Navigating Identities in Text: Towards an Approach for Dementia Care

by

Jess Gano

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2024

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Identity, as a concept, is concerned with the social positioning of the self and the other. It manifests through discourse and interactions, and expressed in relation to other perceived identities. For example, can one be or talk as a *leader* without strictly categorizing those they interact with as *subordinates* or *employees*? Research shows that the onset and progression of dementia may undermine the individual's sense of self and identity. This loss of self or identity has not only been found to cause significant decrease in well-being, but also affect caregiver/care-recipient relationships. However, while identity is compromised in some way, it does not necessarily mean it is completely lost. Autobiographical stories, especially those told repeatedly, may serve as means to reveal significant aspects of the storyteller's self and identity.

In this thesis, we explore the task of persona attribute extraction from dialogues as a proxy for identity cues. We define persona attribute as a triplet *(s, r, o)*, where the relation r indicates the persona attribute type or relationship between the subject s and object o e.g., (I, `has_hobby`, knitting). Employing an information extraction approach, we design a two-stage persona attribute extractor, consisting of a relation predictor and entity extractor. Respectively, we define relation prediction as a multi-label classification task using BERT embeddings and feedforward neural networks, and entity extraction as a template infilling task following the pre-training objective of T5 [39]. We employ our methods on a proxy dataset created by combining Persona-Chat and Dialogue-NLI. Factoring ethical considerations and potential risks, directly evaluating our methods on a dementia use-case is not a feasible task. Therefore, we utilize a dataset consisting of interviews with older adults to assess feasibility within a context more closely resembling the dementia use-case.

Exploring the research problem and developing our methodology highlights the following insights: (1) inferring identities from text, especially considering its nuanced representation in discourse, is challenging due to the abstract nature of identity itself and (2) to our knowledge, there is no available dataset that exhibits the distinct speech characteristics inherent in older adults making training and evaluating models tailored to this demographic very challenging. Furthermore, experiments on the older adults dataset show that a transfer learning approach to solving this problem is insufficient due to significant contrast between the datasets from the source and target domains.

Acknowledgements

I would first like to thank my supervisor, Professor Jesse Hoey, for being willing to take a chance on me and welcoming me as a Master's student since I first joined the CHIL meetings in Winter 2022. It has truly been an honour and a privilege to be one of your students.

I'm very grateful to each of my readers, Professor Jim Wallace and Professor Jimmy Lin, for their feedback on this thesis and for being so accommodating of our timeline.

I would like to thank my NEATO research team — Professor Hoey and Renee Leung, and our collaborators from the English Language and Literature Department — Professor Heather Love, Heather Eustace, and Alyssa Clarkson, for their support and patience in annotating our dataset and tackling this intricate research problem.

I would like to give special thanks to my lab mates turned friends at CHIL — Christopher Risi, Blake VanBerlo, and Joshua Jung, for their guidance and advice on navigating life within and beyond graduate school. I'll never forget our ever-entertaining lab meetings and CHIL's sourdough phase courtesy of Blake and Josh.

Thank you so much to my family and friends for sending their love and support across distant time zones. Especially, for sending me photos of my niece and nephews which never failed to cheer me up.

And final thanks to my boyfriend, William Loh, the only other person to attend the Master's coursework online orientation, for helping me grow in graduate school in so many ways. I appreciate your willingness to be brought along this chaotic journey and jump between cafes and study spaces on campus with me. You are one of the biggest reasons I finished my Master's with my sanity intact and I am forever grateful.

Dedication

To my dad, in loving memory.
To my mom, for granting me the privilege to pursue my goals.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Dementia describes a particular group of symptoms characterized by difficulties with memory, language, and thinking that can affect a person's ability to navigate daily life [2]. In the moderate to severe stages of Alzheimer's disease, the predominant cause of dementia, individuals may exhibit forgetfulness regarding recent events, struggle recognizing family and friends, and experience behavioural changes that may escalate and involve anger and aggression. As a consequence, dementia is often stigmatized and associated with losing one's sense of self or identity [37].

The loss of self or identity for persons with dementia has not only been found to cause significant decrease in well-being [28] but also affect caregiver/care-recipient relationship. Wuest et al. [51] refers to the latter as "the process of becoming strangers", which has been reported to be a significant source of distress for caregivers. Therefore, understanding how persons with dementia perceive themselves and their environment is key to fostering communication and cooperation with them. Caregivers can employ sociological, relational, and individual characteristics about the care-recipient as identity cues to initiate interactions and prompt positive reactions [47]. Referencing an example from Vézina et al. [47], "Sometimes, when he doesn't want to get dressed, just tell him it's Sunday morning and he has to clean up to get to mass, and everything will go fine." The caregiver recognized the individual's spiritual commitment as an identity cue that can be used to encourage cooperation in daily activities. Similarly, greeting expressions like "Hello, mom, it's Emma, your daughter" and terms of endearment such as "my dear" is a means of establishing the social identity of the person with dementia as well as the relationship between the people present. The use of such identity cues help reinforce persons' with dementia self-esteem and identity. Moreover, a study conducted by Cohen-Mansfield et al. [10] also suggests that treatments and interventions that were designed to incorporate persons' with dementia sense of identity was highly effective in reducing disorientation and agitation, as well as increasing positive emotional response and involvement in activities.

In this thesis, we explore the task of extracting *persona attributes* as a proxy method for inferring identity cues within a dementia care context. Persona attributes refer to characteristics, traits, or qualities that define an individual's identity. We define persona attribute as a triplet *(s, r, o)*, where the relation r indicates the persona attribute type or relationship between the subject s and object o e.g., (I, `has_hobby`, knitting). This proves to be a challenging task for several reasons. First and foremost, evaluating our approach on persons with dementia is not a feasible task due to ethical considerations and the potential risks associated with using technology that may not be fully developed and ready. We discuss a framework for developing technology for dementia in detail in Section 3.1. To address this problem, we utilize a dataset consist of interviews with older adults to evaluate our approach and overall assess the feasibility of the task. In Section 4.2, we introduce the OLDER ADULTS INTERVIEW DATASET and provide an overview of its contents and relevance to this study. This step allows us to evaluate the performance of our models on data that somewhat resembles the linguistic patterns found in persons with dementia, but is feasible for our stage of prototyping. This leads us to the second challenging aspect of this research problem: narratives told by persons with dementia, as well as older adults in general, tend to be less chronologically structured and often include stories that are told repeatedly. These stories can also be susceptible to discourse impairments such as disruptive topic shifts and excessive repetitions caused by the decline in cognitive abilities. To tackle this particular issue, we employ methodologies in natural language processing (NLP) that are robust to noise in the data and effective in capturing the semantics of an entire sequence. However, it should also be noted that there does not exists a dataset that captures the linguistic characteristics found in older adults, nor one a dataset that is specifically for extracting persona attributes from dialogues. We therefore follow existing literature and utilize two datasets, PERSONA-CHAT and DIALOGUE-NLI, to train our models and employ a transfer learning approach on the OLDER ADULTS INTERVIEW DATASET. This gap in knowledge inspires additional research both on the task of extracting persona attributes from text and exploring the applicability of NLP on technologies tailored for older adults.

