipeline, for instance, important data preprocessing, feature extraction and

[1] HubSpot
[2] Aircall & Twilio
[3] Refer to Appendix A

training-evaluation cycles that serve as important checkpoints in the model's training process. A well-crafted AI/ML pipeline ensures that model training is not just consistent and efficient, but scalable alongside dataset expansion.

Despite using such AI-empowered tools[4] and productivity toolchains provided by ChatGPT+, StaffAny's sales team do not have such an AI/ML pipeline set up, much less a data pipeline for data aggregation[5]. This introduces complexity to sales activities as it creates an over fragmented context when salespersons attempt to comprehend data across different data repositories whilst preparing for sales events.

As a growth stage startup, StaffAny has limited resources available to engineer such a pipeline as it requires costly operational overhauls and engineering work that is not part of their regular business operations. The most suitable option given these constraints are to engage in costly tooling upgrades — an endeavour StaffAny is unable to support. This supports a strong thesis for an AI/ML pipeline to be crafted to unify insights across these discrete data repositories via pre-training a foundation model, substantially enhancing the ability of the sales team to focus on setting appointments and improving revenue generation.

### 1.2 Pre-training a Foundation Model for Sales

Foundation models are particularly valuable because they can be pre-trained on a rich set of collected data. Such training enables foundation models to understand complex relationships and patterns inherent in data, which are harvested from processes characterised by dynamic environments with multiple factors at play, such as sales. Foundation models are typically pre-trained on a large corpora of textual, aural, and numerical data, resulting

[4] Aircall, Hubspot & Twilio are tools that have some form of AI-integrated tooling
[5] Yet.

in a model with a generalised understanding[6] of the context dictated by the training data's domain. According to Bommasani et al. (2021) [13], this allows foundation models to be further fine-tuned for a wide-array of specific tasks without encountering issues like over-fitting, a feature unique to foundation models. This means that, unlike task-specific models, foundation models possess the flexibility to adapt to various downstream tasks.

This is ideal in the context of sales, where the dataset is often large and varied, where individual pieces of data are usually reused across different sales events. The resultant potential from the development of such a model is limitless, with applications such as lead scoring, customer segmentation, and appointment setting just to name a few, all mission critical applications for a sales team. By leveraging the broad understanding learnt during pre-training, a foundation model can generate more informed predictions about sales events to provide deeper insights post and even during events. This could have an unintended effect on improving the efficiency and effectiveness of StaffAny's Sales Process.

Additionally, by applying novel techniques like DBAL, the foundation model can be continuously improved with minimal labelled data, allowing it to keep pace with evolving sales dynamics and improve its predictive accuracy over time without having to incur the huge costs it takes to pre-train a foundational model.

### 1.3 Other AI/ML Applications for Sales

The market is flooded with AI/ML tools already available to serve the sales domain. These applications come in various forms — chatbots, lead scoring algorithms, sentiment analysis tools, *et cetera*. Chatbots like Drift[7] and

[6] computationally, at least
[7] Include Website

Intercom[8] are useful for automating customer interactions, while tools like Salesforce's Einstein AI[9] assist in lead scoring. Sentiment analysis tools such as MonkeyLearn[10] provide insights into customer attitudes by analysing text from emails or chat logs.

Despite the utility offered by these applications, these tools are often highly specialised and address only one or a few aspects of StaffAny's sales process. In the off-chance that they do meet all the desired needs of StaffAny, it is likely that the attached price tag does not match its utility. Moreover, these tools are likely trained on a corpora of data that do not match the unique features of the data collected by StaffAny. A counter argument can be made that these tools could potentially be fine-tuned to find its place in serving StaffAny's needs within StaffAny's business operational framework. But, this only serves to highlight that these tools by virtue of not being trained on StaffAny's ground truth[11] would require additional effort to be baked into its operations.

In contrast, a foundation model provides a more comprehensive solution by integrating different types of StaffAny's own data — text, audio, and numerical — and learning from them collectively, albeit with its own set of unique technical hurdles. Additionally, the ground truth used to pre-train the foundation model comes directly from StaffAny's own data repositories. This allows the foundation model to understand the specific context of sales events within StaffAny to provision for and better serve the needs of StaffAny's sales team.

My FYP aims to build this foundation model by utilising historical numerical & textual data from HubSpot as well as past sales call recordings to pre-train the

[8] Include Website
[9] Include Website
[10] Include Website
[11] i.e. data

foundation model whilst building the AI/ML pipeline in parallel. This culminates in a sample web application developed to showcase how the foundation model can be utilised in downstream AI applications that fit into the existing sales operations at StaffAny. The AI/ML pipeline will be engineered for the orchestration of a stacked ensemble, based on findings from Gal et. al (2017) [6]. This is an agreeable methodology based on StaffAny's considerations[12].

[12] Refer to Appendix B,C,D for full breakdown of Sales Operating Framework & critical considerations for AI Use

2. Literature Review

NLP is one of the most significant and widely researched aspects of AI/ML, eventually finding its way into the mainstream through the advent of LLM powered chatbots. It is thus unsurprising that the current body of academic work detailing the use of NLP in sales activities is ever growing, with a torrent of research available. These publications exhibit that the application of AI/ML within a sales context is not a novel concept. Salesforce's Einstein AI is known to utilise NLP for lead scoring and sales predictions. According to Khang et al. (2021) [8] , the use of such NLP is known to improve decision-making capabilities. Gong.io[13] offers conversation intelligence by analysing sales calls using advanced NLP and NLP techniques[14], which could provide actionable insights for sales representatives according to Paschen et al. (2020) [11].

However, the application of learning techniques like DBAL for use in developing new foundation models for commercial settings are not as widely explored[15]. Understanding the implications of such research and applying the concepts laid out in landmark studies are crucial in optimising various aspects of StaffAny's sales cycle[16].  My FYP heavily references key findings from various authors and aims to apply the principles laid out in the published papers.

2.1 Natural Language Processing (NLP)

Bidirectional Encoder Representations from Transformers (**BERT**), XLM-R, and other transformer-based models are the obvious choices to emulate when developing a foundation model. BERT uses bidirectional training to understand language nuances, making it highly effective for various NLP

[13] Include Website
[14] In particular, extraction of key phrase, sentiment, and indicators of deal progress
[15] At the time of writing. This is an important caveat as the field of AI/ML research advances rather rapidly
[16] Refer to Appendix C for Sales Cycle Explanation

tasks like text classification, named entity recognition, and question answering, as presented by Devlin et al. (2019) [4]. Not to be outshined, XLM-R extends BERT's capabilities to a multilingual context, providing robust cross-lingual understanding across multiple languages , as presented by Conneau et al. (2020) [3].

These foundation models are pre-trained on a huge dataset, making them ideal candidates for fine-tuning for use in downstream sales applications where diverse language audio and textual data need to be processed efficiently.  This allows the model to generate rich feature representations of textual data featuring strong linguistic nuances and atypical linguistic features (i.e. StaffAny's dataset).

NLP tools are widely used today for feature extraction in both text and audio domains, allowing for a semantic structuring of what was previously unstructured raw data. Techniques like Named Entity Recognition (**NER**) and Part-of-Speech (**POS**) tagging are commonly applied to extract entities, keywords, and sentiments from text data.  Yet, insights from NLP alone are typically not sufficient enough to be used as context, especially when applied as is to sales activities, where context is typically more situational.

### 2.2 Ensembling

Prior research has shown that integrating both textual and audio features significantly enhances the predictive power of models through variational reduction, as researched by Ranganath et al. (2014) [12]. However, closer to reality, it is more common to find the use of numerical models such as RFC, SVM, and LR to be used in place to enhance predictive power. This approach is commonly referred to as hybrid ensembling. It enhances the model's ability to make data-driven decisions by integrating insights from both textual and numerical sources, ultimately providing a richer context for predictive analysis.

Numerical models are already widely used in business intelligence for predictive analytics and classification tasks. Thus, a numerical-NLP ensemble model is a low hanging fruit – these models can be effectively and easily ensembled with NLP models since they are usually already present. It is thus typically the most expedient way to create a more comprehensive solution that leverages both structured and unstructured data to enhance predictive capabilities. Companies like Gong.io and Salesforce have successfully implemented such ensemble approaches to enhance sales forecasting and customer sentiment analysis, as proposed by Paschen et al, (2020) [11] and Khang et. al (2021) [8].

Given the advances in NLP, it should not be too difficult to instead, create a NLP-NLP model ensemble. After all, Ranganath et al. (2020) [12] clearly concludes that the combination of both textual and audio features significantly enhances a model's predictive power. However, there are limitations to implementing a NLP-NLP ensemble. Large data volume and poor feature[17] extraction techniques can lead to disastrous effects during the pre-training process, as researched by Adnan et al. (2019) [14].

Mastering the pre-training process is not simply about the application of efficient NLP techniques. Much like how software design and algorithm optimisation go hand-in-hand, so too should the design of model learning and its counterpart learning techniques. In the face of large data volumes presented during pre-training, a common issue faced is the curse of

[17] Actual wording used was "*information*" but was synonymous in meaning to this report's use of the term *"features"*

dimensionality. Interestingly, much has been researched on how to face this curse during model pre-training.

### 2.3 Deep Bayesian Active Learning (DBAL)

The "*Curse Of Dimensionality*" refers to the exponential increase in computational complexity as the number of features in the data grows. When distilled in its computational representation, features can be thought of as a succinct representation of unstructured information in a matrix. As the number of features increase, this matrix grows exponentially, creating a higher dimensional matrix every time a new feature is added during the learning process. Thus, a single feature increase leads to a more than proportionate increase in dimensions.

This phenomenon often leads to issues such as overfitting, high computational costs, and degraded model performance, particularly when dealing with high-dimensional datasets, as stated by Bellman, 1961 [1], who later coined the term in his book. It is aptly named because the crux of pre-training a model is for the model to learn the features of a dataset to enhance predictive power. Logically, the more features there are, the richer the pre-training and the better the prediction. Yet paradoxically, adding too many features during learning actually risks the model's viability, requiring model revisions to undo time consuming and costly pre-training.

DBAL as a framework is particularly effective in addressing the curse of dimensionality. As demonstrated in Gal et al. (2017) [6], the focus was on reducing computational overhead by actively selecting the most informative data points rather than relying on exhaustive ensembling[18], which was less common at that time. In essence, DBAL incorporates Bayesian methods to reduce the amount of data required to train a model. In DBAL, the focus is on

---

[18] Which Gal et al. (2017) [6] dismissed at the time due to high computational costs

selectively choosing the most informative data points[19] to label and include in model training. This in turn reduces labelling costs and computational overhead, particularly useful when labelling is expensive or time-consuming. The idea of using variation as a source of inference for model performance is a reproducible technique, as recently researched by Giordano et al. (2024) [7], even finding its way into the Journal of Machine Learning Research.

Interestingly, using a Deep Ensemble based on DBAL was shown to be an equally effective tool to improve prediction power. In certain cases, it even outperformed conventional Bayesian approximations as presented by Lakshminarayanan et al. (2017) [9]. This was further supported by Beluch et al. (2018) [2], who showed that Deep Ensemble Active Learning (**DEAL**) could significantly enhance active learning efficiency by improving model uncertainty estimation. Moreover, Beluch et al. (2018) [2] noted that increasing the number of models in the ensemble did not significantly affect performance, as reviewed by Malz (2023) [10]. These results increased the confidence of employing DEAL in model pre-training. The only difference is that with DEAL, the entropy across model predictions is taken as a proxy for the uncertainty itself whereas in DBAL, the exact uncertainty of the prediction is measured. Given this minute difference, both methods are still valid active learning implementations and should be carefully considered in the pursuit of dimensionality reduction.

[19] i.e. data with the highest uncertainty/entropy

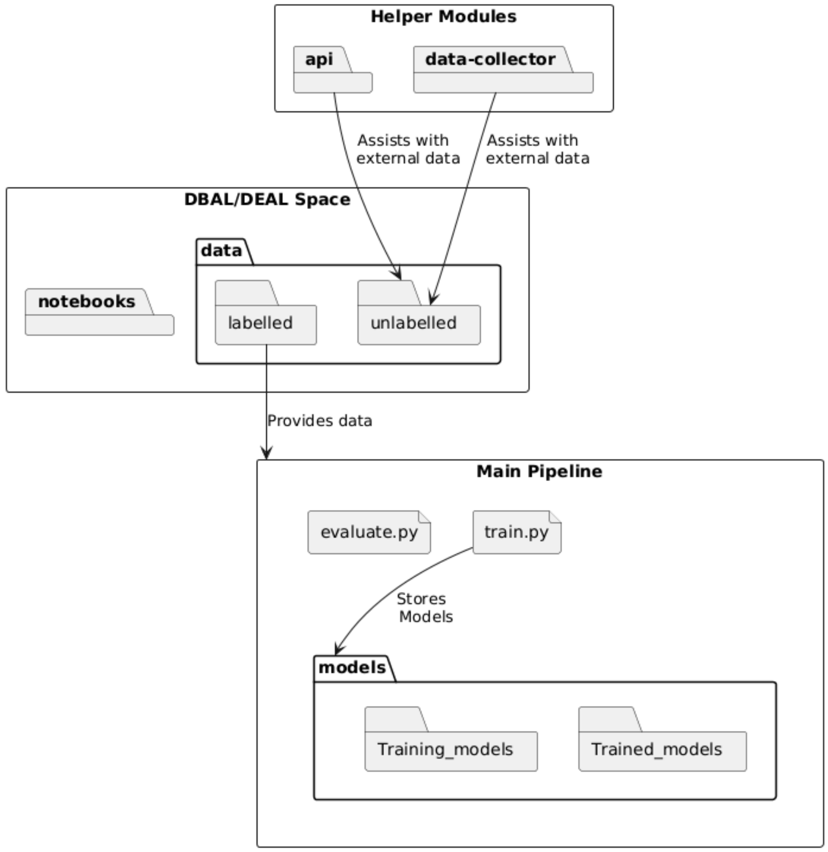



*Figure 1. Pipeline Design Overview*

The design of the AI/ML pipeline was loosely defined by StaffAny. The main criteria were 1) the ability to handle new sources of data without breaking the existing pipeline, 2) a clear separation of concerns between the model training pipeline from the data preparation modules and 3) An entirely local implementation of the pipeline due to the sensitive nature of the data and to maintain compliance with legal statues laid out in Personal Data Protection Act (PDPA)[20]. This inspired the creation of a fine-tuned pipeline for their needs, whilst fitting in the considerations of conventional long-lived AI/ML pipelines and Active Learning.