1.2 Contributions

The contributions we make through this research is two-fold. First, we explore the task of extracting persona attributes from dialogues as a proxy method for inferring identity cues from dialogues. We propose a two-stage attribute extractor, comprised of a relation predictor and and entity extractor model, to extract persona attributes in the form of a triplet *(s, r, o)*, where the relation r indicates the persona attribute type or relationship between the subject s and object o. To address the unavailability of dataset for this task, we combine the PERSONA-CHAT and DIALOGUE-NLI datasets for weak supervised learning.

1.3 Thesis Outline

This thesis is organized as follows:

- In Chapter 2, we present the definition of identity used in this study and explore interpretations and use of the concept in dementia and caregiving. We also introduce fundamental topics in modern natural language processing (NLP), such as Feedforward Neural Networks, Recurrent Neural Networks, and Transformers.
- In Chapter 3, we discuss technologies that have been developed for dementia care (e.g., diagnosis, care delivery, monitoring). This discussion also examines the inherent challenges associated with developing these technologies, particularly focusing on ethical adoption. Additionally, we summarize relevant literature on persona attribute extraction.
- In Chapter 4, we outline our proposed model for Persona Attribute Extraction. We also introduce the datasets that will be used in this study, namely PERSONA-CHAT, DIALOGUE-NLI, and the OLDER ADULTS INTERVIEW DATASET.
- In Chapter 5, we discuss the implementation details of our models and present experimental results on the combined PERSONA-CHAT-DIALOGUE-NLI dataset.
- In Chapter 6, we conduct an experiment on the OLDER ADULTS INTERVIEW DATASET using our model. We also analyze the implications of aging on language and cognition that makes for a challenging research problem. Moreover, we discuss interdisciplinary insights gained from our collaborative efforts for this study.
- In Chapter 7, we summarize our work and discuss possible future work.

Chapter 2

Background

In this chapter, we provide an overview of the literature relevant to this study. We begin by presenting the definition of identity used in this thesis, along with an overview of critical approaches for studying identity. We then cover technical details of Feedforward Neural Networks, Recurrent Neural Networks, and the Transformer architecture which are considered building blocks of modern NLP.

2.1 Dementia & Identity

Identity is a complex and ambiguous concept. Dictionary definitions follow a more traditional usage of the word and fail to capture its meaning as how it's currently used [18]. From the Latin root *idem*, meaning "*the same*", the word identity has had varying interpretations throughout history. In mathematics, it means to remain true for all values of the variables involved (e.g., $a(b+c) = ab + ac$). In a philosophical context emerging from ancient Greek thought, "it is a marker that distinguishes and differentiates one object from another object" [43]. In a somewhat paradoxical manner, it suggests both sameness and difference.

In this thesis, we draw theoretical background from humanities and the social sciences, adopting the concept of identity as *the social positioning of the self and the other* [7]. For example, can one be or talk as a *leader* without strictly categorizing those they interact with as *subordinates* or *employees*? Identity manifests through discourse and interactions, and is expressed in relation to other perceived identities. Moreover, it is through talking that we build and maintain relationships and establish who we are to one another [14].

For persons with dementia, their sense of identity often becomes overshadowed by their medical diagnosis. Additionally, the portrayal of dementia as a loss of self or identity perpetuates a narrative where affected individuals are perceived as the "*demented other*" within society [36]. Memory loss, cognitive decline, and changes in behavior can all contribute to a profound shift in how they perceive themselves and interact with the world around them. Thus, narratives from persons with dementia have become instrumental for studying theories on identity within a dementia context [27]. While identity is compromised in some way with dementia, it does not necessarily mean that it is completely lost. Autobiographical stories, especially those told repeatedly, may serve as means to reveal significant aspects of the storyteller's self and identity. Human beings also are natural storytellers. Therefore, rather than emphasizing loss and dysfunction, we focus on narration and interaction as integral components in shaping the definition of identity used in this study.

2.2 Feedforward Neural Network

Feedforward neural networks, also known as multilayer perceptrons (MLPs), are foundational machine learning models loosely inspired by the interconnected structure of *neurons* in the brain. The core building block of a neural network is a neuron and a collection of neurons is referred to as a *layer*. Typically, a neural network is comprised of an *input layer*, one or more *hidden layers*, and an *output layer*. Each neuron in the hidden layer receives input from the previous layer, calculates a weighted sum of those inputs, that is then passed into an *activation function* to produce an output. This can be expressed mathematically as

$$\hat{\boldsymbol{y}} = \sigma(W\boldsymbol{x} + \boldsymbol{b})$$

where $\boldsymbol{x} \in \mathbb{R}^d$ are the inputs, $W \in \mathbb{R}^{m \times d}$ is a weight matrix, $\boldsymbol{b} \in \mathbb{R}^m$ is the bias, σ is an activation function, and $\hat{\boldsymbol{y}} \in \mathbb{R}^m$ is the output. Activation functions are essential in neural networks as they introduce non-linear properties to the model, allowing it to learn more complex patterns. A number of options for activation functions exists, with Rectified Linear Unit (ReLU) and Sigmoid being among the commonly used ones.

As illustrated in Figure 2.1 , the graph is *acyclic* i.e., it does not have a loop and the input to some neuron can never depend on the same neuron's output. In training a neural network, the goal is to optimize the model parameters (weights and biases) by performing iterative updates using a *gradient-based optimization* comprised of a *forward pass* then a *backward pass*. In a forward pass, the input is passed through each layer of the network to make an inference. The difference between the actual value $\boldsymbol{y}$ and predicted output $\hat{\boldsymbol{y}}$ is computed using a loss function (e.g., Mean Squared Error, Cross Entropy Loss, Hinge

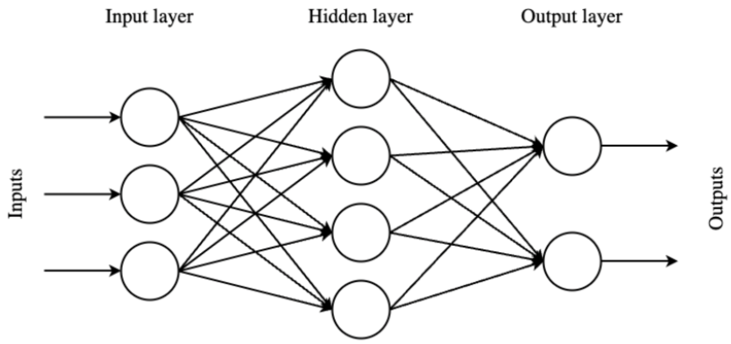

Figure 2.1: Feedforward neural network with one hidden layer.