[20] PDPA is the Singaporean equivalent to the US HIPPA and EU GDPR, with slight legal differences

3.1 Main Pipeline

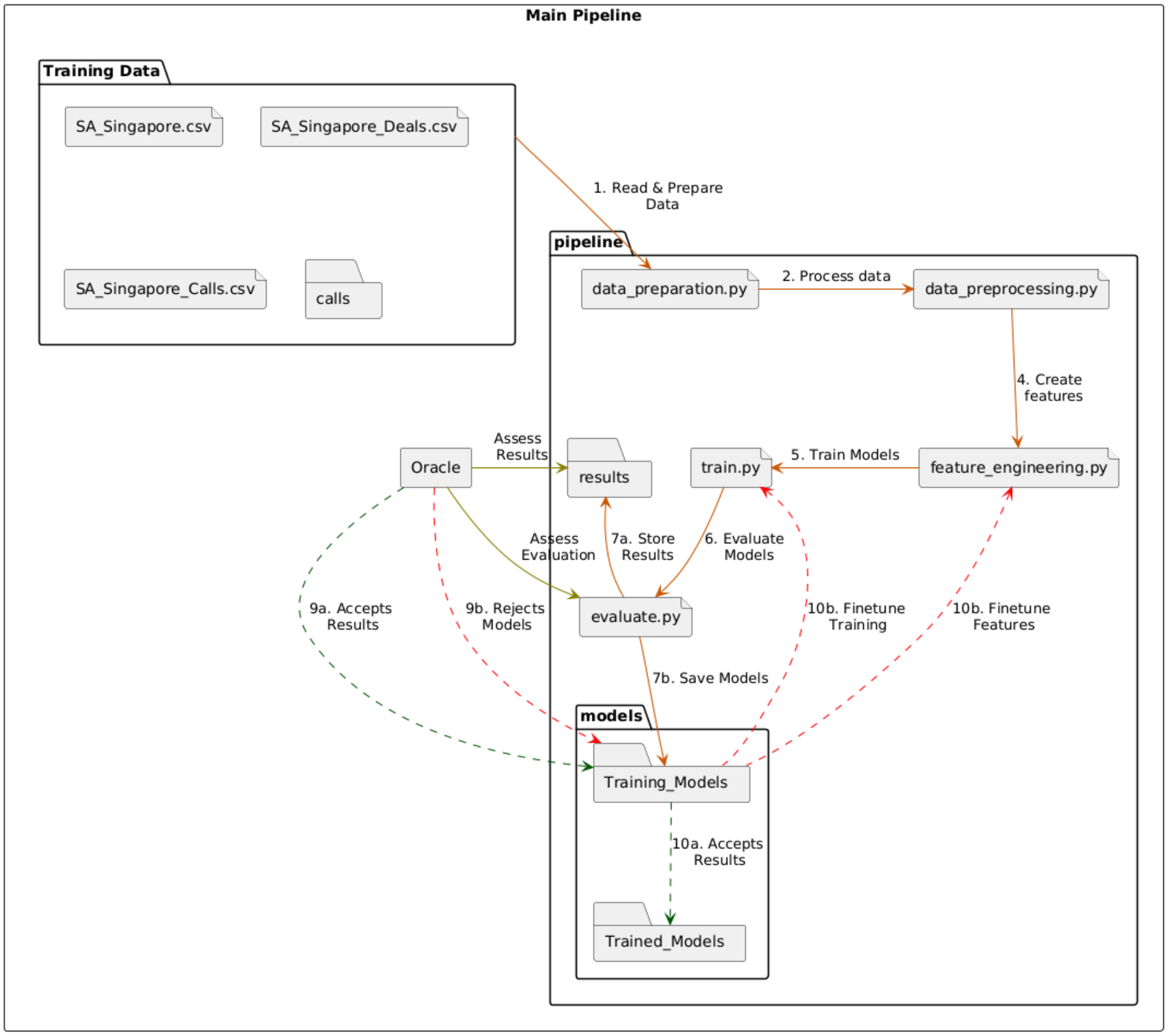



*Figure 2. Main Pipeline Design*

The main pipeline is typical of a conventional Machine Learning Training-Evaluation cycle. The figure above demonstrates data flow from the training data to the main pipeline. The entire process is orchestrated by an Oracle.[21] In my pipeline, the oracle determines if the trained model is accepted, making this a pipeline built on a supervised learning model.

[21] An oracle is a physical trainer in the AI/ML, typically an A/ML engineer. In the context of this FYP, I was the trainer

### 3.2 DBAL/DEAL Integration

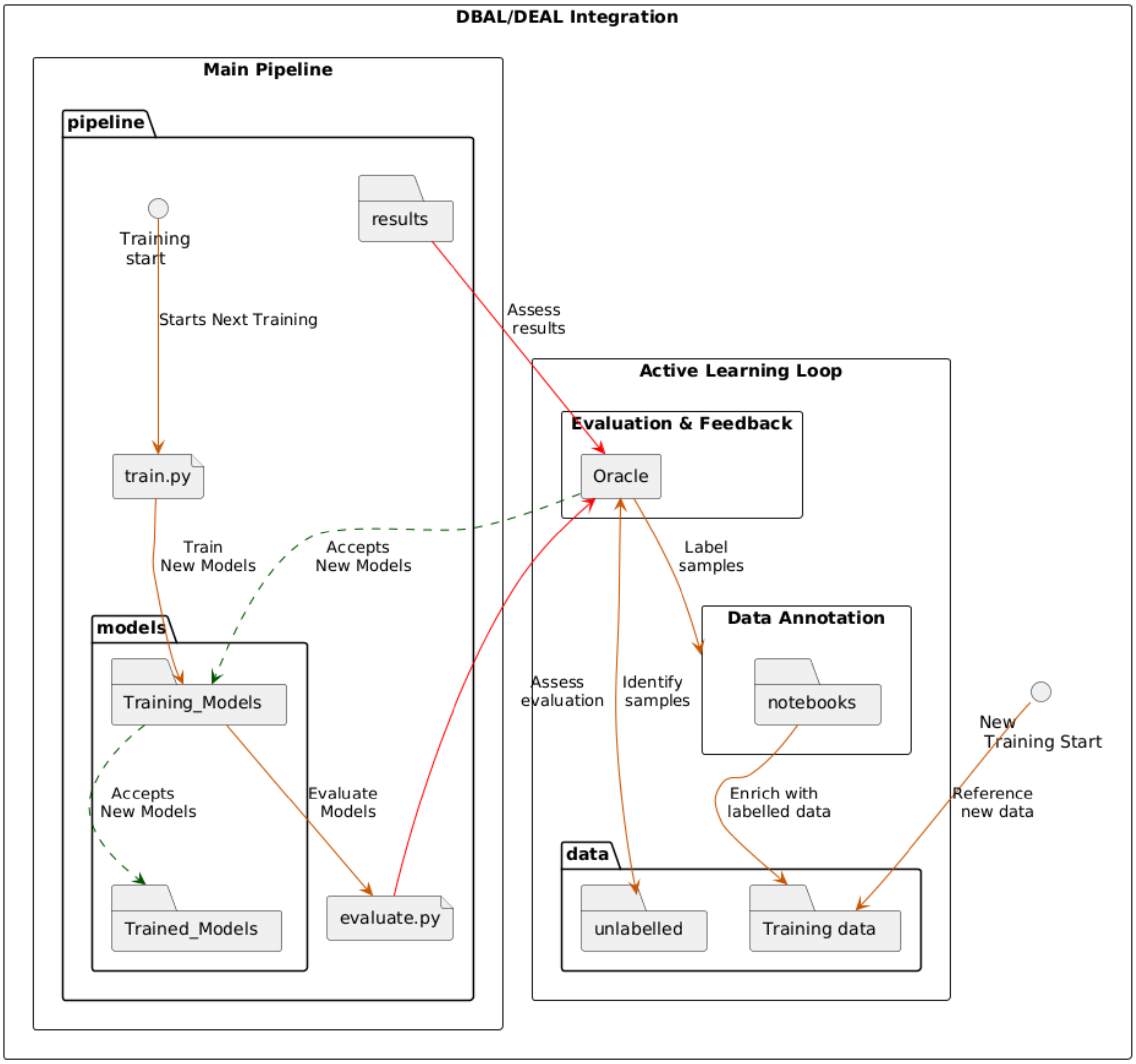



*Figure 3. Integrating Active Learning in AI/ML Pipeline*

Being a pipeline built on a supervised learning model, it integrates the considerations of active learning by design. This is done through allowing an Oracle to view the training results and model evaluations clearly for comparison. A key component here is the ability of the oracle to reject or accept a newly trained model based on these results. The `models` folder thus provides a space for this. Another key component is the readily available `data` folder which the Oracle can use to choose new unlabelled data from.

Additionally, the necessary data and data retrieval methods are available to the Oracle for consideration and use should new unlabelled data be deemed by the Oracle to be required. Storing of Old EDA[22] reports in the `notebooks` folder also allows previous work to be referenced to guide the labelling process.

### 3.3 Continuous Learning Extensibility

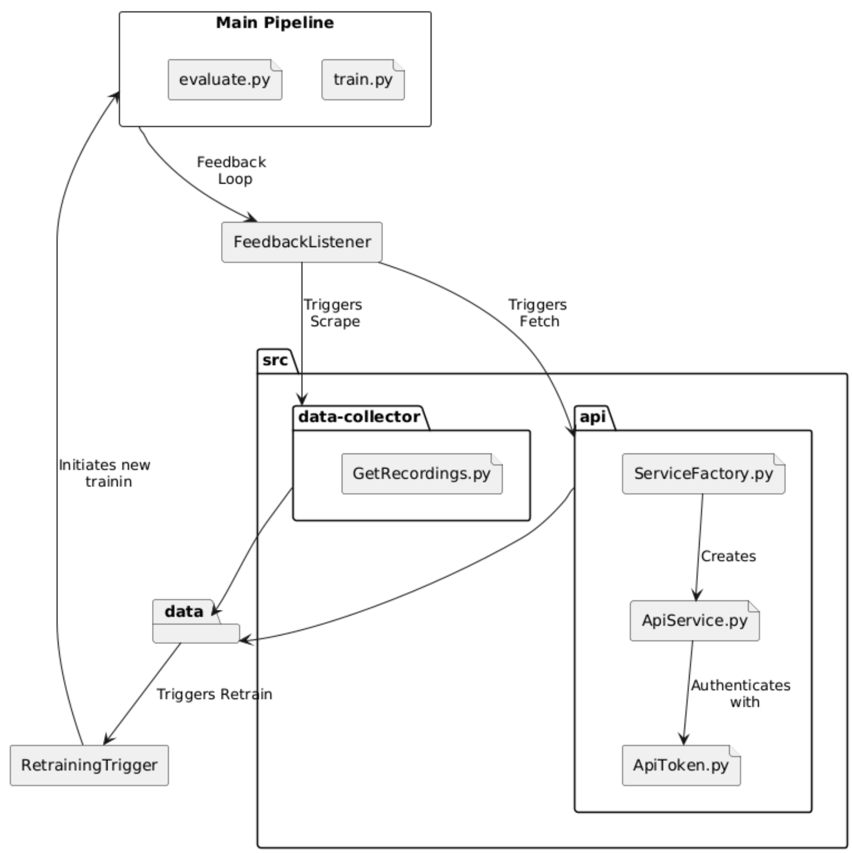



*Figure 4. Extending AI/ML Pipeline for Continuous Learning*

The existing packages found in the `src` folder are useful in extending this pipeline for continuous learning. The 2 packages `data-collector` and `api` use 2 different approaches for collecting data - web scraping with selenium and fetching data through interaction with CRM APIs. Additionally, the use of

[22] Exploratory Data Analysis

the Factory Pattern in the api folder provides an extensible framework for adding new data sources and retrieval methods.

2 potential listeners, namely the `FeedbackListener` and `RetrainingTrigger`, can fit into this pipeline design easily. They can be crafted alongside the main pipeline components or as separate components. This would then serve as a more complete AI/ML pipeline by integrating the continuous learning aspect of model training.

For purposes related to resource considerations, these 2 components were not implemented. Nonetheless, they could still be implemented in the future because of the extensibility of the pipeline's architecture.

The ground truth was extracted from sources[23] outlined in the introduction. They were loaded into a Jupyter Notebook for further analysis. Below are the notable overviews of the ground truth.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3774 entries, 0 to 3773
Data columns (total 28 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Record ID                     3774 non-null   int64
 1   Company name                  3774 non-null   object
 2   Campaign                      1261 non-null   object
 3   Ideal Customer Profile        2276 non-null   object
 4   Company Country               3774 non-null   object
 5   Last Activity Date            3390 non-null   object
 6   Industry (StaffAny Official)  3103 non-null   object
 7   Associated Contact            3492 non-null   object
 8   Deal with Primary Company     2651 non-null   object
 9   Child Company                 385 non-null    object
 10  Parent Company                471 non-null    object
 11  Associated Note               2000 non-null   object
 12  Churn Date                    319 non-null    object
 13  Likelihood to close           2617 non-null   float64
 14  Number of Associated Contacts 3768 non-null   float64
 15  First Deal Created Date       2667 non-null   object
 16  Billing Entities              525 non-null    object
 17  Last Logged Call Date         2263 non-null   object
 18  Number of times contacted     3393 non-null   float64
 19  First Contact Create Date     3079 non-null   object
 20  Is Billing Entity             528 non-null    object
 21  Create Date                   3774 non-null   object
 22  Associated Contact IDs        3492 non-null   object
 23  Deal with Primary CompanyIDs  2651 non-null   object
 24  Child CompanyIDs              385 non-null    object
 25  Parent CompanyIDs             471 non-null    float64
 26  Associated Note IDs           2000 non-null   object
 27  Billing EntitiesIDs           525 non-null    object
dtypes: float64(4), int64(1), object(23)
memory usage: 825.7+ KB
```

*Figure 5. Raw Numerical Data Snapshot*

For numerical data, the dataset has a total of 98124[24] individual entries. However, the presence of many null values coupled with the dimensionality of the data set suggested that the ground truth quality was lacklustre.

[23] Aircall, Twilio & HubSpot CRM
[24] 26 columns * 3774 Rows = 98124 entries. We do not factor the use of the Record ID & Company name column

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8326 entries, 0 to 8325
Data columns (total 16 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Record ID                8326 non-null   int64  
 1   Call Title               8324 non-null   object 
 2   Activity date            8326 non-null   object 
 3   Activity assigned to     8243 non-null   object 
 4   Call notes               6195 non-null   object 
 5   Associated Contact       6785 non-null   object 
 6   Associated Company       7975 non-null   object 
 7   Associated Deal          2465 non-null   object 
 8   Call outcome             8326 non-null   object 
 9   Recording URL            8326 non-null   object 
 10  To Number                8326 non-null   object 
 11  Call duration (HH:mm:ss) 8326 non-null   object 
 12  Voicemail Available      0 non-null      float64
 13  Associated Contact IDs   6785 non-null   object 
 14  Associated Company IDs   7977 non-null   object 
 15  Associated Deal IDs      2465 non-null   object 
dtypes: float64(1), int64(1), object(14)
memory usage: 1.0+ MB
```

*Figure 6. Raw Audio Data Snapshot*

The audio recordings, although having lesser null values, lacked the complete ground truth as the actual audio recordings were unavailable through a single export. The `data-collector` module mentioned was employed[25] to retrieve the recordings[26] to support initial data analysis work. All 8326 recordings were exported over the course of 18 hours through referencing the `Recording URL` column of the audio recordings data set. Transcriptions of these recordings were either not generated or not available for extraction.

At a glance, the ground truth required a significant amount of engineering work — from data imputation to feature engineering. I employed the use of EDAs and labelling inspired by Gal et al, (2017) [6] to analyse the data set and develop new features. Below, I highlight the most notable aspects of the EDAs and the labelling exercise.

[25] Selenium Web Scraping Tool
[26] In .mp3 format

## 4.1 Dimensionality Reduction

```
<class 'pandas.core.frame.DataFrame'>
Index: 3239 entries, 0 to 3773
Data columns (total 30 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Record ID                          3239 non-null   int64
 1   Company name                       3239 non-null   string
 2   Campaign                           3239 non-null   object
 3   Ideal Customer Profile             3239 non-null   object
 4   Industry (StaffAny Official)       3239 non-null   category
 5   Associated Contact                 3239 non-null   object
 6   Associated Note                    3239 non-null   object
 7   Likelihood to close                3239 non-null   float64
 8   Number of Associated Contacts      3239 non-null   int64
 9   Number of times contacted          3239 non-null   int64
 10  Is Billing Entity                  3239 non-null   bool
 11  Parent CompanyIDs                  3239 non-null   int64
 12  Child Company Count                3239 non-null   int64
 13  Is Parent                          3239 non-null   bool
 14  Has Parent                         3239 non-null   bool
 15  Number of Appointments             3239 non-null   int64
 16  Appointments Strength              3239 non-null   float64
 17  Has Appointments                   3239 non-null   bool
 18  Associated Note Count              3239 non-null   int64
 19  Is Lacking Note                    3239 non-null   bool
 20  Associated Note Strength           3239 non-null   float64
 21  Weighted Associated Note Strength  3239 non-null   float64
 22  Is Lacking Contact                 3239 non-null   bool
 23  Associated Contacts Strength       3239 non-null   float64
 24  Billing Entities Count             3239 non-null   int64
 25  Has Churned                        3239 non-null   bool
 26  Time Known                         3048 non-null   float64
 27  Time Index                         3048 non-null   float64
 28  Weighted Time Index                3048 non-null   float64
 29  Year                               3239 non-null   category
dtypes: bool(7), category(2), float64(8), int64(8), object(4), string(1)
memory usage: 586.1+ KB
```

*Figure 7. Cleaned Numerical Data Snapshot*

Due to the redundancy of a large number of features in my numerical ground truth, many of the initial features were removed in the data cleaning process or repurposed into more useful features with feature engineering. Additionally, the final data set also saw the removal of 535 rows of data due to contamination of the ground truth from poor data management processes at StaffAny. As a result, the initial dimensionality of 98124 data points was reduced to 90692[27] data points – an almost 7.5%[28] reduction in dimensionality[29]. This is in spite of the addition of more features to enrich the poorly labelled ground truth, a notable feat.

[27] 28 columns * 3239 Rows = 90692 entries. We do not factor the use of the Record ID & Company name column

[28] 7.57% more accurately

[29] Since one hot encoding will be applied to both the raw and cleaned data set, its effect on dimensionality was nullified and thus not considered


*Figure 8. Sample of 28 Extracted Calls*

Due to the redundancy of the audio ground truth, as well as the poor feature value[30], the initial raw audio data set was entirely replaced with the recordings extracted from the `Call Recording URL` column. The key to reducing the dimensionality of this new dataset was to apply the theoretical findings of Gal et al. (2017) [6] and Beluch et al. (2018) [2]. In particular, an important detail to

[30] It should be noted that although the raw audio snapshot provided a large number of relational mappings, poor data governance within StaffAny's business pipeline rendered these mappings useless.

note was the mere size of the initial sample used in Gal et al. (2017) [6] for the initial training cycle.

In Gal et. al (2017) [6], a mere 0.028%[31] of the entire MNIST data was randomly sampled and labelled before proceeding with the first cycle of model training. I chose to select an initial batch of 28 postprocessed call recordings, a much higher proportion of 0.34%[32] of my data set to label before proceeding with the first training cycle.

This is largely influenced by the resource constraints of my FYP as well as the differences in Gal et al. (2017) [6]..

The goal of Gal et. al (2017) [6] was to judge the efficacy of the active learning method. Random sampling is thus an important control factor in influencing the reliability of the end results. This is in stark contrast to my FYP, where the goal was to pre-train a model quickly for practical use. The exact confidence in the final results was of minute importance.[33]

Additionally, I had a lack of computational power available due to the requirement for a local implementation of the AI/ML pipeline. This meant that in theory, I had lesser iterations to fine tune my model compared to Gal et al. (2017) [6]. As sampling was done prior to pre-training, a critical decision was made to add more samples on the first training cycle to reduce the need for subsequent cycles, potentially at the cost of overfitting the model.

Nonetheless, based on the application of the methods used in Gal et al. (2017) [6], and further replicated by Beluch et al. (2018) [2], I was able to significantly reduce the dimensionality of the initial audio training data used, narrowing down the initial 8326 data points to a much smaller sample of 28.

[31] 20 of the 70000 data points available
[32] 28 of the 8326 audio recordings
[33] This is quite a common practice commercially, as the goal is typically to craft something that works over something that works *right*

## 4.2 NLP in Labelling & Sampling

Aside from common mundane preprocessing steps taken for data before model pre-training like data normalisation, accurate data conversion and data imputation, an additional but crucial step was also taken for certain data segments and the audio data. These steps heavily featured the use of NLP, showing that NLP itself as a trained model can be applied in the important but labour intensive task of labelling.

*Figure 9. Textual Data in Numerical Data Set Snapshot*

During the EDA, it was discovered that textual data was present in the numerical data set. A zero-shot binary classification text transformer was used to extract semantic meaning from the textual data and then subsequently converted into a numerical feature in the data set.

```
[ENGLISH:Do you remember?]  [MANDARIN:]
[ENGLISH:I think it's okay for now.]    [MANDARIN:I think it's okay for us r
[ENGLISH:Because we are all both very busy as an exercise.] [MANDARIN:Becaus
[ENGLISH:Very busy I can't.]    [MANDARIN:]
[ENGLISH:So cannot get a new commit.]   [MANDARIN:]
[ENGLISH:We have a new upload planning and we are.] [MANDARIN:Actually, it's
```

```
[ENGLISH:It was supposed to be 5th of March, but I think I didn't contact]
[ENGLISH:Because last week I was a bit sick.]   [MANDARIN:Because last week
[ENGLISH:Wanted to know when you do want to reschedule this or?]    [MANDAR
[ENGLISH:I think it's okay for now.]    [MANDARIN:Because we are all both v
[ENGLISH:Because we are all both very busy.]    [MANDARIN:]
[ENGLISH:Next time it's very busy.] [MANDARIN:Next time it's very busy.]
[ENGLISH:I cannot get an upload to come in.]    [MANDARIN:I cannot get an u
[ENGLISH:We have a new upload opening.] [MANDARIN:We have a new upload oper
```

```
[ENGLISH:Because last week I was a bit sick.]   [MANDARIN:I wanted t
[ENGLISH:I wanted to know when you, when you, do you want to, we sho
[ENGLISH:Because we are all both very busy.]    [MANDARIN:]
[ENGLISH:As an exercise, very busy.]    [MANDARIN:As an exercise, ve
[ENGLISH:I can't, so I cannot get, uh, how do you come in?] [MANDARI
[ENGLISH:We have a new upload, opening at Vian.]    [MANDARIN:We hav
[ENGLISH:Uh, I see.]    [MANDARIN:]
[ENGLISH:Oh, we just open, it's big.]   [MANDARIN:We just open... It
```

*Figure 10. Transcription Differences – WhisperX Tiny, Small & Base (Top to Bottom)*

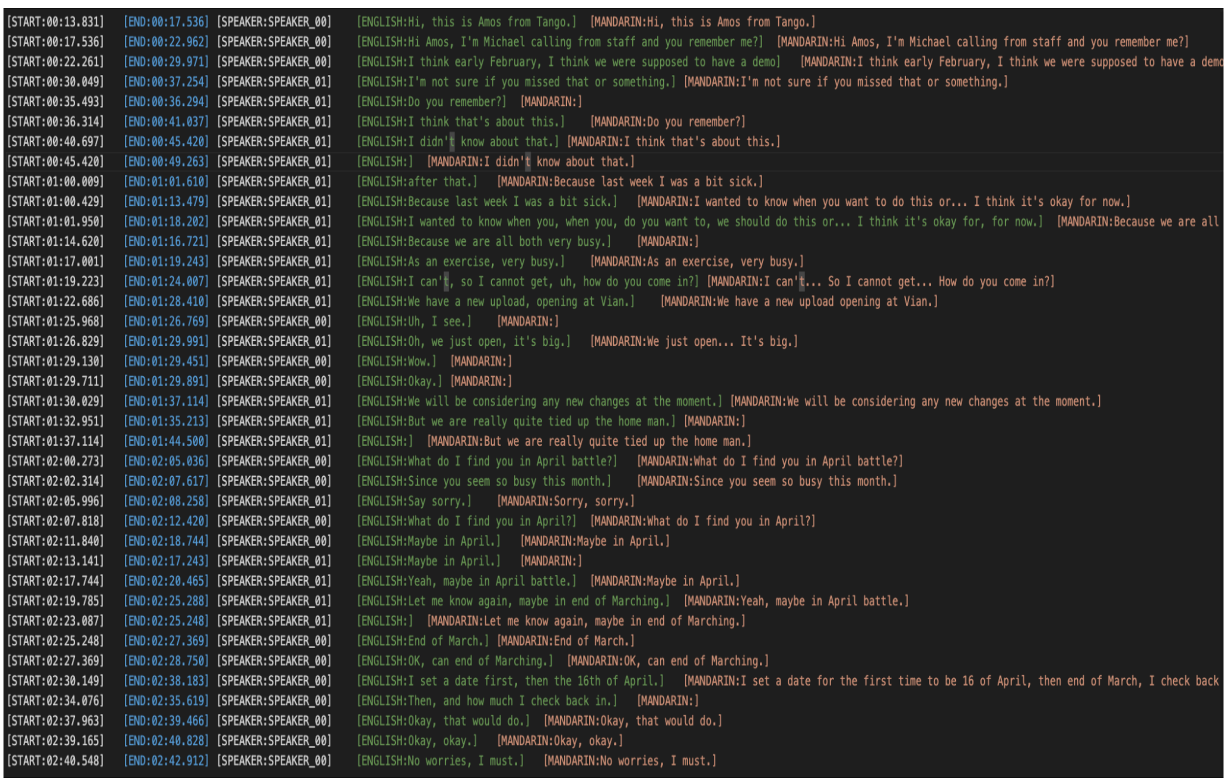
The use of NLP featured heavily in the selectively sampled 28 postprocessed call recordings. These recordings were known to lead to successful results[34] at sampling time. Thus, they were labelled[35] inherently by definition. They were selectively sampled, unlike in Gal et al. (2017) [6], Beluch et al. (2018) [2] and Lakshminarayanan et al. (2018) [9].

To compound the issue of a biased sampling method, my data set had a much lower likelihood of an accurately labelled call recording due to the poor

[34] In StaffAny's context, a successful result for a call is one that leads to an appointment being set with a potential customer
[35] The call led to an appointment being set

quality and volume of the ground truth.[36] In contrast, the MNIST data set also has a higher volume and more balanced spread, with more data points to sample from for subsequent iterations, weight decay optimisation[37] and test validation.

To combat the inherent issues in a poor ground truth, I employed the use of multiple transcription model variants to generate different transcriptions of correctly labelled audio. This had the impact of creating differing qualities of the same transcription text bounded by the correctly labelled context. This created a good foundation for preventing overfitting of the NLP model, without facing the issue of feeding the model with wrongly labelled data.



```
[START:00:13.831]  [END:00:17.536] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi, this is Amos from Tango.]  [MANDARIN:Hi, this is Amos from Tango.]
[START:00:17.536]  [END:00:22.962] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi Amos, I'm Michael calling from staff and you remember me?]  [MANDARIN:Hi Amos, I'm Michael calling from staff and you remember me?]
[START:00:22.261]  [END:00:29.971] [SPEAKER:SPEAKER_00]  [ENGLISH:I think early February, I think we were supposed to have a demo]  [MANDARIN:I think early February, I think we were supposed to have a dem
[START:00:30.049]  [END:00:37.254] [SPEAKER:SPEAKER_01]  [ENGLISH:I'm not sure if you missed that or something.] [MANDARIN:I'm not sure if you missed that or something.]
[START:00:35.493]  [END:00:36.294] [SPEAKER:SPEAKER_01]  [ENGLISH:Do you remember?]  [MANDARIN:]
[START:00:36.314]  [END:00:41.037] [SPEAKER:SPEAKER_01]  [ENGLISH:I think that's about this.]  [MANDARIN:Do you remember?]
[START:00:40.697]  [END:00:45.420] [SPEAKER:SPEAKER_01]  [ENGLISH:I didn't know about that.] [MANDARIN:I think that's about this.]
[START:00:45.420]  [END:00:49.263] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:I didn't know about that.]
[START:01:00.009]  [END:01:01.610] [SPEAKER:SPEAKER_01]  [ENGLISH:after that.]  [MANDARIN:Because last week I was a bit sick.]
[START:01:00.429]  [END:01:13.479] [SPEAKER:SPEAKER_01]  [ENGLISH:Because last week I was a bit sick.]  [MANDARIN:I wanted to know when you want to do this or... I think it's okay for now.]
[START:01:01.950]  [END:01:18.202] [SPEAKER:SPEAKER_01]  [ENGLISH:I wanted to know when you, when you, do you want to, we should do this or... I think it's okay for, for now.]  [MANDARIN:Because we are all
[START:01:14.620]  [END:01:16.721] [SPEAKER:SPEAKER_01]  [ENGLISH:Because we are all both very busy.]  [MANDARIN:]
[START:01:17.001]  [END:01:19.243] [SPEAKER:SPEAKER_01]  [ENGLISH:As an exercise, very busy.]  [MANDARIN:As an exercise, very busy.]
[START:01:19.223]  [END:01:24.007] [SPEAKER:SPEAKER_01]  [ENGLISH:I can't, so I cannot get, uh, how do you come in?] [MANDARIN:I can't... So I cannot get... How do you come in?]
[START:01:22.686]  [END:01:28.410] [SPEAKER:SPEAKER_01]  [ENGLISH:We have a new upload, opening at Vian.]  [MANDARIN:We have a new upload opening at Vian.]
[START:01:25.968]  [END:01:26.769] [SPEAKER:SPEAKER_00]  [ENGLISH:Uh, I see.]  [MANDARIN:]
[START:01:26.829]  [END:01:29.991] [SPEAKER:SPEAKER_01]  [ENGLISH:Oh, we just open, it's big.]  [MANDARIN:We just open... It's big.]
[START:01:29.130]  [END:01:29.451] [SPEAKER:SPEAKER_00]  [ENGLISH:Wow.]  [MANDARIN:]
[START:01:29.711]  [END:01:29.891] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay.] [MANDARIN:]
[START:01:30.029]  [END:01:37.114] [SPEAKER:SPEAKER_01]  [ENGLISH:We will be considering any new changes at the moment.] [MANDARIN:We will be considering any new changes at the moment.]
[START:01:32.951]  [END:01:35.213] [SPEAKER:SPEAKER_01]  [ENGLISH:But we are really quite tied up the home man.] [MANDARIN:]
[START:01:37.114]  [END:01:44.500] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:But we are really quite tied up the home man.]
[START:02:00.273]  [END:02:05.036] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April battle?]  [MANDARIN:What do I find you in April battle?]
[START:02:02.314]  [END:02:07.617] [SPEAKER:SPEAKER_00]  [ENGLISH:Since you seem so busy this month.]  [MANDARIN:Since you seem so busy this month.]
[START:02:05.996]  [END:02:08.258] [SPEAKER:SPEAKER_01]  [ENGLISH:Say sorry.]  [MANDARIN:Sorry, sorry.]
[START:02:07.818]  [END:02:12.420] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April?]  [MANDARIN:What do I find you in April?]
[START:02:11.840]  [END:02:18.744] [SPEAKER:SPEAKER_00]  [ENGLISH:Maybe in April.]  [MANDARIN:Maybe in April.]
[START:02:13.141]  [END:02:17.243] [SPEAKER:SPEAKER_01]  [ENGLISH:Maybe in April.]  [MANDARIN:]
[START:02:17.744]  [END:02:20.465] [SPEAKER:SPEAKER_01]  [ENGLISH:Yeah, maybe in April battle.]  [MANDARIN:Maybe in April.]
[START:02:19.785]  [END:02:25.288] [SPEAKER:SPEAKER_01]  [ENGLISH:Let me know again, maybe in end of Marching.]  [MANDARIN:Yeah, maybe in April battle.]
[START:02:23.087]  [END:02:25.248] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:Let me know again, maybe in end of Marching.]
[START:02:25.248]  [END:02:27.369] [SPEAKER:SPEAKER_00]  [ENGLISH:End of March.] [MANDARIN:End of March.]
[START:02:27.369]  [END:02:28.750] [SPEAKER:SPEAKER_00]  [ENGLISH:OK, can end of Marching.]  [MANDARIN:OK, can end of Marching.]
[START:02:30.149]  [END:02:38.183] [SPEAKER:SPEAKER_00]  [ENGLISH:I set a date first, then the 16th of April.]  [MANDARIN:I set a date for the first time to be 16 of April, then end of March, I check back
[START:02:34.076]  [END:02:35.619] [SPEAKER:SPEAKER_00]  [ENGLISH:Then, and how much I check back in.]  [MANDARIN:]
[START:02:37.963]  [END:02:39.466] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, that would do.]  [MANDARIN:Okay, that would do.]
[START:02:39.165]  [END:02:40.828] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, okay.]  [MANDARIN:Okay, okay.]
[START:02:40.548]  [END:02:42.912] [SPEAKER:SPEAKER_00]  [ENGLISH:No worries, I must.]  [MANDARIN:No worries, I must.]
```

*Figure 11. Sample Diarized Transcription from WhisperX*

[36] For context, only about 1% of sampled StaffAny's recorded calls were calls that led to a successful result. I personally sieved through over 3000 calls before selecting the initial data

[37] Or regularisation

The MNSIT dataset is also a numerical data set. This is in stark contrast to my data set, which consisted of audio data. The goals were also very different, with my goal being the use of the audio data to feed the processed audio data to an NLP model and pipe the prediction results to a T5 generative text meta-model. Accurately labelled data samples are much more important in my context given that the NLP model should be learning the correct conversational patterns necessary for the T5 model to generate the relevant contextual responses based on the new input conversational patterns, rather than noisy input.

Thus, additional labelling was also done on the audio data to convert the data into a suitable textual format for processing by downstream NLP models to extract semantic meaning. This involved the use of WhisperX, a transcription model, to diarize the output, directly applying a known NLP technique POS, to improve and enrich the labelling of audio data. This also had the unintentional impact of adding features to the original textual data.

Finally, optimisation for results earlier in the training cycle was also of lesser concern in Gal et. al (2017) [6] as the goal of the study was to judge the efficacy of the active learning method. This is in stark contrast to my FYP, where the goal was to pre-train a model quickly for practical use at StaffAny, although at the cost of potentially overfitting the model. Nonetheless, the process of using variance sampling in subsequent training iterations during the active learning process should serve as an appropriate mitigation for this initial overfitting attributed to the firm theoretical basis of the active learning method.