Loss). In a backward pass, the computed loss is propagated through each of the network's layers through *backpropagation*, adjusting the model's parameters to minimize the difference between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$. This gradient update step can be expressed as

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ is the network parameter, $\boldsymbol{\eta}$ is the leaning rate, and $\boldsymbol{\mathcal{L}}$ is the loss function.

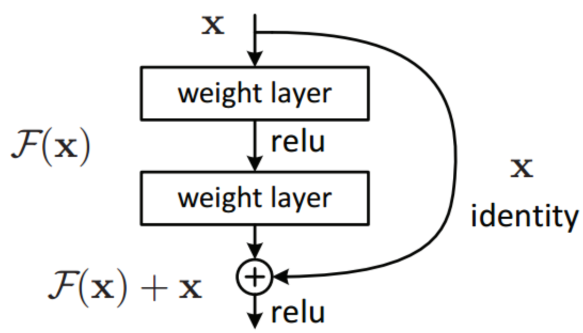

Figure 2.2: Residual block illustration from He et al. [25]

The depth of the network is defined by its number of layers. Deeper networks can suffer from the *vanishing gradient problem*, where gradients diminish exponentially (approaching zero) or become too large (approaching infinity) as they are propagated through the network layers during backpropagation. This issue can affect the training process, causing slow convergence or hindering learning entirely for exceptionally deep networks. To mitigate this issue, He et al. [25] propose *residual learning* which enables training of deep networks by incorporating "shortcut" connections. As shown in Figure 2.2, the input $\boldsymbol{x}$ traverses each layer of the network, while simultaneously bypassing certain layers without any transformations. Incorporating the original input into the output through the shortcut connections facilitate the flow of gradients during training; therefore, mitigating the vanishing gradient problem.

2.3 Recurrent Neural Networks

Recurrent neural networks (RNNs) is a family of neural networks for handling sequential data, which involves variable length inputs and outputs. Unlike feedforward neural networks, RNNs have connections that form cycles which allows them to process sequences of data. In their basic configuration, RNNs maintain a *hidden state* vector that is propagated through the network as each input in the sequence is processed, enabling the model to capture dependency relations. This feature is essential for NLP tasks, such as text generation, where the prediction for the next token depends on the context provided by the preceding word(s). Figure 2.3 shows an illustration of an RNN (a) and an unfolded representation of the network (b). At each time step t, the previous hidden state $\boldsymbol{h}_{t-1}$ and the current input $\boldsymbol{x}_t$ are used to update the current hidden state $\boldsymbol{h}_t$ and produce the current output $\boldsymbol{y}_t$. This process can be expressed mathematically as

$$\begin{aligned} \boldsymbol{h}_t &= g(W_{hh}\boldsymbol{h}_{t-1} + W_{hx}\boldsymbol{x}_t) \\ \hat{\boldsymbol{y}}_t &= f(W_{hy}\boldsymbol{h}_t) \end{aligned}$$

where W_{hh}, W_{hx}, and W_{hy} are learnable parameters shared across all time steps and f,g are activation functions.

While RNNs have the ability to deal with sequential data and capturing temporal dependencies, their sequential architecture limits their ability to perform parallel computations efficiently. This makes them less efficient for tasks that could benefit from parallelization, such as processing large datasets or training on multiple GPUs simultaneously. RNNs also encounter challenges in compressing previous inputs into the hidden states. This is known as the long-range dependency problem in RNNs. Because the dimensions of the hidden state is finite and the sequence length can get arbitrarily long, RNNs can "forget" information from distant time steps. This is partially a side-effect of the vanishing gradient problem discussed in Section 2.2. In the case of RNNs, the gradients will either grow or

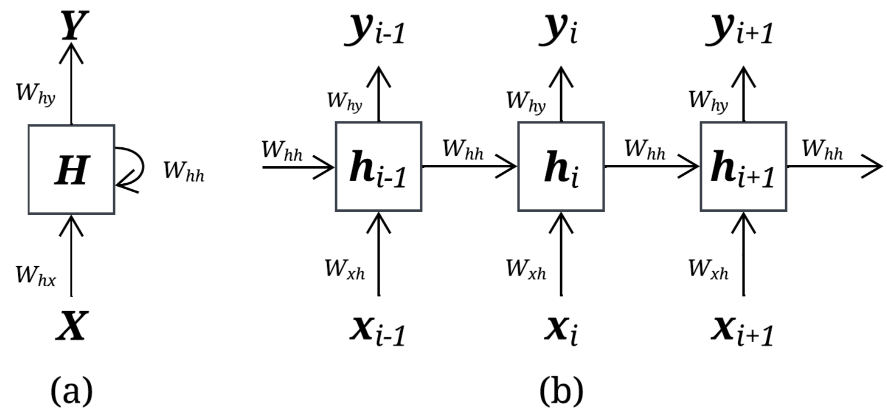

Figure 2.3: (a) vanilla RNN (b) unrolled RNN forward computation

shrink exponentially with the length of the sequence, making it very difficult to train. This leads to parameter updates that are either negligible or unstable.

Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) networks address these challenges by by incorporating gating mechanisms to regulate the flow of information through the network. Concretely, an LSTM maintains a *cell state* $\boldsymbol{c}_t$, in addition to a hidden state $\boldsymbol{h}_t$. The update step for LSTM can be expressed mathematically as

$$
\begin{aligned}
f_t &= \sigma(W_{xf}\boldsymbol{x}_t + W_{hf}\boldsymbol{h}_{t-1} + b_f) && \text{(forget gate)} \\
i_t &= \sigma(W_{xi}\boldsymbol{x}_t + W_{hi}\boldsymbol{h}_{t-1} + b_i) && \text{(input gate)} \\
g_t &= \tanh(W_{xc}\boldsymbol{x}_t + W_{hc}\boldsymbol{h}_{t-1} + b_c) && \text{(cell state update)} \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t && \text{(cell state)} \\
o_t &= \sigma(W_{xo}\boldsymbol{x}_t + W_{ho}\boldsymbol{h}_{t-1} + b_o) && \text{(output gate)} \\
\boldsymbol{h}_t &= o_t \odot \tanh(c_t) && \text{(output state)}
\end{aligned}
$$

where the *forget gate* f_t determines whether the cell state should retain or forget information, the *input gate* i_t controls the amount of input information incorporated to the cell state, and the *output gate* o_t regulates which information from the cell state is passed to the subsequent LSTM unit.