## 4.3 Feature Engineering

Feature engineering serves to enrich the current data set and is often used to derive deeper/hidden meaning inherent in the data. Undoubtedly, some of the labelling efforts presented in the previous section also serve as feature engineering efforts due to the inherent overlap in their domains, even though they serve different purposes for the training data.

```
[START:00:13.831]  [END:00:17.536] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi, this is Amos from Tango.]  [MANDARIN:Hi, this is Amos from Tango.]
[START:00:17.536]  [END:00:22.962] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi Amos, I'm Michael calling from staff and you remember me?]  [MANDARIN:Hi Amos, I'm Michael calling from staff and you remember me?]
[START:00:22.261]  [END:00:29.971] [SPEAKER:SPEAKER_00]  [ENGLISH:I think early February, I think we were supposed to have a demo]  [MANDARIN:I think early February, I think we were supposed to have a demo
[START:00:30.049]  [END:00:37.254] [SPEAKER:SPEAKER_01]  [ENGLISH:I'm not sure if you missed that or something.] [MANDARIN:I'm not sure if you missed that or something.]
[START:00:35.493]  [END:00:36.294] [SPEAKER:SPEAKER_01]  [ENGLISH:Do you remember?]  [MANDARIN:]
[START:00:36.314]  [END:00:41.037] [SPEAKER:SPEAKER_01]  [ENGLISH:I think that's about this.]  [MANDARIN:Do you remember?]
[START:00:40.697]  [END:00:45.420] [SPEAKER:SPEAKER_01]  [ENGLISH:I didn't know about that.] [MANDARIN:I think that's about this.]
[START:00:45.420]  [END:00:49.263] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:I didn't know about that.]
[START:01:00.009]  [END:01:01.610] [SPEAKER:SPEAKER_01]  [ENGLISH:after that.]  [MANDARIN:Because last week I was a bit sick.]
[START:01:00.429]  [END:01:13.479] [SPEAKER:SPEAKER_01]  [ENGLISH:Because last week I was a bit sick.]  [MANDARIN:I wanted to know when you want to do this or... I think it's okay for now.]
[START:01:01.950]  [END:01:18.202] [SPEAKER:SPEAKER_01]  [ENGLISH:I wanted to know when you, when you, do you want to, we should do this or... I think it's okay for, for now.]  [MANDARIN:Because we are all
[START:01:14.620]  [END:01:16.721] [SPEAKER:SPEAKER_01]  [ENGLISH:Because we are all both very busy.]  [MANDARIN:]
[START:01:17.001]  [END:01:19.243] [SPEAKER:SPEAKER_01]  [ENGLISH:As an exercise, very busy.]  [MANDARIN:As an exercise, very busy.]
[START:01:19.223]  [END:01:24.007] [SPEAKER:SPEAKER_01]  [ENGLISH:I can't, so I cannot get, uh, how do you come in?] [MANDARIN:I can't... So I cannot get... How do you come in?]
[START:01:22.686]  [END:01:28.410] [SPEAKER:SPEAKER_01]  [ENGLISH:We have a new upload, opening at Vian.]  [MANDARIN:We have a new upload opening at Vian.]
[START:01:25.968]  [END:01:26.769] [SPEAKER:SPEAKER_00]  [ENGLISH:Uh, I see.]  [MANDARIN:]
[START:01:26.829]  [END:01:29.991] [SPEAKER:SPEAKER_01]  [ENGLISH:Oh, we just open, it's big.]  [MANDARIN:We just open... It's big.]
[START:01:29.130]  [END:01:29.451] [SPEAKER:SPEAKER_00]  [ENGLISH:Wow.]  [MANDARIN:]
[START:01:29.711]  [END:01:29.891] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay.] [MANDARIN:]
[START:01:30.029]  [END:01:37.114] [SPEAKER:SPEAKER_01]  [ENGLISH:We will be considering any new changes at the moment.] [MANDARIN:We will be considering any new changes at the moment.]
[START:01:32.951]  [END:01:35.213] [SPEAKER:SPEAKER_01]  [ENGLISH:But we are really quite tied up the home man.] [MANDARIN:]
[START:01:37.114]  [END:01:44.500] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:But we are really quite tied up the home man.]
[START:02:00.273]  [END:02:05.036] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April battle?]  [MANDARIN:What do I find you in April battle?]
[START:02:02.314]  [END:02:07.617] [SPEAKER:SPEAKER_00]  [ENGLISH:Since you seem so busy this month.]  [MANDARIN:Since you seem so busy this month.]
[START:02:05.996]  [END:02:08.258] [SPEAKER:SPEAKER_01]  [ENGLISH:Say sorry.]  [MANDARIN:Sorry, sorry.]
[START:02:07.818]  [END:02:12.420] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April?]  [MANDARIN:What do I find you in April?]
[START:02:11.840]  [END:02:18.744] [SPEAKER:SPEAKER_00]  [ENGLISH:Maybe in April.]  [MANDARIN:Maybe in April.]
[START:02:13.141]  [END:02:17.243] [SPEAKER:SPEAKER_01]  [ENGLISH:Maybe in April.]  [MANDARIN:]
[START:02:17.744]  [END:02:20.465] [SPEAKER:SPEAKER_01]  [ENGLISH:Yeah, maybe in April battle.]  [MANDARIN:Maybe in April.]
[START:02:19.785]  [END:02:25.288] [SPEAKER:SPEAKER_01]  [ENGLISH:Let me know again, maybe in end of Marching.]  [MANDARIN:Yeah, maybe in April battle.]
[START:02:23.087]  [END:02:25.248] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:Let me know again, maybe in end of Marching.]
[START:02:25.248]  [END:02:27.369] [SPEAKER:SPEAKER_00]  [ENGLISH:End of March.] [MANDARIN:End of March.]
[START:02:27.369]  [END:02:28.750] [SPEAKER:SPEAKER_00]  [ENGLISH:OK, can end of Marching.]  [MANDARIN:OK, can end of Marching.]
[START:02:30.149]  [END:02:38.183] [SPEAKER:SPEAKER_00]  [ENGLISH:I set a date first, then the 16th of April.]  [MANDARIN:I set a date for the first time to be 16 of April, then end of March, I check back
[START:02:34.076]  [END:02:35.619] [SPEAKER:SPEAKER_00]  [ENGLISH:Then, and how much I check back in.]  [MANDARIN:]
[START:02:37.963]  [END:02:39.466] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, that would do.]  [MANDARIN:Okay, that would do.]
[START:02:39.165]  [END:02:40.828] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, okay.]  [MANDARIN:Okay, okay.]
[START:02:40.548]  [END:02:42.912] [SPEAKER:SPEAKER_00]  [ENGLISH:No worries, I must.]  [MANDARIN:No worries, I must.]
```

*Figure 12. Features Extracted with OpenSMILE*

CompareE2016 is a well-established feature set in the field of computational paralinguistics, specifically designed to extract a wide range of audio features that are crucial for understanding speech characteristics beyond just the spoken words. Compared to other feature sets, such as eGeMAPS or MFCCs, CompareE2016 offers a more extensive suite of features, including energy, spectral, and voicing-related parameters, which makes it particularly suited for analysing nuanced aspects of sales call interactions, such as emotion, stress, and engagement level. These features help in adding further context of the audio data, by enriching the downstream XLM-R and T5 model with otherwise lost features as a result of transcription. It should be noted that the features from ComparE2016 were only used when downstream NLP models considered the entirety of the transcription rather than the diarized transcription. Due to a pressing time constraint, the plan to embed features extracted from ComparE2016 directly into the transcriptions was eventually unimplemented. The opportunity to improve transcription labelling was lost as a result.

```
[START:00:13.831]  [END:00:17.536] [SPEAKER:SPEAKER_00]   [ENGLISH:Hi, this is Amos from Tango.]  [MANDARIN:Hi, this is Amos from Tango.]
[START:00:17.536]  [END:00:22.962] [SPEAKER:SPEAKER_00]   [ENGLISH:Hi Amos, I'm Michael calling from staff and you remember me?]  [MANDARIN:Hi Amos, I'm Michael calling from staff and you remember me?]
[START:00:22.261]  [END:00:29.971] [SPEAKER:SPEAKER_00]   [ENGLISH:I think early February, I think we were supposed to have a demo]   [MANDARIN:I think early February, I think we were supposed to have a demo
[START:00:30.049]  [END:00:37.254] [SPEAKER:SPEAKER_01]   [ENGLISH:I'm not sure if you missed that or something.] [MANDARIN:I'm not sure if you missed that or something.]
[START:00:35.493]  [END:00:36.294] [SPEAKER:SPEAKER_01]   [ENGLISH:Do you remember?]  [MANDARIN:]
[START:00:36.314]  [END:00:41.037] [SPEAKER:SPEAKER_01]   [ENGLISH:I think that's about this.]   [MANDARIN:Do you remember?]
[START:00:40.697]  [END:00:45.420] [SPEAKER:SPEAKER_01]   [ENGLISH:I didn't know about that.] [MANDARIN:I think that's about this.]
[START:00:45.420]  [END:00:49.263] [SPEAKER:SPEAKER_01]   [ENGLISH:]  [MANDARIN:I didn't know about that.]
[START:01:00.009]  [END:01:01.610] [SPEAKER:SPEAKER_01]   [ENGLISH:after that.]   [MANDARIN:Because last week I was a bit sick.]
[START:01:00.429]  [END:01:13.479] [SPEAKER:SPEAKER_01]   [ENGLISH:Because last week I was a bit sick.]   [MANDARIN:I wanted to know when you want to do this or... I think it's okay for now.]
[START:01:01.950]  [END:01:18.202] [SPEAKER:SPEAKER_01]   [ENGLISH:I wanted to know when you, when you, do you want to, we should do this or... I think it's okay for, for now.]  [MANDARIN:Because we are all
[START:01:14.620]  [END:01:16.721] [SPEAKER:SPEAKER_01]   [ENGLISH:Because we are all both very busy.]   [MANDARIN:]
[START:01:17.001]  [END:01:19.243] [SPEAKER:SPEAKER_01]   [ENGLISH:As an exercise, very busy.]   [MANDARIN:As an exercise, very busy.]
[START:01:19.223]  [END:01:24.007] [SPEAKER:SPEAKER_01]   [ENGLISH:I can't, so I cannot get, uh, how do you come in?] [MANDARIN:I can't... So I cannot get... How do you come in?]
[START:01:22.686]  [END:01:28.410] [SPEAKER:SPEAKER_01]   [ENGLISH:We have a new upload, opening at Vian.]   [MANDARIN:We have a new upload opening at Vian.]
[START:01:25.968]  [END:01:26.769] [SPEAKER:SPEAKER_00]   [ENGLISH:Uh, I see.]   [MANDARIN:]
[START:01:26.829]  [END:01:29.991] [SPEAKER:SPEAKER_01]   [ENGLISH:Oh, we just open, it's big.]   [MANDARIN:We just open... It's big.]
[START:01:29.130]  [END:01:29.451] [SPEAKER:SPEAKER_00]   [ENGLISH:Wow.]  [MANDARIN:]
[START:01:29.711]  [END:01:29.891] [SPEAKER:SPEAKER_00]   [ENGLISH:Okay.] [MANDARIN:]
[START:01:30.029]  [END:01:37.114] [SPEAKER:SPEAKER_01]   [ENGLISH:We will be considering any new changes at the moment.] [MANDARIN:We will be considering any new changes at the moment.]
[START:01:32.951]  [END:01:35.213] [SPEAKER:SPEAKER_01]   [ENGLISH:But we are really quite tied up the home man.] [MANDARIN:]
[START:01:37.114]  [END:01:44.500] [SPEAKER:SPEAKER_01]   [ENGLISH:]  [MANDARIN:But we are really quite tied up the home man.]
[START:02:00.273]  [END:02:05.036] [SPEAKER:SPEAKER_00]   [ENGLISH:What do I find you in April battle?]   [MANDARIN:What do I find you in April battle?]
[START:02:02.314]  [END:02:07.617] [SPEAKER:SPEAKER_00]   [ENGLISH:Since you seem so busy this month.]   [MANDARIN:Since you seem so busy this month.]
[START:02:05.996]  [END:02:08.258] [SPEAKER:SPEAKER_01]   [ENGLISH:Say sorry.]   [MANDARIN:Sorry, sorry.]
[START:02:07.818]  [END:02:12.420] [SPEAKER:SPEAKER_00]   [ENGLISH:What do I find you in April?]  [MANDARIN:What do I find you in April?]
[START:02:11.840]  [END:02:18.744] [SPEAKER:SPEAKER_00]   [ENGLISH:Maybe in April.]   [MANDARIN:Maybe in April.]
[START:02:13.141]  [END:02:17.243] [SPEAKER:SPEAKER_01]   [ENGLISH:Maybe in April.]   [MANDARIN:]
[START:02:17.744]  [END:02:20.465] [SPEAKER:SPEAKER_01]   [ENGLISH:Yeah, maybe in April battle.]  [MANDARIN:Maybe in April.]
[START:02:19.785]  [END:02:25.288] [SPEAKER:SPEAKER_01]   [ENGLISH:Let me know again, maybe in end of Marching.]  [MANDARIN:Yeah, maybe in April battle.]
[START:02:23.087]  [END:02:25.248] [SPEAKER:SPEAKER_01]   [ENGLISH:]  [MANDARIN:Let me know again, maybe in end of Marching.]
[START:02:25.248]  [END:02:27.369] [SPEAKER:SPEAKER_00]   [ENGLISH:End of March.] [MANDARIN:End of March.]
[START:02:27.369]  [END:02:28.750] [SPEAKER:SPEAKER_00]   [ENGLISH:OK, can end of Marching.]  [MANDARIN:OK, can end of Marching.]
[START:02:30.149]  [END:02:38.183] [SPEAKER:SPEAKER_00]   [ENGLISH:I set a date first, then the 16th of April.]   [MANDARIN:I set a date for the first time to be 16 of April, then end of March, I check back
[START:02:34.076]  [END:02:35.619] [SPEAKER:SPEAKER_00]   [ENGLISH:Then, and how much I check back in.]   [MANDARIN:]
[START:02:37.963]  [END:02:39.466] [SPEAKER:SPEAKER_00]   [ENGLISH:Okay, that would do.]  [MANDARIN:Okay, that would do.]
[START:02:39.165]  [END:02:40.828] [SPEAKER:SPEAKER_00]   [ENGLISH:Okay, okay.]   [MANDARIN:Okay, okay.]
[START:02:40.548]  [END:02:42.912] [SPEAKER:SPEAKER_00]   [ENGLISH:No worries, I must.]   [MANDARIN:No worries, I must.]
```

*Figure 13. Features Extracted from HubSpot API Deal Object*

Enriching the existing numerical data set involved a study of an external data set, namely the `Deal` object of HubSpot CRM. As relational mappings existed within the extracted numerical data set to the `Deal` object, it was decided through the course of the EDA that data from this data source could potentially enrich the lacklustre ground truth. This was confirmed by

StaffAny's officers. Through the employment of the `api` module, I was able to extract the necessary data in the `Deal` object to create an entirely new column `Number of Appointments`, which eventually became the basis of the y-predicate used by the numerical models for prediction.

Additionally, the use of `Deal`objects in StaffAny's context was directly complementary in my data set as it informed the data set with fresh data that did not exist in the data set prior. This is in line with best practices for feature engineering, to prevent tainting the existing data set with by using best effort approximations from within the data set itself[38], which often leads to over-optimistic model predictions and target leakage.

An exhaustive list of features are available here. [39]

[38] For instance, using the existing `Deal with Primary IDs` column to approximate the Number of Appointments over getting the accurate observations from the `Deal`object
[39] Refer to Appendix E

With the processed data, I begin training the foundation model for StaffAny's Sales Team. Below is a diagram of the Foundation Model and its components.