GRU is similar to the LSTM, but is simplified such that the cell state is omitted and only two gates are used to control information flow. The *update gate* z_i controls how much information from the previous time steps and how much of the candidate hidden state is kept. The *reset gate* r_t determines how much information from previous time steps should be forgotten or reset. The update step for GRU can be expressed mathematically as

$$
\begin{aligned}
r_t &= \sigma(W_{xr}\boldsymbol{x}_t + W_{hr}\boldsymbol{h}_{t-1} + b_r) && \text{(reset gate)} \\
z_t &= \sigma(W_{xz}\boldsymbol{x}_t + W_{hz}\boldsymbol{h}_{t-1} + b_z) && \text{(update gate)} \\
\bar{h}_t &= \tanh(W_{xh}\boldsymbol{x}_t + W_{hh}(r_t \odot \boldsymbol{h}_{t-1}) + b_h) && \text{(candidate hidden state)} \\
h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \bar{h}_t && \text{(hidden state update)}
\end{aligned}
$$

While both the LSTM and GRU are designed to mitigate the vanishing gradient problem, they still struggle in handling long-range dependencies and parallelization. Both models may still struggle to effectively capture and retain information for very long sequences despite their gating mechanisms. Additionally, their sequential nature limits their ability to process inputs in parallel, as computations must be performed step-by-step. This potentially could lead to slower training and inference times.

2.4 Transformers

2.4.1 Attention Mechanism

Attention mechanism, first introduced by Bahdanau et al. [4], allows a network to "attend" to specific parts of the sequence over other parts based on relative importance. Conceptually, attention can be described as a function of *query* and *key-value* pairs. The attention mechanism adopted by Transformers is called *scaled dot-product attention.* Given the embedding of the input sequence $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, we create the query, key, and value vectors for each token by

$$
\begin{aligned}
\boldsymbol{q}_i &= W^Q\, \boldsymbol{x}_i \\
\boldsymbol{k}_i &= W^K\, \boldsymbol{x}_i \\
\boldsymbol{v}_i &= W^V\, \boldsymbol{x}_i
\end{aligned}
$$

where W^Q, W^K, W^V are learnable projection matrices. We calculate the attention scores between each pair of tokens using the dot product of the query and key vectors, followed by a scaling factor to stabilize gradients by

$$
\boldsymbol{s}_{ij} = \frac{\boldsymbol{q}_i \cdot \boldsymbol{k}_j}{\sqrt{d_k}}
$$

where d_k represents the dimensionality of the key vectors. We normalize the attention scores with a softmax function

$$\boldsymbol{\alpha}_{ij} = \text{softmax}(\boldsymbol{s}_{ij}) = \frac{\exp(\boldsymbol{s}_{ij})}{\sum_{j'=1}^{n} \exp(\boldsymbol{s}_{ij}))}$$

The final step is to compute a weighted sum of the value vectors using the normalized attention scores by

$$\boldsymbol{z}_i = \sum_{j=1}^{d_k} \alpha_{ij} v_j$$

In practice, the attention scores are computed using matrix operations, which allow for efficient parallelization across multiple tokens and batches. Given an embedding matrix X, the query, key, and value matrices are computed by

$$\begin{aligned} Q &= XW^Q \\ K &= XW^K \\ V &= XW^V \end{aligned}$$

We calculate the attention scores by

$$\text{Attention}(Q; K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

2.4.2 Transformers Architecture

RNN-based models struggled with capturing dependencies across distant tokens due to their sequential nature. Since these models process each token in a sequence one by one, their capacity for parallelization is inherently limited. Introduced by Vaswani et al. [46], the Transformer architecture addresses these limitations by avoiding recurrence, relying on the attention mechanism, leveraging parallelization techniques that allow the model to handle longer sequences without significant increase in computational cost. As illustrated in Figure 2.4, the Transformer architecture consists of two primary components: an *encoder* and a *decoder*. Conceptually, the encoder maps the input sequence $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ into a fixed-length contextual representation $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$, which is then passed to the decoder to generate the output sequence $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$.

The encoder consists of a stack of N identical layers, each composed of two sub-layers: the multi-head self-attention layer and a feedforward neural network. As discussed in

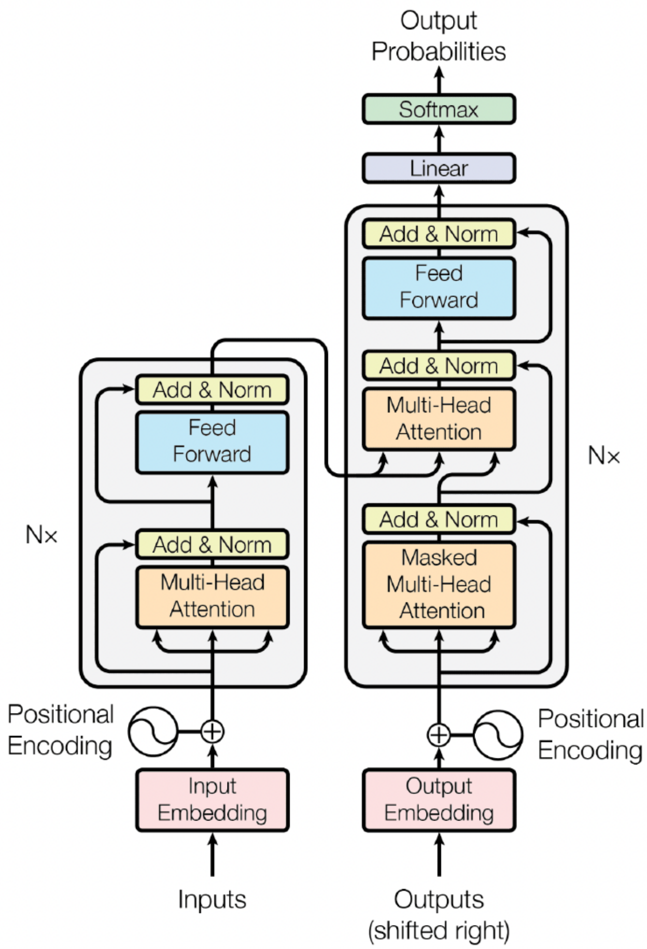

Figure 2.4: Transformer model architecture from Vaswani et al. [46]. On the left is the encoder component which is responsible for processing the input data, while the right side shows the decoder component which handles generating output sequences.