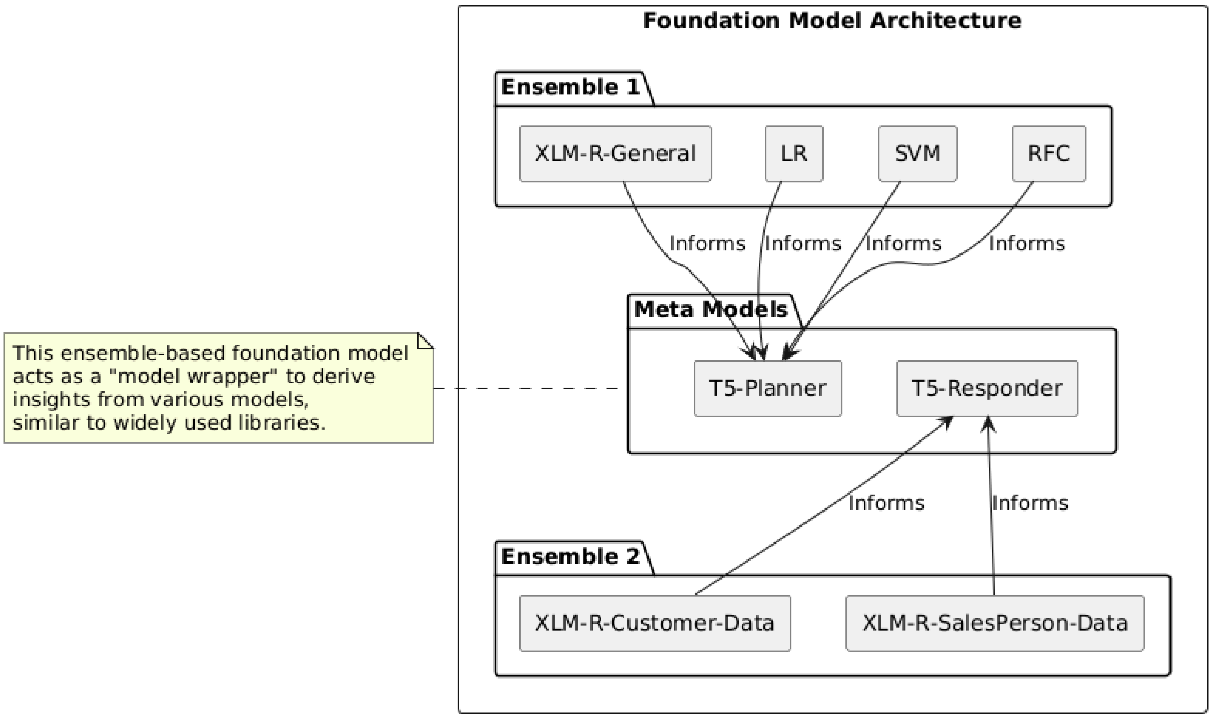



*Figure 14. Foundation Model Architecture*

The foundation model for StaffAny's sales unit was pre-trained with a large corpora of data types, classifying it as a foundation model. A unique feature of this model is the use of ensembling to derive insights from data, making this a “model wrapper" of sorts, similar to many widely available libraries in the software engineering field. The models used in this ensemble are XLM-R, T5, RFC, SCM & LR.

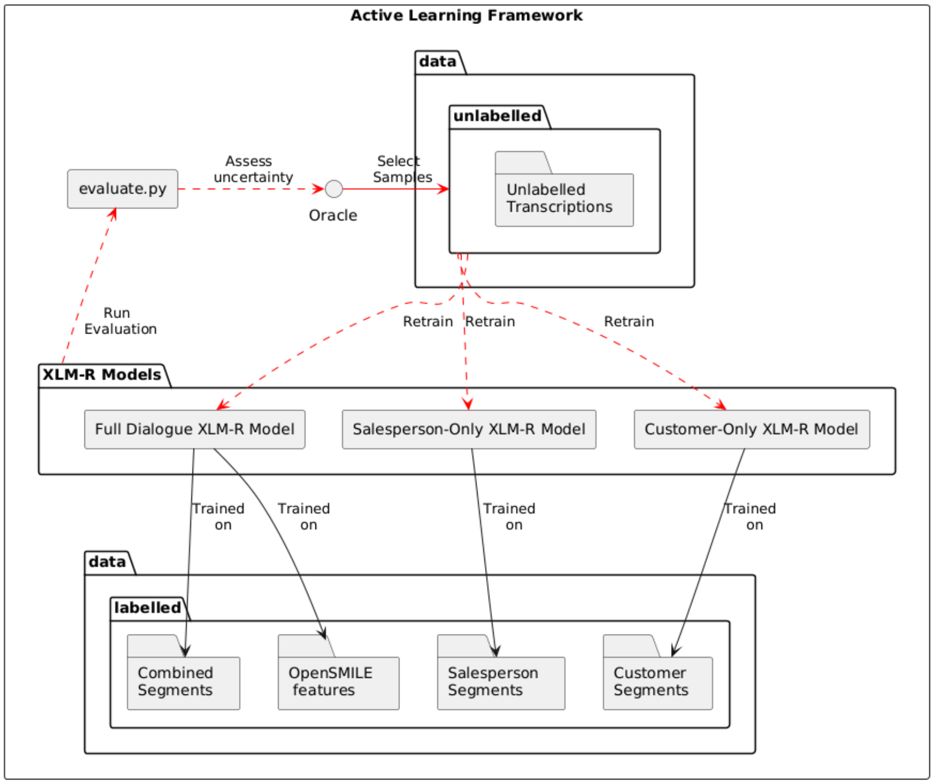



*Figure 15. Training XLMR Model*

The first XLMR model was fed customer-only interactions from the diarized Transcriptions. It focuses on deriving semantic meaning from the customer's transcriptions to gauge typical customer responses.

The second XLMR model was fed salesperson-only interactions from the diarized transcriptions. It focuses on deriving semantic meaning from how salespersons typically craft their responses in a conversational flow.

The third, and final, XLMR model was fed entire diarized transcriptions. It focuses on deriving semantic meaning from the overall dialogue flow. Its input data also included audio features derived from OpenSMILE'

Active Learning was employed utilising a Deep Ensemble in the training of these 3 models.

5.2 Numerical Data Models (RFC, SVM, LR)

Prior to training the numerical models, necessary data transformations and postprocessing steps were applied. This included applying one-hot encoding and improving dataset normality as the EDA revealed severe skew and kurtosis in the preprocessed data set

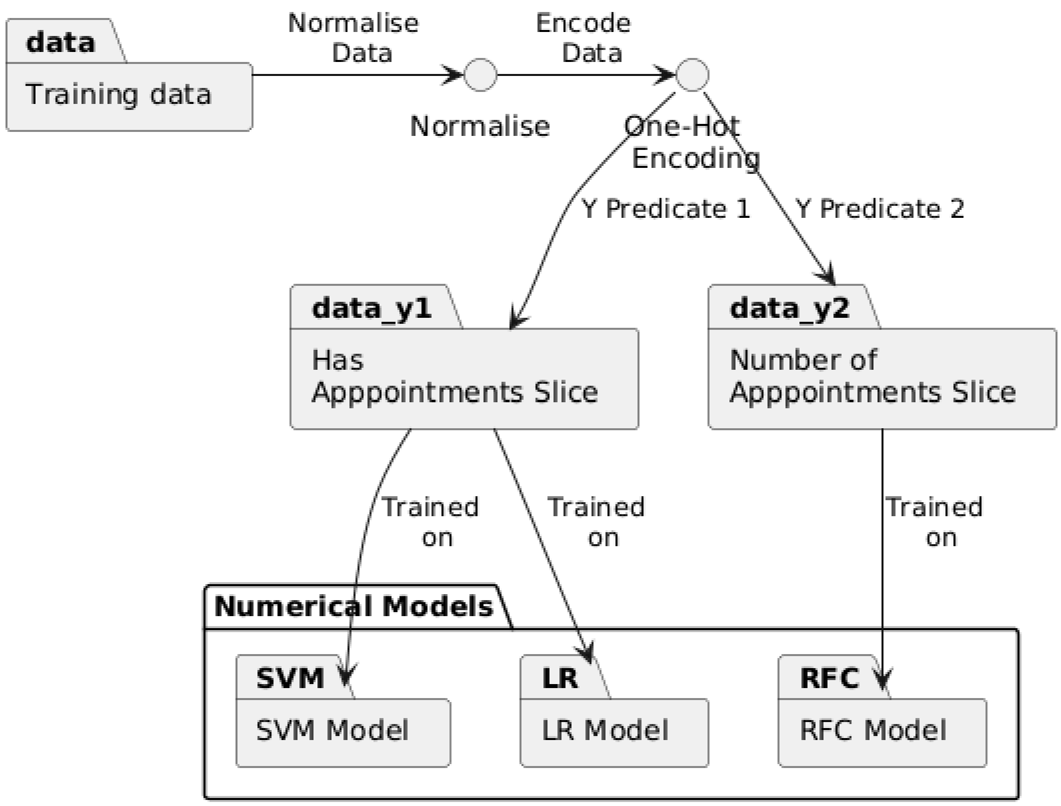



*Figure 16. Training Numerical Models*

All models were fed with data from the same data set, but with different y predicates and data segments.

SVM & LR Models were trained to predict the value of the boolean `Has Appointments` classifier in the numerical data set. Additionally, data segments where `Is Billing Entity` is True were not included in their training data. This meant that the SVM and LR models were trained to predict based on the existing data, whether a customer would have appointments set with

StaffAny, without being contaminated with the obvious relation that if a potential customer were to be a paying customer already, it would no doubt have a positive prediction in `Has Appointments`, which reduces the prediction reliability of the model.

This meant that the SVM or LR model were able to predict if a new hubspot customer would have appointments based on its current data, even before the appointment has been set. This has applications in lead scoring and lead retargeting based on what was already known in the data set.

RFC Model was trained to predict the value of the numerical `Number of Appointments` value in the numerical data set. Additionally, only a segment of the numerical data was used as it captured the context more completely. This segment was namely rows in the data where the `Is Billing Entity` is True. This meant that the RFC model learnt the number of appointments required before a customer became a client, given other data points.

5.3 T5 as a Meta-Model

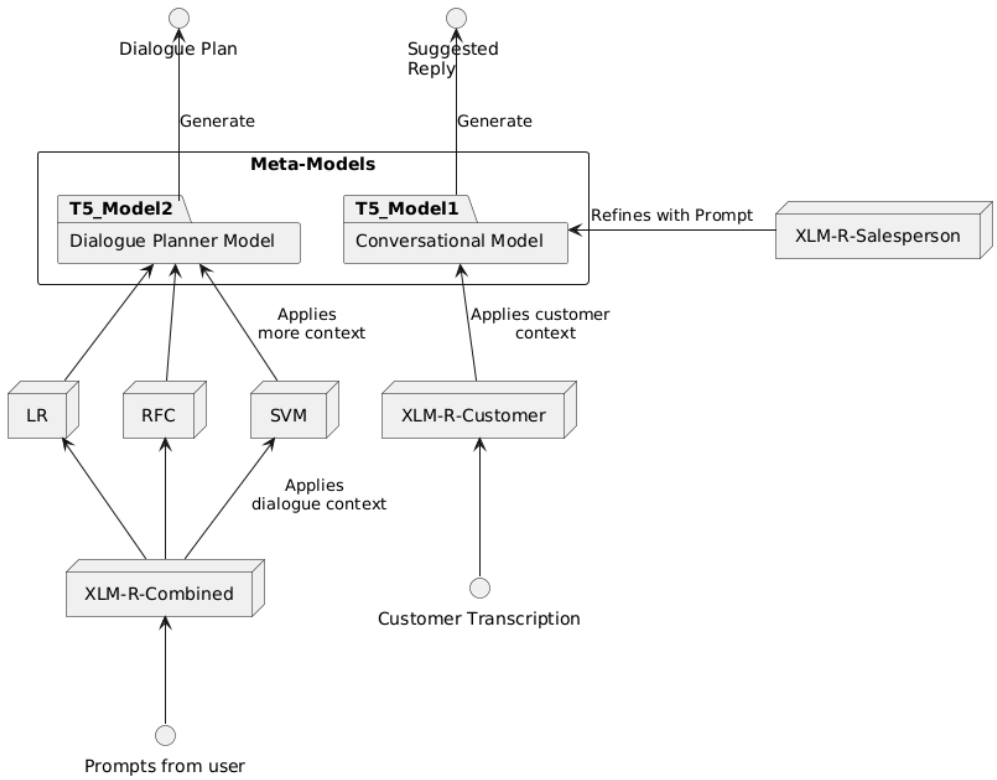



*Figure 17. T5 in combination with other models*

2 T5 models were trained.

The first T5 model was ensembled with the first and second XLMR models to understand dialogue based on the turn-based transcription input. It then generated replies based on the context fed to it from both XLMR models. The input text of this model was a customer's transcription whilst the output was what a salesperson should reply based on what the customer transcription was. This is an application of BERT's 2-way contextual understanding.

The second T5 model was ensembled with the third XLMR model to understand the entire dialogue. Prompts were fine-tuned with input from the numerical models before being tasked to generate sample dialogue plans. The output was not fine-tuned.

## 6. Implementing DEAL/DBAL

Active learning was featured heavily in the training process of the NLP models. This was apt due to the sheer dimensionality and weak labelling inherent in the resultant processed data from the audio data.

### 6.1 Ensemble Models used for Entropy

This simple active learning workflow enables efficient tracking and visualisation of the entropy-based selection process, ensuring that the labelling effort focuses on samples with the greatest potential to improve the model's understanding.

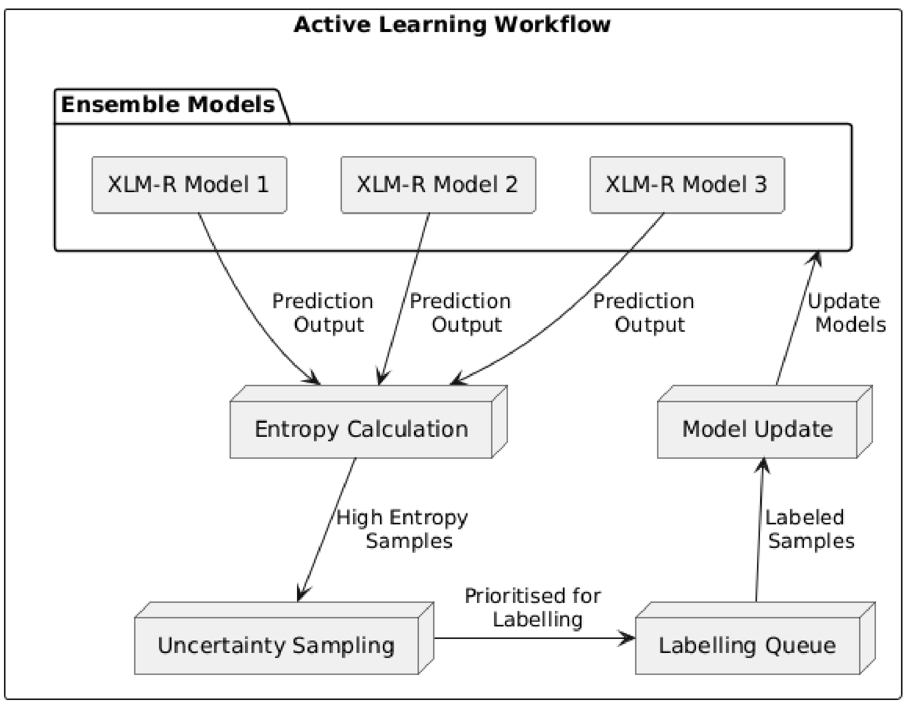



*Figure 18. Active Learning Workflow*

Ensemble models were employed as a key component of the active learning process to estimate uncertainty through entropy. By training a diverse set of XLM-R models, each specialising in different segments of the customer-sales dialogues, I constructed an ensemble capable of providing a more comprehensive assessment of model confidence in each prediction.

Entropy was calculated based on the ensemble's output, capturing the uncertainty across predictions. Higher entropy values indicated ambiguous samples, which were flagged for further inspection and labelling. This method follows the principle of uncertainty sampling, where samples that demonstrate high entropy are prioritised for labelling to maximise information gain in subsequent model updates.

6.2 Variance Sampling

In cases where it was possible, I also employed the use of Variance Sampling alongside traditional entropy-based methods as outlined by previous DBAL implementations. This was implemented as a complementary strategy to entropy-based selection to capture instances where individual model predictions in the ensemble diverged significantly.

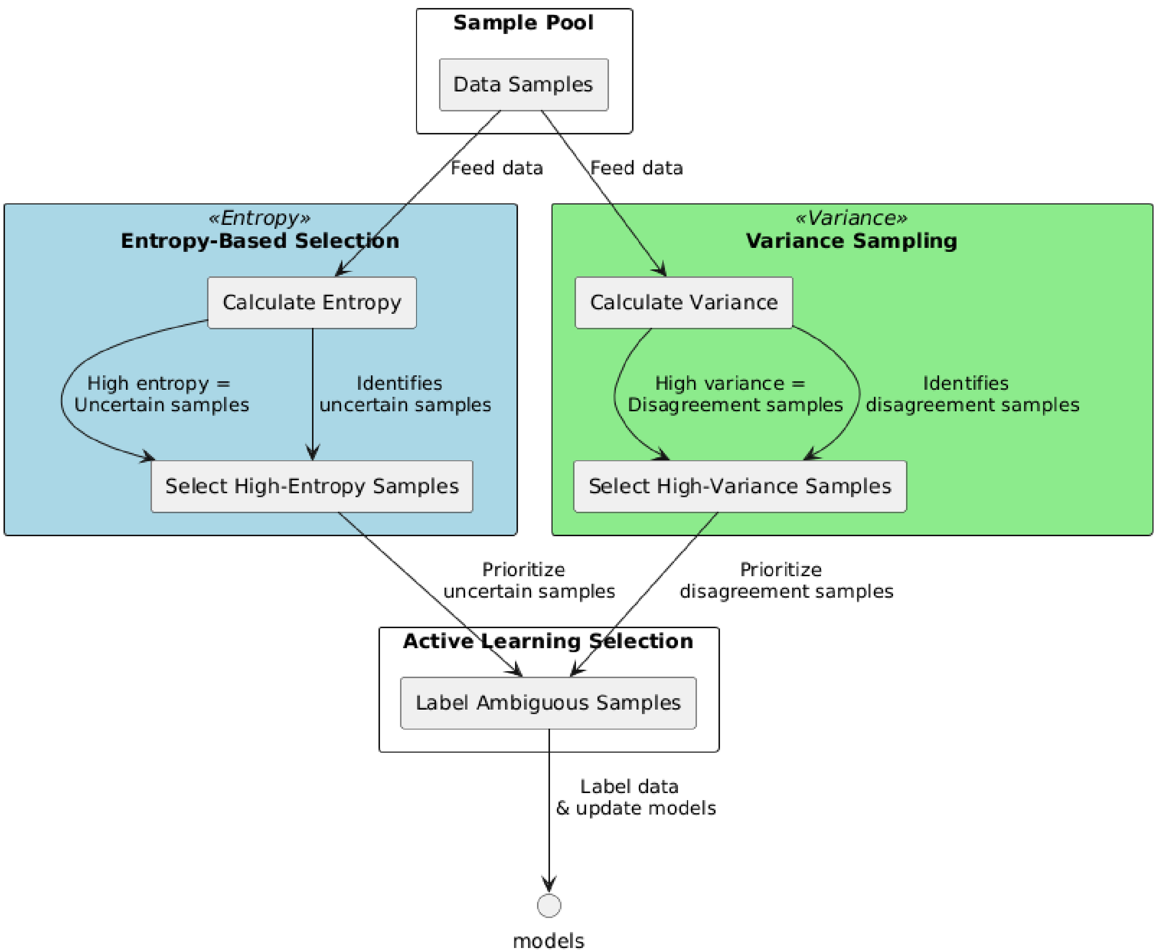
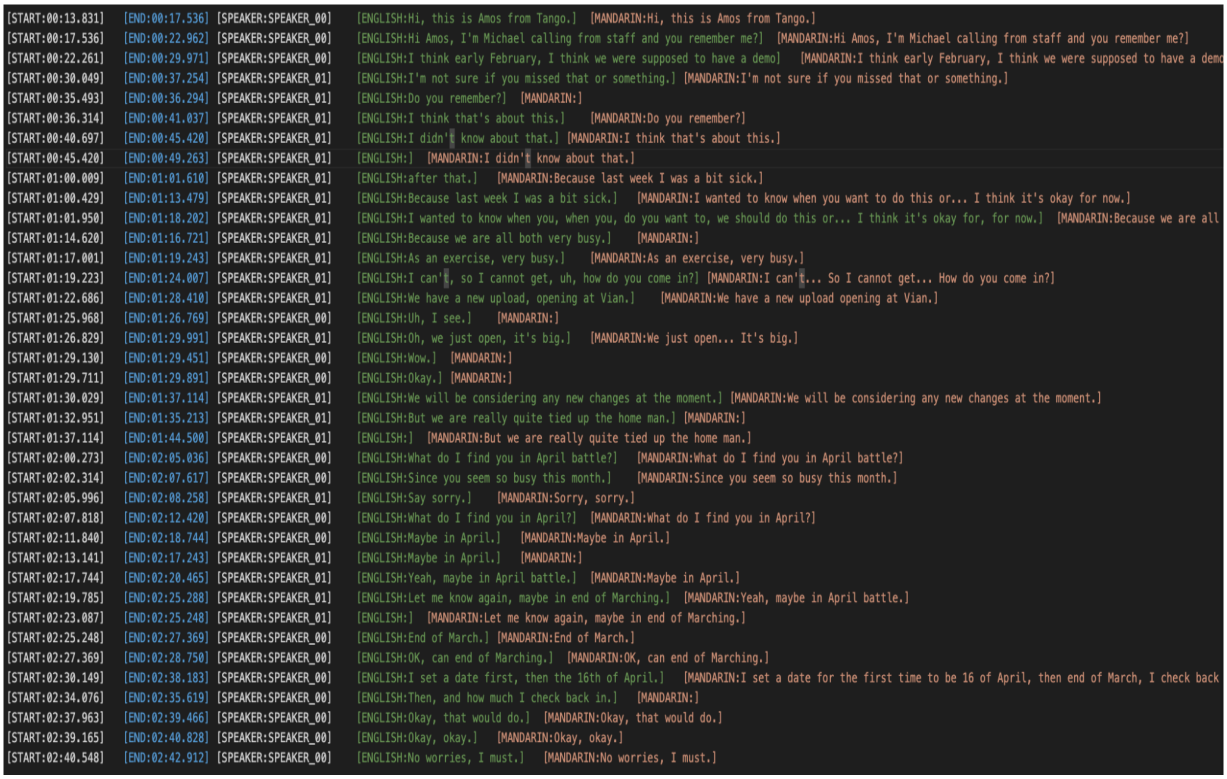



*Figure 19. Sampling Process*

Variance was calculated across the predictions of each ensemble model for a given sample, and high-variance samples were flagged for labelling. This was particularly valuable in cases where models exhibited differing interpretations of the same data point — often an indicator of ambiguous or nuanced sentiment in customer speech. This approach ensured the models received targeted training on samples where the ensemble lacked consensus.

### 6.3 Outcome of Active Learning



```
[START:00:13.831]  [END:00:17.536] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi, this is Amos from Tango.]  [MANDARIN:Hi, this is Amos from Tango.]
[START:00:17.536]  [END:00:22.962] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi Amos, I'm Michael calling from staff and you remember me?]  [MANDARIN:Hi Amos, I'm Michael calling from staff and you remember me?]
[START:00:22.261]  [END:00:29.971] [SPEAKER:SPEAKER_00]  [ENGLISH:I think early February, I think we were supposed to have a demo]  [MANDARIN:I think early February, I think we were supposed to have a demo
[START:00:30.049]  [END:00:37.254] [SPEAKER:SPEAKER_01]  [ENGLISH:I'm not sure if you missed that or something.] [MANDARIN:I'm not sure if you missed that or something.]
[START:00:35.493]  [END:00:36.294] [SPEAKER:SPEAKER_01]  [ENGLISH:Do you remember?]  [MANDARIN:]
[START:00:36.314]  [END:00:41.037] [SPEAKER:SPEAKER_01]  [ENGLISH:I think that's about this.]  [MANDARIN:Do you remember?]
[START:00:40.697]  [END:00:45.420] [SPEAKER:SPEAKER_01]  [ENGLISH:I didn't know about that.] [MANDARIN:I think that's about this.]
[START:00:45.420]  [END:00:49.263] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:I didn't know about that.]
[START:01:00.009]  [END:01:01.610] [SPEAKER:SPEAKER_01]  [ENGLISH:after that.]  [MANDARIN:Because last week I was a bit sick.]
[START:01:00.429]  [END:01:13.479] [SPEAKER:SPEAKER_01]  [ENGLISH:Because last week I was a bit sick.]  [MANDARIN:I wanted to know when you want to do this or... I think it's okay for now.]
[START:01:01.950]  [END:01:18.202] [SPEAKER:SPEAKER_01]  [ENGLISH:I wanted to know when you, when you, do you want to, we should do this or... I think it's okay for, for now.]  [MANDARIN:Because we are all
[START:01:14.620]  [END:01:16.721] [SPEAKER:SPEAKER_01]  [ENGLISH:Because we are all both very busy.]  [MANDARIN:]
[START:01:17.001]  [END:01:19.243] [SPEAKER:SPEAKER_01]  [ENGLISH:As an exercise, very busy.]  [MANDARIN:As an exercise, very busy.]
[START:01:19.223]  [END:01:24.007] [SPEAKER:SPEAKER_01]  [ENGLISH:I can't, so I cannot get, uh, how do you come in?] [MANDARIN:I can't... So I cannot get... How do you come in?]
[START:01:22.686]  [END:01:28.410] [SPEAKER:SPEAKER_01]  [ENGLISH:We have a new upload, opening at Vian.]  [MANDARIN:We have a new upload opening at Vian.]
[START:01:25.968]  [END:01:26.769] [SPEAKER:SPEAKER_00]  [ENGLISH:Uh, I see.]  [MANDARIN:]
[START:01:26.829]  [END:01:29.991] [SPEAKER:SPEAKER_01]  [ENGLISH:Oh, we just open, it's big.]  [MANDARIN:We just open... It's big.]
[START:01:29.130]  [END:01:29.451] [SPEAKER:SPEAKER_00]  [ENGLISH:Wow.]  [MANDARIN:]
[START:01:29.711]  [END:01:29.891] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay.] [MANDARIN:]
[START:01:30.029]  [END:01:37.114] [SPEAKER:SPEAKER_01]  [ENGLISH:We will be considering any new changes at the moment.] [MANDARIN:We will be considering any new changes at the moment.]
[START:01:32.951]  [END:01:35.213] [SPEAKER:SPEAKER_01]  [ENGLISH:But we are really quite tied up the home man.] [MANDARIN:]
[START:01:37.114]  [END:01:44.500] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:But we are really quite tied up the home man.]
[START:02:00.273]  [END:02:05.036] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April battle?]  [MANDARIN:What do I find you in April battle?]
[START:02:02.314]  [END:02:07.617] [SPEAKER:SPEAKER_00]  [ENGLISH:Since you seem so busy this month.]  [MANDARIN:Since you seem so busy this month.]
[START:02:05.996]  [END:02:08.258] [SPEAKER:SPEAKER_01]  [ENGLISH:Say sorry.]  [MANDARIN:Sorry, sorry.]
[START:02:07.818]  [END:02:12.420] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April?]  [MANDARIN:What do I find you in April?]
[START:02:11.840]  [END:02:18.744] [SPEAKER:SPEAKER_00]  [ENGLISH:Maybe in April.]  [MANDARIN:Maybe in April.]
[START:02:13.141]  [END:02:17.243] [SPEAKER:SPEAKER_01]  [ENGLISH:Maybe in April.]  [MANDARIN:]
[START:02:17.744]  [END:02:20.465] [SPEAKER:SPEAKER_01]  [ENGLISH:Yeah, maybe in April battle.]  [MANDARIN:Maybe in April.]
[START:02:19.785]  [END:02:25.288] [SPEAKER:SPEAKER_01]  [ENGLISH:Let me know again, maybe in end of Marching.]  [MANDARIN:Yeah, maybe in April battle.]
[START:02:23.087]  [END:02:25.248] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:Let me know again, maybe in end of Marching.]
[START:02:25.248]  [END:02:27.369] [SPEAKER:SPEAKER_00]  [ENGLISH:End of March.] [MANDARIN:End of March.]
[START:02:27.369]  [END:02:28.750] [SPEAKER:SPEAKER_00]  [ENGLISH:OK, can end of Marching.]  [MANDARIN:OK, can end of Marching.]
[START:02:30.149]  [END:02:38.183] [SPEAKER:SPEAKER_00]  [ENGLISH:I set a date first, then the 16th of April.]  [MANDARIN:I set a date for the first time to be 16 of April, then end of March, I check back
[START:02:34.076]  [END:02:35.619] [SPEAKER:SPEAKER_00]  [ENGLISH:Then, and how much I check back in.]  [MANDARIN:]
[START:02:37.963]  [END:02:39.466] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, that would do.]  [MANDARIN:Okay, that would do.]
[START:02:39.165]  [END:02:40.828] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, okay.]  [MANDARIN:Okay, okay.]
[START:02:40.548]  [END:02:42.912] [SPEAKER:SPEAKER_00]  [ENGLISH:No worries, I must.]  [MANDARIN:No worries, I must.]
```

*Figure 20. Preliminary Results of Iteration 1*

The outcome of the active learning process in iteration 1 demonstrated promising improvements in model performance. Labelling high-entropy and high-variance samples enabled the model to gain a deeper understanding of challenging samples, leading to more accurate predictions. The preliminary results indicated increased accuracy in model predictions.

Coupled with the iterative approach of active learning, DEAL ensures that each cycle refines the model's ability to capture relevant patterns in customer-sales dialogues, gradually building a more robust predictive system.

## 7. Testing & Preliminary Results

The testing methodology was limited due to time constraints. As a result, testing under realistic conditions was not accomplished. I defaulted to using conventional benchmark tests to gauge to evaluate the trained models. Additionally, a sample web application was deployed to showcase the use of the Foundation Model in StaffAny's business operations.

### 7.1 Evaluating the Numerical models

```
[START:00:13.831]   [END:00:17.536] [SPEAKER:SPEAKER_00]   [ENGLISH:Hi, this is Amos from Tango.]  [MANDARIN:Hi, this is Amos from Tango.]
[START:00:17.536]   [END:00:22.962] [SPEAKER:SPEAKER_00]   [ENGLISH:Hi Amos, I'm Michael calling from staff and you remember me?]  [MANDARIN:Hi Amos, I'm Michael calling from staff and you remember me?]
[START:00:22.261]   [END:00:29.971] [SPEAKER:SPEAKER_00]   [ENGLISH:I think early February, I think we were supposed to have a demo]   [MANDARIN:I think early February, I think we were supposed to have a demo
[START:00:30.049]   [END:00:37.254] [SPEAKER:SPEAKER_01]   [ENGLISH:I'm not sure if you missed that or something.] [MANDARIN:I'm not sure if you missed that or something.]
[START:00:35.493]   [END:00:36.294] [SPEAKER:SPEAKER_01]   [ENGLISH:Do you remember?]  [MANDARIN:]
[START:00:36.314]   [END:00:41.037] [SPEAKER:SPEAKER_01]   [ENGLISH:I think that's about this.]    [MANDARIN:Do you remember?]
[START:00:40.697]   [END:00:45.420] [SPEAKER:SPEAKER_01]   [ENGLISH:I didn't know about that.] [MANDARIN:I think that's about this.]
[START:00:45.420]   [END:00:49.263] [SPEAKER:SPEAKER_01]   [ENGLISH:]  [MANDARIN:I didn't know about that.]
[START:01:00.009]   [END:01:01.610] [SPEAKER:SPEAKER_01]   [ENGLISH:after that.]   [MANDARIN:Because last week I was a bit sick.]
[START:01:00.429]   [END:01:13.479] [SPEAKER:SPEAKER_01]   [ENGLISH:Because last week I was a bit sick.]   [MANDARIN:I wanted to know when you want to do this or... I think it's okay for now.]
[START:01:01.950]   [END:01:18.202] [SPEAKER:SPEAKER_01]   [ENGLISH:I wanted to know when you, when you, do you want to, we should do this or... I think it's okay for, for now.]  [MANDARIN:Because we are all
[START:01:14.620]   [END:01:16.721] [SPEAKER:SPEAKER_01]   [ENGLISH:Because we are all both very busy.]    [MANDARIN:]
[START:01:17.001]   [END:01:19.243] [SPEAKER:SPEAKER_01]   [ENGLISH:As an exercise, very busy.]    [MANDARIN:As an exercise, very busy.]
[START:01:19.223]   [END:01:24.007] [SPEAKER:SPEAKER_01]   [ENGLISH:I can't, so I cannot get, uh, how do you come in?] [MANDARIN:I can't... So I cannot get... How do you come in?]
[START:01:22.686]   [END:01:28.410] [SPEAKER:SPEAKER_01]   [ENGLISH:We have a new upload, opening at Vian.]    [MANDARIN:We have a new upload opening at Vian.]
[START:01:25.968]   [END:01:26.769] [SPEAKER:SPEAKER_00]   [ENGLISH:Uh, I see.]    [MANDARIN:]
[START:01:26.829]   [END:01:29.991] [SPEAKER:SPEAKER_01]   [ENGLISH:Oh, we just open, it's big.]   [MANDARIN:We just open... It's big.]
[START:01:29.130]   [END:01:29.451] [SPEAKER:SPEAKER_00]   [ENGLISH:Wow.]  [MANDARIN:]
[START:01:29.711]   [END:01:29.891] [SPEAKER:SPEAKER_00]   [ENGLISH:Okay.] [MANDARIN:]
[START:01:30.029]   [END:01:37.114] [SPEAKER:SPEAKER_01]   [ENGLISH:We will be considering any new changes at the moment.] [MANDARIN:We will be considering any new changes at the moment.]
[START:01:32.951]   [END:01:35.213] [SPEAKER:SPEAKER_01]   [ENGLISH:But we are really quite tied up the home man.] [MANDARIN:]
[START:01:37.114]   [END:01:44.500] [SPEAKER:SPEAKER_01]   [ENGLISH:]  [MANDARIN:But we are really quite tied up the home man.]
[START:02:00.273]   [END:02:05.036] [SPEAKER:SPEAKER_00]   [ENGLISH:What do I find you in April battle?]   [MANDARIN:What do I find you in April battle?]
[START:02:02.314]   [END:02:07.617] [SPEAKER:SPEAKER_00]   [ENGLISH:Since you seem so busy this month.]    [MANDARIN:Since you seem so busy this month.]
[START:02:05.996]   [END:02:08.258] [SPEAKER:SPEAKER_01]   [ENGLISH:Say sorry.]    [MANDARIN:Sorry, sorry.]
[START:02:07.818]   [END:02:12.420] [SPEAKER:SPEAKER_00]   [ENGLISH:What do I find you in April?]  [MANDARIN:What do I find you in April?]
[START:02:11.840]   [END:02:18.744] [SPEAKER:SPEAKER_00]   [ENGLISH:Maybe in April.]   [MANDARIN:Maybe in April.]
[START:02:13.141]   [END:02:17.243] [SPEAKER:SPEAKER_01]   [ENGLISH:Maybe in April.]   [MANDARIN:]
[START:02:17.744]   [END:02:20.465] [SPEAKER:SPEAKER_01]   [ENGLISH:Yeah, maybe in April battle.]  [MANDARIN:Maybe in April.]
[START:02:19.785]   [END:02:25.288] [SPEAKER:SPEAKER_01]   [ENGLISH:Let me know again, maybe in end of Marching.]  [MANDARIN:Yeah, maybe in April battle.]
[START:02:23.087]   [END:02:25.248] [SPEAKER:SPEAKER_01]   [ENGLISH:]  [MANDARIN:Let me know again, maybe in end of Marching.]
[START:02:25.248]   [END:02:27.369] [SPEAKER:SPEAKER_00]   [ENGLISH:End of March.] [MANDARIN:End of March.]
[START:02:27.369]   [END:02:28.750] [SPEAKER:SPEAKER_00]   [ENGLISH:OK, can end of Marching.]  [MANDARIN:OK, can end of Marching.]
[START:02:30.149]   [END:02:38.183] [SPEAKER:SPEAKER_00]   [ENGLISH:I set a date first, then the 16th of April.]   [MANDARIN:I set a date for the first time to be 16 of April, then end of March, I check back
[START:02:34.076]   [END:02:35.619] [SPEAKER:SPEAKER_00]   [ENGLISH:Then, and how much I check back in.]   [MANDARIN:]
[START:02:37.963]   [END:02:39.466] [SPEAKER:SPEAKER_00]   [ENGLISH:Okay, that would do.]  [MANDARIN:Okay, that would do.]
[START:02:39.165]   [END:02:40.828] [SPEAKER:SPEAKER_00]   [ENGLISH:Okay, okay.]   [MANDARIN:Okay, okay.]
[START:02:40.548]   [END:02:42.912] [SPEAKER:SPEAKER_00]   [ENGLISH:No worries, I must.]   [MANDARIN:No worries, I must.]
```