Section 2.4.1, the attention mechanism allows the model to weigh the significance of each word relative to every other word therefore efficiently capturing long-range dependencies. Self-attention implies that the key, query, and value vectors are derived from the same sequence. Multi-head attention is adapted by having multiple sets of W^Q, W^K, W^V matrices that are learned in parallel and concatenated afterwards as shown in Figure 2.5. The multi-head attention can be computed by

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O$$

where each head is computed is computed as

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Between sub-layers are residual connections followed by layer normalization modules. In the original Transformer architecture proposed by Vaswani et al. [46], there is a total of 6 encoder layers stacked and 8 attention heads; however, these parameters can vary depending on the implementation or variant of the model. Despite each layer being identical, parameters are not shared between them.

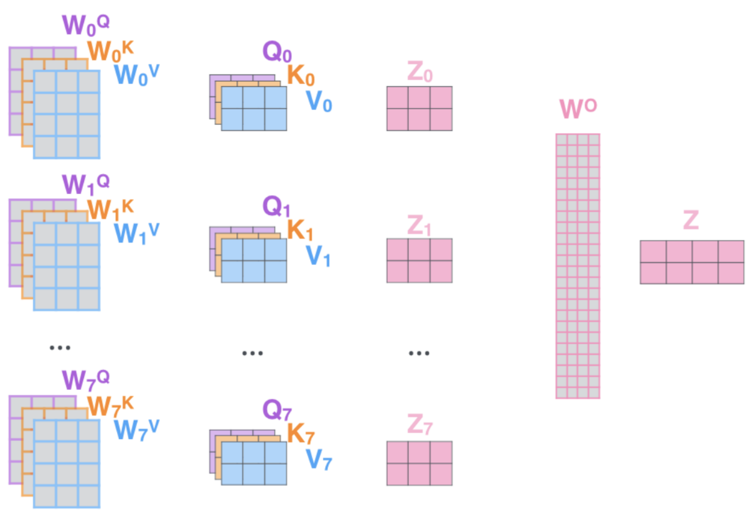

Figure 2.5: Multi-head attention illustration

The decoder is similarly composed of three sub-layers: a masked self-attention layer, an encoder-decoder attention layer, and a feedforward neural network. Masked self-attention is used to ensure that each position in a sequence attends only to positions before it. The reason for this is that the decoder is auto-regressive, which means that it generates one token at a time while incorporating the previously generated token as an additional input. This process is replicated while computing attention scores for each token in the sequence by setting the values to $-\infty$ before applying the softmax function. As a result, during the softmax operation, the masked values effectively get a weight close to zero, ensuring that the current token attends only to positions before it in the sequence. The encoder-decoder attention layer is a form of cross-attention which performs multi-head attention over the output of the encoder. Unlike self-attention, there are two different input sequences that serve as input for computing the query, key, and value vectors in cross attention. The attention score for the decoder is calculated by

$$\begin{aligned} Q &= XW^Q \\ K &= ZW^K \\ V &= ZW^V \end{aligned}$$

where Z is the output of the encoder layer and X is the input to the decoder.

2.4.3 Transformer-based models

BERT [12], or Bidirectional Encoder Representations from Transformers, is an encoder-only language model renowned for its capability to capture bidirectional contextual information from text. This is mainly achieved through the model's pre-training objectives: masked language modeling (MLM) and next sentence prediction (NSP). In MLM, a subset of the input tokens are masked and the model is trained to predict the masked tokens based on the surrounding context. On the other hand, NSP involves predicting whether a pair of sentences appear consecutively in the text. Put simply, MLM allows the model to learn relationships between words, while NSP allows it to learn the relationships between sentences. A special token `[CLS]`, which stands for classification, is prepended to every input sequence. During the pre-training phase of BERT, the model is trained to generate a representation for the `[CLS]` token that captures the overall semantics of the input sequence. This `[CLS]` token representation is then used as an aggregate representation of the entire sequence, which can be passed through additional layers for downstream tasks such as text classification or sentiment analysis.

T5 [39], short for Text-to-Text Transfer Transformer, uses the standard encoder-decoder structure of the original Transformer architecture and consists of 12 encoder-decoder layers. It follows a "text-to-text" framework such that text is both the input and output to the model. This entails that the same model, hyperparameters, and loss function can be used across different tasks such as text summarization, language translation, text

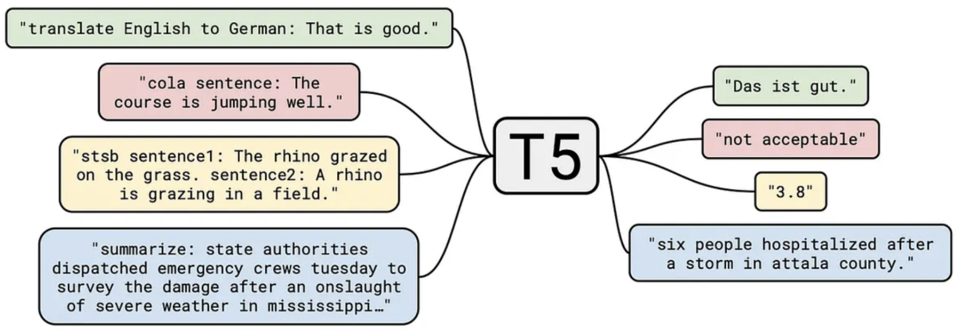

Figure 2.6: Diagram of the T5 framework from Raffel et al. [39].

classification, and sentence similarity. As shown in Figure 2.6, a task-specific prefix is prepended to the input sequence during model pre-training to obtain the desired output. For example, language translation tasks would require adding the prefix "`translate French to English:`" such that the input to the model is "`translate French to English:` Je étudiant l'informatique". For a textual entailment classification task, the pair of input sequences are prepended with "`mnli premise:`" and "`hypothesis`:" and the expected output is either `entailment`, `contradiction`, or `neutral`.

Chapter 3

Related Work

3.1 Technology in Dementia Care

Persons with dementia require a substantial amount of support and assistance, and this need increases with the progression of the disease. More recent innovations in technology offer promising solutions to address the challenges posed by an aging population affected by dementia, as well as the shortage of both family and professional caregivers. Based on the current technology developments in dementia reported by Astell et al. [3], there are four categories of technology-based solutions: (1) diagnosis, assessment, and monitoring, (2) maintenance of functioning, (3) leisure and activity, and (4) caregiving and management.