*Figure 21. Preliminary Results of Numerical Models*

The evaluation metrics and criteria for testing the foundation model focused on understanding its performance in terms of accuracy, reliability, and effectiveness in a sales environment. Key metrics used included accuracy, precision, recall, F1-score, and ROC-AUC, which are standard measures for classification models.

## 7.2 Evaluating the T5 Model

```
[START:00:13.831]  [END:00:17.536] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi, this is Amos from Tango.]  [MANDARIN:Hi, this is Amos from Tango.]
[START:00:17.536]  [END:00:22.962] [SPEAKER:SPEAKER_00]  [ENGLISH:Hi Amos, I'm Michael calling from staff and you remember me?]  [MANDARIN:Hi Amos, I'm Michael calling from staff and you remember me?]
[START:00:22.261]  [END:00:29.971] [SPEAKER:SPEAKER_00]  [ENGLISH:I think early February, I think we were supposed to have a demo]  [MANDARIN:I think early February, I think we were supposed to have a demo
[START:00:30.049]  [END:00:37.254] [SPEAKER:SPEAKER_01]  [ENGLISH:I'm not sure if you missed that or something.] [MANDARIN:I'm not sure if you missed that or something.]
[START:00:35.493]  [END:00:36.294] [SPEAKER:SPEAKER_01]  [ENGLISH:Do you remember?]  [MANDARIN:]
[START:00:36.314]  [END:00:41.037] [SPEAKER:SPEAKER_01]  [ENGLISH:I think that's about this.]  [MANDARIN:Do you remember?]
[START:00:40.697]  [END:00:45.420] [SPEAKER:SPEAKER_01]  [ENGLISH:I didn't know about that.] [MANDARIN:I think that's about this.]
[START:00:45.420]  [END:00:49.263] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:I didn't know about that.]
[START:01:00.009]  [END:01:01.610] [SPEAKER:SPEAKER_01]  [ENGLISH:after that.]  [MANDARIN:Because last week I was a bit sick.]
[START:01:00.429]  [END:01:13.479] [SPEAKER:SPEAKER_01]  [ENGLISH:Because last week I was a bit sick.]  [MANDARIN:I wanted to know when you want to do this or... I think it's okay for now.]
[START:01:01.950]  [END:01:18.202] [SPEAKER:SPEAKER_01]  [ENGLISH:I wanted to know when you, when you, do you want to, we should do this or... I think it's okay for, for now.]  [MANDARIN:Because we are all
[START:01:14.620]  [END:01:16.721] [SPEAKER:SPEAKER_01]  [ENGLISH:Because we are all both very busy.]  [MANDARIN:]
[START:01:17.001]  [END:01:19.243] [SPEAKER:SPEAKER_01]  [ENGLISH:As an exercise, very busy.]  [MANDARIN:As an exercise, very busy.]
[START:01:19.223]  [END:01:24.007] [SPEAKER:SPEAKER_01]  [ENGLISH:I can't, so I cannot get, uh, how do you come in?] [MANDARIN:I can't... So I cannot get... How do you come in?]
[START:01:22.686]  [END:01:28.410] [SPEAKER:SPEAKER_01]  [ENGLISH:We have a new upload, opening at Vian.]  [MANDARIN:We have a new upload opening at Vian.]
[START:01:25.968]  [END:01:26.769] [SPEAKER:SPEAKER_00]  [ENGLISH:Uh, I see.]  [MANDARIN:]
[START:01:26.829]  [END:01:29.991] [SPEAKER:SPEAKER_01]  [ENGLISH:Oh, we just open, it's big.]  [MANDARIN:We just open... It's big.]
[START:01:29.130]  [END:01:29.451] [SPEAKER:SPEAKER_00]  [ENGLISH:Wow.]  [MANDARIN:]
[START:01:29.711]  [END:01:29.891] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay.] [MANDARIN:]
[START:01:30.029]  [END:01:37.114] [SPEAKER:SPEAKER_01]  [ENGLISH:We will be considering any new changes at the moment.] [MANDARIN:We will be considering any new changes at the moment.]
[START:01:32.951]  [END:01:35.213] [SPEAKER:SPEAKER_01]  [ENGLISH:But we are really quite tied up the home man.] [MANDARIN:]
[START:01:37.114]  [END:01:44.500] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:But we are really quite tied up the home man.]
[START:02:00.273]  [END:02:05.036] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April battle?]  [MANDARIN:What do I find you in April battle?]
[START:02:02.314]  [END:02:07.617] [SPEAKER:SPEAKER_00]  [ENGLISH:Since you seem so busy this month.]  [MANDARIN:Since you seem so busy this month.]
[START:02:05.996]  [END:02:08.258] [SPEAKER:SPEAKER_01]  [ENGLISH:Say sorry.]  [MANDARIN:Sorry, sorry.]
[START:02:07.818]  [END:02:12.420] [SPEAKER:SPEAKER_00]  [ENGLISH:What do I find you in April?]  [MANDARIN:What do I find you in April?]
[START:02:11.840]  [END:02:18.744] [SPEAKER:SPEAKER_00]  [ENGLISH:Maybe in April.]  [MANDARIN:Maybe in April.]
[START:02:13.141]  [END:02:17.243] [SPEAKER:SPEAKER_01]  [ENGLISH:Maybe in April.]  [MANDARIN:]
[START:02:17.744]  [END:02:20.465] [SPEAKER:SPEAKER_01]  [ENGLISH:Yeah, maybe in April battle.]  [MANDARIN:Maybe in April.]
[START:02:19.785]  [END:02:25.288] [SPEAKER:SPEAKER_01]  [ENGLISH:Let me know again, maybe in end of Marching.]  [MANDARIN:Yeah, maybe in April battle.]
[START:02:23.087]  [END:02:25.248] [SPEAKER:SPEAKER_01]  [ENGLISH:]  [MANDARIN:Let me know again, maybe in end of Marching.]
[START:02:25.248]  [END:02:27.369] [SPEAKER:SPEAKER_00]  [ENGLISH:End of March.] [MANDARIN:End of March.]
[START:02:27.369]  [END:02:28.750] [SPEAKER:SPEAKER_00]  [ENGLISH:OK, can end of Marching.]  [MANDARIN:OK, can end of Marching.]
[START:02:30.149]  [END:02:38.183] [SPEAKER:SPEAKER_00]  [ENGLISH:I set a date first, then the 16th of April.]  [MANDARIN:I set a date for the first time to be 16 of April, then end of March, I check back
[START:02:34.076]  [END:02:35.619] [SPEAKER:SPEAKER_00]  [ENGLISH:Then, and how much I check back in.]  [MANDARIN:]
[START:02:37.963]  [END:02:39.466] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, that would do.]  [MANDARIN:Okay, that would do.]
[START:02:39.165]  [END:02:40.828] [SPEAKER:SPEAKER_00]  [ENGLISH:Okay, okay.]  [MANDARIN:Okay, okay.]
[START:02:40.548]  [END:02:42.912] [SPEAKER:SPEAKER_00]  [ENGLISH:No worries, I must.]  [MANDARIN:No worries, I must.]
```

*Figure 22. Using the T5 Model in the Streamlit Web Application*

While no formal testing of the meta-model was conducted, I was able to test the generative capabilities with a simple web application. Although this may not qualify as a rigorous test benchmark, it shows promise in the use of the meta-model in a simulated context and supports the potential of the model.

## 7.3 Test Results

The outcome of these preliminary evaluations are promising. At the time of writing, no formal tests were conducted on all models. However, implementation of the model's use in the context of a web application, coupled with evaluation metrics from the numerical models, show promise that the model has the potential to be tested even more thoroughly and under real-world scenarios.

## 8. Discussion

As stated, due to time constraints, the resulting model was only tested minimally. As such the focus of this section would be on detailing the challenges faced and critiquing the overall quality of my FYP.

### 8.1 Noisy & Low Quality Data

Throughout development and testing, handling noisy data emerged as a major challenge. The quality of audio recordings varied widely, with some samples suffering from substantial background noise that complicated transcription and analysis. Although denoising techniques were applied to improve Whisper transcription outputs, their effectiveness was limited by the time and computational resources available. While these methods helped marginally improve transcription quality, they did not completely eliminate noise issues, leading to residual inaccuracies that impacted downstream analysis. Ideally, advanced deep learning-based denoising models (e.g., Wave-U-Net or Speech Enhancement GANs) could have been implemented, but resource constraints prevented their integration.

Additionally, the variability in audio quality led to inconsistencies in transcription accuracy, affecting feature extraction and model performance. To address this, two distinct XLM-R models were used, one specialized for lower-quality audio and the other for higher-quality recordings. However, this dual-model approach introduced additional computational overhead and increased model complexity. A more efficient solution, such as a unified model with adaptive training for diverse audio qualities, might have reduced resource strain, but developing and tuning such a model was infeasible given the project's resource and time constraints.

8.2 Data Dimensionality

The large dataset posed significant challenges for efficient data processing and management. Although batch processing was implemented to manage data in chunks, this solution was not fully optimised for handling larger-than-memory datasets, resulting in processing delays. While workspaces were utilised to persist intermediate data, minimising redundant computations, they required frequent input-output operations that impacted overall performance. A more sophisticated solution, such as distributed processing across multiple nodes or parallelized data handling, could have improved efficiency. However, the limitations in computational resources restricted me to a single-machine setup, hindering the ability to fully leverage these approaches.

The high dimensionality of extracted features from audio and numerical data further complicated the process, making it challenging to maintain manageable model complexity. While Principal Component Analysis (PCA) and basic feature selection were applied to reduce feature space, these methods may have led to information loss, potentially impacting model performance. Advanced dimensionality reduction techniques like autoencoders or t-SNE were considered but ultimately dismissed due to the lack of resources needed to fine-tune and validate these methods. This constraint limited the ability to explore the full potential of dimensionality reduction, resulting in models that could have been more interpretable and less computationally expensive if a more rigorous feature selection approach had been possible.

## 8.3 Overfitting

Overfitting presented a considerable challenge in the modelling process, largely due to deviations from a more rigorous uncertainty-based methodology outlined by Gal et al., which emphasises model robustness and generalizability over raw predictive power. In an effort to maximise immediate performance metrics, I prioritised predictive accuracy and feature extraction capacity, but this approach came at the expense of model stability. By not adhering strictly to Gal et al.'s (2018) [6] Bayesian approximation techniques, the models become more prone to overfitting, especially in high-dimensional feature spaces.

Resource limitations also played a role in this decision. Techniques proposed by Gal et al., such as Monte Carlo Dropout, demand significant computational power due to repeated forward passes during inference to estimate uncertainty. Given the constraints in available hardware and processing time, fully implementing this methodology proved impractical. Consequently, simpler regularisation techniques were employed, including basic dropout and early stopping, which partially mitigated overfitting but did not offer the robustness of Gal et al.'s uncertainty-driven approach.

This decision to optimise for predictive power resulted in models that performed well on training data but exhibited reduced generalizability to unseen data, especially in cases where the underlying data distributions varied. A stricter adherence to uncertainty-based methods would likely have fostered more robust models capable of capturing epistemic uncertainty, but such rigour was compromised due to practical limitations in computational resources and the focus on immediate predictive gains. A more balanced approach that accounts for both predictive power and model uncertainty would have likely yielded better and more reliable results.

# 9. Conclusion

This FYP provides a preliminary evaluation of the results achieved through the implementation of transcription, feature extraction, and classification models within a sales-support context. The developed foundation model aims to improve predictive accuracy and assist sales personnel in setting appointments through machine learning techniques. Despite various advancements, this report recognizes both the strengths and limitations of the model in addressing the unique challenges of real-world sales environments. While the foundation model demonstrates strong potential for assisting in appointment setting within sales contexts, further refinements are necessary to fully overcome its current limitations. Future research should aim to balance predictive power with real-time applicability and bias mitigation, ensuring that the model achieves robustness and generalizability in diverse sales environments.

## 9.1 Summary of Work Done

The project centred on implementing an ensemble of models designed to capture and analyse both customer and sales representative dialogues. These models included components for transcription, feature extraction, and classification, ultimately creating a robust system with applications in appointment setting. The integration of diverse data sources, combined with advanced machine learning techniques, contributed to a comprehensive foundation model that enhances predictive capabilities in sales contexts. Active learning was incorporated as a strategy to selectively improve model performance by focusing on uncertain or high-value samples, which aligns with methodologies proposed by Lakshminarayanan et al. (2018) [9] and Beluch et al. (2018) [2]. This iterative labelling approach refined model predictions, reducing noise and enhancing accuracy in key areas like customer sentiment and engagement analysis.

Despite meaningful progress, several limitations remain within the current model's design, implementation, and testing.

Firstly, the dependency on audio quality. Although various noise reduction techniques were applied, model performance is still heavily dependent on audio quality. When these techniques underperform, transcription and downstream analyses suffer, impacting the reliability of the final predictions. The attempt to integrate Beluch et al.'s (2018) [2] active learning approach helped improve feature selection by prioritising uncertain samples. However, computational and time constraints limited its full application, particularly in scenarios with low-quality audio.

Secondly, the lack of real-time Processing. The model currently operates offline, which restricts its applicability for live sales interactions where real-time feedback is crucial. The work of Gal et al. (2017) [6] on uncertainty-based techniques highlighted the importance of reliable, real-time predictions, especially in dynamic environments. However, resource limitations and the need for high computational power posed challenges in achieving real-time implementation, thereby reducing the model's responsiveness in real-world applications.