Within the category of diagnosis, assessment, and monitoring, advanced diagnostic technologies, such as neuroimaging and genetic testing, enable earlier detection and more accurate identification of dementia types. Big data research also enables in-depth analysis on large datasets to uncover valuable insights into the disease's progression, risk factors, and potential treatment strategies [22, 34]. Maintenance of functioning is supported by adaptive technologies designed to support daily tasks and promote independent living. Additionally, cognitive assistant technologies such as COACH [35], which utilizes computer vision technology to guide persons with dementia through hand washing using verbal and visual prompts, have garnered increased research attention in recent years. Technologies supporting leisure and activity encompass interactive games [17], social robots [38], and digital reminiscence therapy [15]. These not only provide entertainment but also foster mental stimulation and emotional well-being. Lastly, caregiving and management technology play a crucial role in supporting informal caregivers and facilitating remote care delivery, especially considering that the majority of persons with dementia are primarily cared for by family members.

Despite the potential benefits and significant development efforts made over the past decade, adaptive and assistive technologies for dementia mostly remain in the research stage and see limited commercialization. One of the main reasons for this are the ethical considerations on their potential risks and benefits. To guide future developments of technology for dementia care, Robillard et al. [42] introduce the concept of *ethical adoption*, defined as the integration of ethical principles into the design, development, deployment, ongoing usage, and management of technology. It is characterized by five core principles:

1. **Inclusive participatory design.** Including end users in every stage of the development process can help in evaluating the trade-off between risks and benefits for solutions that raise ethical considerations (e.g., wander safety products, in-home sensors). Moreover, diversity in user characteristics, language, community culture and environment should also be taken into consideration to ensure the relevance and feasibility of the solution for its target population.

2. **Emotional alignment.** Humans naturally seek social interaction in environments where their sense of self is acknowledged and respected. By incorporating emotional alignment in assistive technologies (e.g., robot companions, reminder/task management apps), we can promote positive relationship with the technology and enhance the well-being of persons with dementia who, otherwise, would regard these technologies with mistrust and suspicion.

3. **Adoption modeling.** When developing technologies for persons with dementia, it is important to recognize factors that can influence an individual to disengage from certain activities or situations. This can include personal preferences, physical and cognitive limitations, cultural norms, and the presence of coercion or undue influence.

4. **Ethical standards assessment.** It is important that ethical standards, particularly concerning privacy, confidentiality, and informed consent be assessed both during implementation and ongoing use. This entails ensuring that claims about the technology's capabilities are accurate and transparently communicated to all users. Upholding these ethical standards not only protects the rights of the persons with dementia but also fosters trust and confidence in the technology among caregivers.

5. **Education and training.** Several types of assistive technologies still face stigma from use. For example, wander safety products such as GPS trackers and monitoring

bracelets may be seen as intrusive due to concerns about privacy or surveillance. Therefore, education and training play a crucial role in the effective use of dementia technologies, especially when considering the potential negative impacts of poorly designed interfaces (e.g., aggression, reduced interaction).

3.2 Persona Attribute Extraction

Understanding user behaviour and preference plays an important role in creating personalized human-computer interactions. In dialogue systems, incorporating persona attributes enable dialogue models to generate personalized and engaging responses which has then increased reliance of dialogue systems for chit-chat [32, 52] or task-oriented tasks [24, 26].

Such persona attributes can either be predefined (i.e., collected through a demographic questionnaire) or revealed either implicitly or explicitly during a dialogue, with more recent advancements focused on the latter approach. Dialogues are a rich source of persona attributes that reveal insights about a person's hobbies, interests, and many more. Therefore, it is important that dialogue systems are capable of recognizing and extracting such information from a user's utterance. Earlier works typically used latent representations of such information to generate personalized or context-aware responses. Li et al. [32] encodes speaker-specific information (e.g., age, gender, country of residence) through a persona vector that is then integrated into the hidden layers of their response generation model. Inspired by knowledge-grounding, Zhao et al. [53] incorporates context-relevant documents into a pre-trained language model to create knowledge-grounded dialogue generation models. While effective in certain contexts, latent representations may not always translate to meaningful results when used in downstream tasks due to its lack of interpretability.

To support development of engaging chit-chat dialogue agents, Zhang et al. [52] constructed Persona-Chat which consists of dialogue-persona pairs collected through crowdsourcing on Amazon Mechanical Turk. Each persona is defined through multiple textual descriptions (e.g., *"I used work at a carnival"*, *"I like to drink scotch to relax"*) referred to as profile sentences. While the primary objective for Persona-Chat is to serve as a benchmark for how well dialogue models can maintain and embody their assigned personas in a conversation, the authors also present preliminary experiments for predicting user profiles from dialogue history. Several other works have since then explored persona attribute extraction using the Persona-Chat dataset. Gu et al. [23] frame their approach as a Speaker Persona Detection task where they experiment with different aggregation and encoding approach to identify the best-matching persona given a dialogue. A study by

Tigunova et al. [44] experiment with incorporating attention mechanism into their models to extract more interpretable attributes. However, their study only focuses on a specific set of attributes, namely profession, age, and family status and the need to formulate individual models for each attribute is not ideal due to scalability issues. Leung [31] conducted a brief investigation of the Persona Generation Task (PGTask) [41], comparing BART against the baseline GPT-2 model for generating profile sentences given an utterance. Compared to our study which centers on persona attribute extraction, the primary objective of Leung [31] is to explore how integrating persona (or profile sentences) into dialogue generation models can enhance conversational agents specialized in directing users to specific target topics.

More recent works redefine persona attribute extraction as a triplet extraction task, enabling the use of such information in downstream tasks. Extracted persona attributes would be in the form of triplet *(s, r, o)*, where the relation r indicates the persona attribute type or relationship between the subject entity s and object entity o.

Since there is no dialogue dataset available for this particular task, prior studies combined Persona-Chat and Dialogue-NLI [48] — a dataset containing manually annotated triplets for each dialogue utterance and profile sentence from Persona-Chat. The dataset was originally created for developing models that can infer the logical relationship (i.e., *entailment, contradiction,* or *neutral*) between utterances in a dialogue through their annotated triples. For instance, "*I just adopted a puppy*" and "*My dog is named Ori*" would have the same triplet annotation (I, `have_pet`, dog) and therefore, be classified as having an *entailment* relationship.

Wu et al. [50] leverage both datasets and develop a model consists of an end-to-end memory network and a gated recurrent unit (GRU) for generating (s, r, o) triplets given a dialogue utterance. To evaluate model performance, they compare the extracted triplet against the human annotated triplets and calculate the F1 and BLEU-1 scores. F_1 score was computed by comparing each entity in the triplet (i.e., compare $\hat{s}$ to s, $\hat{r}$ to r, and $\hat{o}$ to o). BLEU-1 scores where reported to account for variations in text that still convey the same meaning, e.g., (I, `favorite_season`, fall) and (I, `favorite_season`, autumn). Their method achieved an F_1 score of 0.2868 and BLEU-1 score of 51.87. However, the authors did not publish the scripts necessary to reproduce their results in their code repository hence, we cannot consider it as a reliable baseline. To tackle the issue of inconsistent annotations, Zhu et al. [54] manually re-annotate and correct triplet labels creating PersonaExt. While their study showed reasonable results, their method can only handle triplet extraction for one relation at a time which overlooks the dynamic nature of conversations wherein multiple topics and information can coexist. For instance, during a discussion about travel plans, there may be multiple topics discussed such as transportation plans, sights to visit, local cuisine, etc.

Chapter 4

Methodology

4.1 Task Definition

Given a dialogue with N utterances defined as $D = \{u_1, u_2, u_3, \ldots, u_N\}$, where even-numbered turns denote user utterances and odd-numbered turns represent the dialogue agent responses, we want to obtain persona attributes of the format *(subject, relation, object)* specific to the user. We design a two-stage persona attribute extractor, consisting of models for *relation prediction* and *entity extraction*. We define relation prediction as a multi-label classification problem since there can be multiple attributes that are detected in the dialogue. If a relation is triggered, the entity extractor then identifies the subject and object associated to that specific relation.

4.2 Datasets

Persona-Chat

Persona-Chat [52] is a crowd-sourced dialogue dataset collected by having a pair of speakers chat and engage with each other while conditioning their dialogues on their assigned persona. Each persona is consists of 4 to 6 profile sentences, while each dialogue consists of 6 to 8 turns, each containing a maximum of 15 words. An example dialogue from the dataset is shown in Table 4.1. The dataset has a total of 162,064 utterances over 10,907 dialogues that is then split into 131,438 utterances over 8,939 dialogues for training, 15,602 utterances over 1,000 dialogues for validation, and 15,024 utterances over 968 dialogues for testing. Additionally, for personas, there are a total of 1,115 possible personas, with 955 designated for training, 100 for validation, and 100 for testing.

Persona 1	**Persona 2**
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat Cheetos	I love watching Game of Thrones
Speaker 1: Hi	
Speaker 2: Hello! How are you doing today?	
Speaker 1: I am good thank you, how are you.	
Speaker 2: Great, thanks ! My children and I were just about to watch Game of Thrones.	
Speaker 1: Nice ! How old are your children?	
Speaker 2: I have for that range in age from 10 to 21. You ?	
Speaker 1: I do not have children at the moment.	
Speaker 2: That just means you get to keep all the popcorn for yourself.	
Speaker 1: And Cheetos at the moment!	
Speaker 2: Good choice. Do you want Game of Thrones?	
Speaker 1: No, I do not have much time for TV.	
Speaker 2: I usually spend my time painting: but I love the show.	

Table 4.1: Example dialogue from Persona-Chat. Persona 1 is assigned to Speaker 1 while Persona 2 is assigned to Speaker 2.

Dialogue-NLI

Dialogue-NLI [48] is a dataset built upon Persona-Chat for improving the consistency of dialogue models. It consists of sentence pairs that can be labeled either as *entailment*, *neutral*, or *contradiction*. Given a persona P comprised of profile sentences for each speaker $P_A = \{p_1^A, \dots, p_n^A\}$ and $P_B = \{p_1^B, \dots, p_n^B\}$ and a dialogue defined as a sequence of utterances $D = \{u_1^A, u_2^B, u_3^A, u_4^B, \dots, u_T^B\}$, the goal of training a model on Dialogue-NLI is to understand the logical relationship between utterances in a dialogue context. Specifically, this involves identifying utterances that contradict a previous utterance or an agent's persona by comparing the triplets *(subject, relation, object)* of input pairs (u, u') or (u, p). As shown in Table 4.2, a pair of inputs are classified as having an entailment relationship if they share the exact same triplet. If they share the same relation but have a different subject or object in the triplet, they are classified as neutral. Otherwise, they are classified

Sentence Pairs	**Triplets**	**Label**
i am only 22 so i would not know . i am twenty two years old	(i, `has_age`, 22) (i, `has_age`, 22)	Entailment
i play guitar in my spare time . i play the violin .	(i, `has_ability`, play instrument) (i, `has_ability`, play instrument)	Entailment
i work at a daycare . i work for a government agency .	(i, `employed_by_general`, daycare) (i, `employed_by_general`, government)	Neutral
my locks are chesnut . i am blonde	(i, `physical_attribute`, brunette) (i, `physical_attribute`, blonde)	Neutral
my vehicle is an older model car . i have pets .	(i, `have_vehicle`, car) (i, `have_pet`, pets)	Contradiction
my favorite band is imagine dragons . i am from texas	(i, `favorite_music_artist`, imagine dragons) (i, `place_origin`, texas)	Contradiction

Table 4.2: Examples from the Dialogue-NLI train set.

as contradiction. Triplet annotations (s, r, o) were obtained through crowd-sourcing using Amazon Mechanical Turk where the relation r is chosen from a predefined list of relation types. A comprehensive list of the relation types from Dialogue-NLI is provided in Appendix A.

Older Adults Interview Dataset

As discussed in Section 3.2, research specifically focused on the extraction of structured persona attributes is rather limited, with prior studies combining the Persona-Chat and Dialogue-NLI datasets to overcome the lack of available data. It should also be emphasized that both Persona-Chat and Dialogue-NLI were created for an entirely different task i.e., building personalized conversational agents, and so its triplet annotations many not precisely fulfill the objectives of this study. As shown in Table 4.1, some of the dialogues may appear unnatural such that there is not much engagement between the speakers and they are more focused on talking about themselves [52]. Moreover, the characteristics of the dialogues from these datasets are also inherently different from those typically observed from older adults.

To accommodate the specific needs of our research, we utilize transcripts from an interview podcast called Senior Storytelling[1] as our dataset. Podcast episodes were transcribed using Descript[2] transcription. The podcast is an initiative from the Elderly Embrace Care Network[3], a youth-led senior-focused nonprofit media organization that aims to foster intergenerational dialogue and culture respect for the elderly. The podcast features a conversation between a senior and their grandchild, during which they reminisce about the senior's childhood experiences and share wisdom gained from a lifetime of lived experiences. Through storytelling, the interview serves as an opportunity for seniors to reflect on their own identity and help the interviewer understand the senior's identity. We show descriptive statistics on the dataset in Table 4.3 and present the list of interview questions featured in the podcast in Table 4.4.