Lastly, inherent data bias. The model was trained primarily on specific types of successful calls, potentially leading to biased predictions that do not generalise across diverse sales scenarios. The study of Lakshminarayanan et al. (20180 [9] underscores the impact of training data diversity on generalizability, yet time and resource constraints limited the ability to diversify the dataset further. Without explicit bias mitigation, the model may underperform in scenarios that deviate significantly from the training data.

### 9.3 Recommendations for Future Enhancements

To address these limitations and expand the model's applicability, some of the following enhancements are recommended.

Firstly is the implementation of real-time processing with WebSockets. Integrating WebSocket support could enable real-time transcription and analysis during sales calls, providing live recommendations to sales personnel based on the model's outputs. This addition would allow the system to leverage dynamic customer cues, aligning with Gal et al.'s findings on real-time uncertainty estimation in high-stakes decision-making.

Next, is the incorporation of emotional labels to the transcription input. Adding emotional tone recognition to the model could enable a more nuanced analysis of customer interactions. By detecting emotional cues, the system could tailor its recommendations based on non-verbal indicators, improving rapport-building efforts with customers. The addition of such emotional labels would require further adaptation of the active learning framework to optimise feature selection and labelling in this domain.

Finally, to improve generalisability, efforts should be made to expand and diversify the training data to include a broader range of sales scenarios. Implementing explicit bias mitigation techniques, as demonstrated in Lakshminarayanan et al.'s and Beluch et al.'s research, would help address over-representation issues and enhance the model's ability to handle diverse customer interactions more effectively. Additionally, combining these techniques with the active learning approach could ensure that the labelling process remains efficient, focusing on high-impact samples that enhance the model's predictive accuracy across a variety of contexts.

## References

[1] Bellman, R. (1961). *Adaptive control processes : a guided tour*. Princeton University Press.

[2] Beluch, W. H., Genewein, T., Nürnberger, A., & Kohler, J. (2018). The Power of Ensembles for Active Learning in Image Classification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, *1*(10.1109/CVPR40276.2018), 9368, 9377. https://doi.org/10.1109/cvpr.2018.00976

[3] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440, 8451. https://doi.org/10.18653/v1/2020.acl-main.747

[4] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 4171–4186. https://doi.org/10.18653/v1/n19-1423

[5] Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: The Munich Versatile and Fast Open-source Audio Feature Extractor. *MM '10: Proceedings of the 18th ACM International Conference on Multimedia*, 1459–1462. https://doi.org/10.1145/1873951.1874246

[6] Gal, Y., Islam, R., & Ghahramani, Z. (2017). Deep Bayesian Active Learning with Image Data. *Proceedings of the 34th International Conference on Machine Learning* , *70*, 1183–1192. https://dl.acm.org/doi/pdf/10.5555/3305381.3305504

[7] Giordano, R., Ingram, M., & Broderick, T. (2024). Black Box Variational Inference with a Deterministic Objective: Faster, More Accurate, and Even More Black Box. *Journal of Machine Learning Research*, *25*, 1–39. https://jmlr.org/papers/volume25/23-1015/23-1015.pdf

[8] Khang, A., Abdullayev, V., Jadhav, B., Gupta, S. K., & Morris, G. (2023). *AI-Centric Modeling and Analytics: Concepts, Technologies, and Applications* (1st ed.). CRC Press.

[9] Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, *3*, 6405–6416. https://dl.acm.org/doi/pdf/10.5555/3295222.3295387

[10] Malz, N. (2024). Deep Bayesian Active Learning with High-Dimensional Data - A Critical Review [Review of *Deep Bayesian Active Learning with High-Dimensional Data - A Critical Review*]. *Machine Learning with Limited Labels Seminar*. https://nicolasmalz.com/resources/dbal_paper.pdf

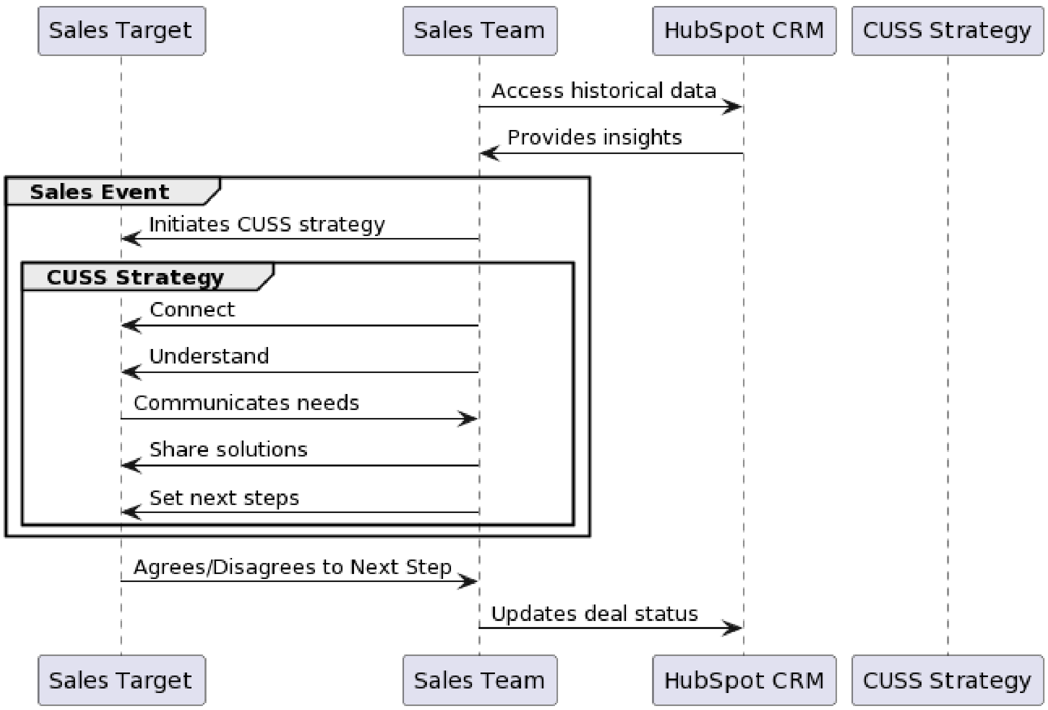
[11] Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, *63*(3), 403, 414. https://doi.org/10.1016/j.bushor.2020.01.003

[12] Ranganath, R., Gerrish, S., & Blei, D. M. (2014). Black Box Variational Inference. In S. Kaski & J. Corander (Eds.), *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* (Vol. 33, pp. 814, 822). Machine Learning Research Press. https://proceedings.mlr.press/v33/ranganath14.html

[13] Rishi Bommasani, Hudson, D. A., Ehsan Adeli, Altman, R. B., Arora, S., Sydney von Arx, Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. S., Chen, A. T., Creel, K., Davis, J., Demszky, D., & Donahue, C. (2021). On the Opportunities and Risks of Foundation Models. In *The Center for Research on Foundation Models (CRFM)* (Vol. 1, Issue 1, pp. 152–160). Stanford Institute for Human-Centered Artificial Intelligence (HAI). https://doi.org/10.48550/arxiv.2108.07258

[14] Adnan, K., & Akbar, R. (2019). Limitations of information extraction methods and techniques for heterogeneous unstructured big data. *International Journal of Engineering Business Management*, *11*(1), 184797901989077. https://doi.org/10.1177/1847979019890771

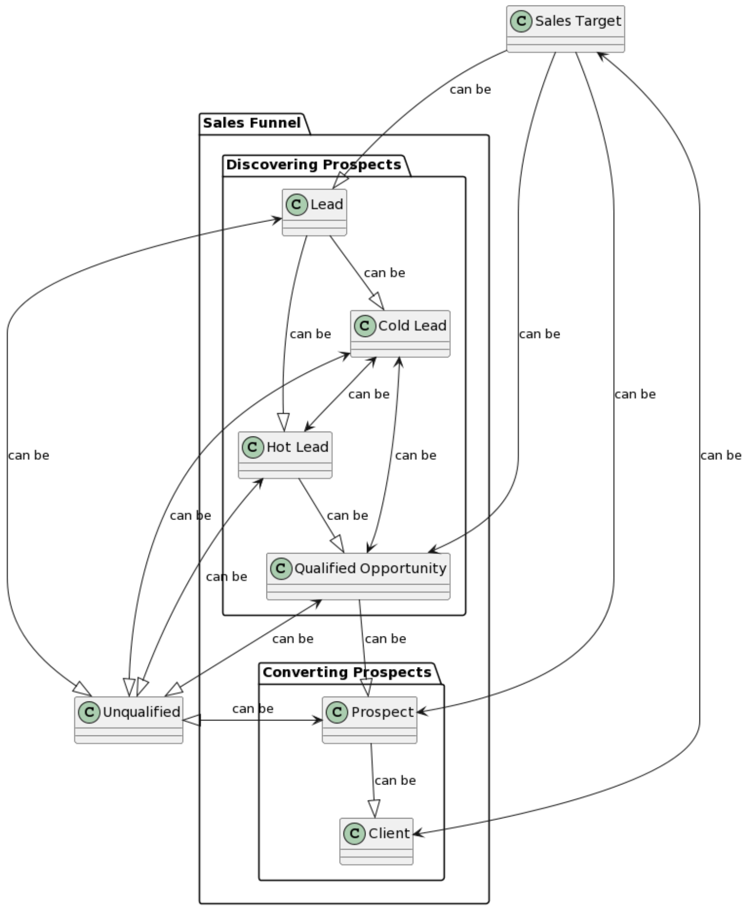
Appendix A - Sales Event Explanation




*Sales Event*

A sales event involves communication between a Sales Team and their sales targets via any means. An event is defined by a 4 point communication protocol or the CUSS strategy,

1. Connecting: Building rapport with prospects.
2. Understanding: Learning about their needs.
3. Sharing: Explaining how StaffAny can meet their needs.
4. Setting: Planning the next steps




*Sales Funnel*

Each sales event categorises a sales target. This can be visualised as a state transition pathway from the starting 'Sales Target' class to 'Client' class or 'Unqualified' class. This pathway is known as the sales funnel.

'Sales Target' class is the entry point and 'Unqualified' class is the exit point of the sales funnel. 'Client' class signifies the end of the sales funnel.

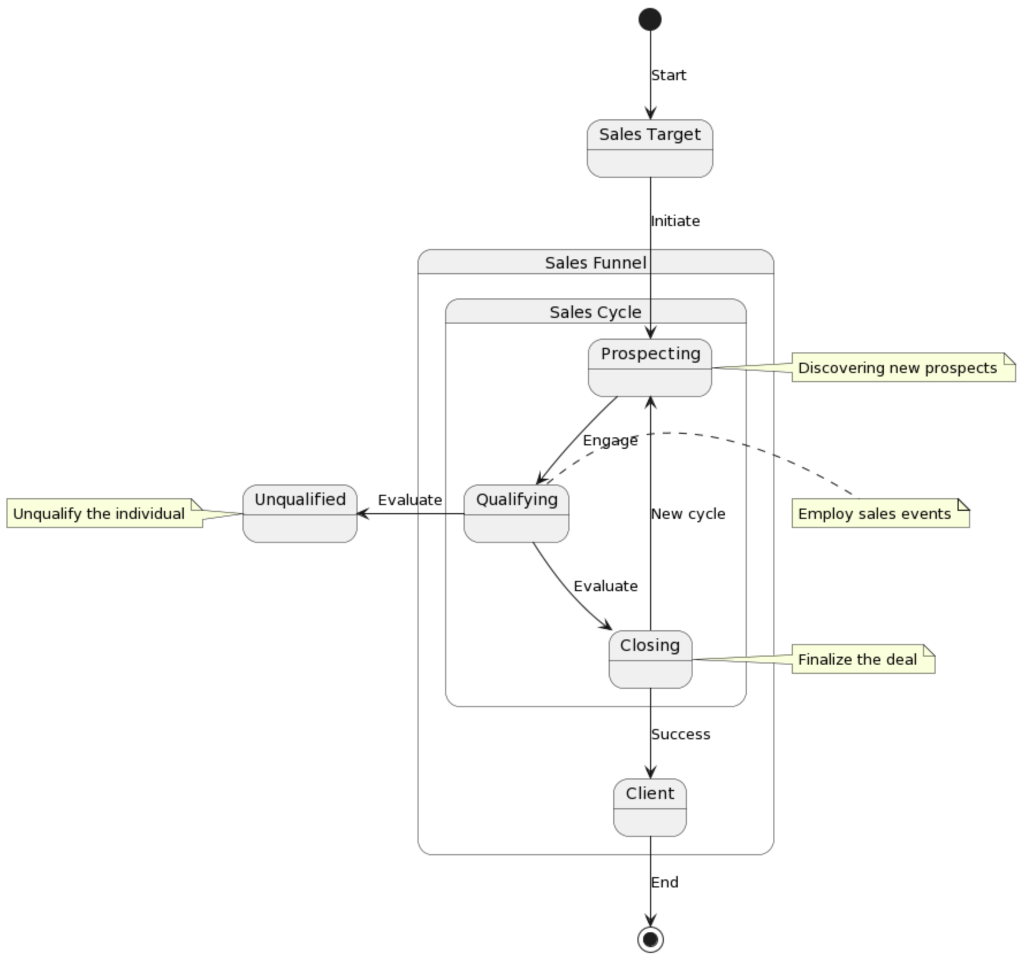
White arrowheads reflect a typical state transition pathway in the sales funnel. Represented as a series of class castings in the context of a state pattern,

1. 'Sales Target' class casted to 'Lead' class.
2. If warmed up to a sales person,
    a. Then, cast to 'Hot Lead' class
    b. Else if requires more warming, cast to 'Cold Lead' class
    c. Else, cast to 'Unqualified' class
3. If strong product fit and/or likelihood for purchase,
    a. Then, cast to 'Qualified Opportunity' class
    b. Else, cast to 'Unqualified' class
4. If agree to a product demo,
    a. Then, cast to 'Prospect' class
    b. Else, cast to 'Unqualified' class
5. If purchase after demo
    a. Then, cast to the 'Client'' class
    b. Else, cast to 'Unqualified' class

Black arrowheads reflect atypical state transitions. Bidirectional transitions between 'Lead', 'Cold Lead', 'Hot Lead', 'Qualified Opportunity' or 'Prospect' classes to 'Unqualified' class is possible. For instance, if the reasons for leaving the sales pipeline are temporary, then a transition back to any class in the sales funnel is possible once these reasons are resolved. Bidirectional transitions between 'Client' class and 'Sales Target' class is also possible. For instance, if service A but not service B was purchased, then 'Client' class for service A can transit to 'Sales Target' class for service B.

'Sales Target' class can transit to any class in the sales funnel without a sales event. For instance, if an individual uses a sign up form, then it is not considered a sales event. Multiple transitions can occur in a single sales event. For instance, if an individual arranges for a product demo in a single sales event, then 'Sales Target' class is casted to 'Lead' class, to 'Hot Lead' class, to 'Qualified Opportunity' class and to 'Prospect' class. A single transition may require multiple sales events. For instance, if an individual of the 'Qualified Opportunity' class agrees to a product demo after 2 sales events, then 2 sales events are needed to transition to the 'Prospect' class.

Sales events activate state transitions. Sales events are classified into 2 categories - discovering prospects and converting prospects - depending on the desired state transition.




*Sales Cycle*

The sales funnel can be generalised into 3 stages - Prospecting, Qualifying and Closing. This is the sales cycle.

The sales cycle is sequential. It must start with prospecting and end with closing. The sales cycle is cyclical. A failed or successful closing can always lead back to prospecting. The sales cycle's duration varies by prospect/client size, often longer for bigger prospects/clients.

Appendix D - StaffAny Sales Operations

Sales operations consist of 3 key components - sales events, sales funnels and sales cycles. Refer to Appendix A, B or C for more information.

| Metric | Description | Formula |
| --- | --- | --- |
| Total Appointments Set | Total number of appointments scheduled with potential leads. | Count of appointments |
| Appointment Setting Rate | Percentage of cold contacts that result in an appointment. | (Appointments set / Cold contacts) * 100 |
| Total Number of Sales Activities | Total count of sales-related actions taken (calls made, emails sent, etc.). | Count of sales activities |
| Conversion Rate | Percentage of appointments that result in a sale. | (Sales closed / Appointments set) * 100 |
| Sales Cycle Duration | The average time taken from initial contact to closing a sale. | (Sum of all sales cycles) / (Number of sales closed) |

Above are StaffAny's Sales Metrics which serves as a quantifiable basis to judge salespersons on.