	Avg num of turns	**Avg num of words**	**Max num of words**
Mr. Mao	32	68.22	250
Mr. Stamper	35	139.6	717
Mrs. Mean	36	107.56	349
Mrs. Rechenmacher	28	174.03	691

Table 4.3: Statistics from the OLDER ADULTS INTERVIEW DATASET. Reported word counts were measured per utterance.

4.3 Relation Predictor

Classification tasks can be categorized based on the number of class labels that can be assigned to each sample or data point. In binary classification, each sample can be assigned to one of two mutually exclusive classes $\{0, 1\}$. An example of binary classification in real-world applications is spam email detection where the model is tasked to classify emails into one of two categories: "spam" or "not spam". Multi-class classification extends this approach to scenarios where a sample can belong to one of k classes $\{0, 1, \ldots, k-1\}$, with each class being mutually exclusive. Natural language inference (NLI), introduced in Section 4.2, is an example of a multi-class classification problem. By training a model on an NLI dataset such as DIALOGUE-NLI, it develops the ability to discern relationship (i.e., entailment, neutrality, contradiction) between sentences. A more general formulation would be to remove the mutual exclusivity constraint, resulting in multi-label classification.

[1]Senior Storytelling podcast: https://podcasters.spotify.com/pod/show/seniorstorytelling
[2]Descript: https://www.descript.com/transcription.
[3]Elderly Embrace Care Network website: https://elderlyembrace.org/

	Interview Prompts/Questions
1	Introduce yourself.
2	When you were younger, who was your model and why?
3	If you could relive one year or decade of your life, what would it be and why?
4	If you could speak to your younger self, what advice would you give?
5	Was there anything that you wanted to do but never did? Why did you never do said things?
6	Do you have any regrets from your younger years that you would feel is a good lesson for the younger generation?
7	What advice did you get that you wish you would've listen to or taken more seriously?
8	Are there any habits or skills that you regret not learning or picking up?
9	What is something about today's youth that really surprises you or is fundamentally different from when you were growing up?
10	Is there something about today's world that you never expected to become this way?
11	What is something that was common in your youth that is now obsolete?
12	What has changed most in society since when you were in college?
13	What were your fondest memories of school/college? How do you think it compares to today's education?
14	What is something that is available today that you would have liked to have in your childhood?
15	What is something that you really enjoy doing and brings you joy?
16	Do you have a go-to comfort food? Has that change over time?
17	Do you have any funny stories that were legal back then but not anymore?
18	What are some things you do to relax or take your mind off of things?
19	What was the biggest milestone in your life to date?
20	What is your best childhood memory?
21	What was your favourite thing to do when you were younger?
22	What career would you have pursued if you if there wasn't any concern with money or practicality?
23	How did you meet your spouse? What do you remember most about your wedding?
24	What is the easiest or hardest part about getting older?
25	After seeing how the world has changed until now, do you have any predictions on what the future will look like?
26	What did you think 2020 was going to be like when you were younger?
27	Do you have any advice for kids considering how much technology really impacts their relationships?
28	Do you have a message to share with other seniors?

Table 4.4: Interview questions from the Senior Storytelling podcast.

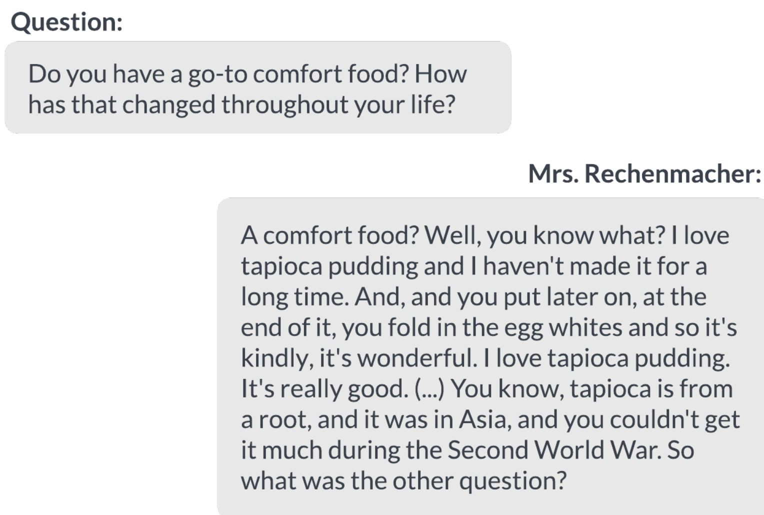

Figure 4.1: Transcript excerpt from the Older Adults Interview Dataset.

In this scenario, samples are not constrained to just a single class; rather, they can be associated to multiple classes simultaneously. Object detection in images is an example of multi-label classification in the computer vision domain. This task involves identifying and categorizing the contents of an image which is particularly significant for applications in medical imaging and self-driving technologies.

Prior research on persona attribute extraction primarily focused on processing input information at the utterance level under an implicit assumption that only a single relation is present in each utterance. However, by the very nature of dialogues, it is common for topics, ideas, and information to coexist and interplay, reflecting the complexity of human interactions. This becomes particularly pronounced in older adults and especially for persons with dementia when higher cortisol functions i.e., processes related to language, memory, and perception, decline with old age. We can observe from the excerpt from the Older Adults Interview Dataset shown in Figure 4.1 discourse impairments such disruptive topic shifts, repetitions, and empty phrases. A study conducted by Dijkstra et al. [13] revealed that persons with dementia experience more difficulties maintaining conversation than cognitively healthy older adults. Considering these factors, we therefore approach persona attribute extraction as a multi-label classification task.

The input $\boldsymbol{x}$ to the model is derived by converting each word in the input sequence (i.e., unstructured text) into high-dimensional vector representations called *embeddings*. After obtaining the embedding for each word in the input text, they are aggregated to produce a single vector representation of the entire sentence. For k given relations, the relation predictor $f : \mathbb{R}^{768} \rightarrow [0, 1]^k$ computes $\hat{y} = f(\boldsymbol{x})$. For each relation r, $\hat{y}_r$ represents $\Pr(y_r = 1|\boldsymbol{x}) = \hat{y}_r$ i.e., the estimated probability of r being assigned to the input. Since this is a multi-label classification task, $\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_k\}$ are independent. For example, given Mrs. Rechenmacher's response in Figure 4.1, we expect the model to predict the relation `like_food`.

4.4 Entity Extractor

Information extraction (IE) is a broad concept in natural language processing covering a wide range of tasks concerned with extracting structured information from unstructured text data. Some commons tasks in IE include named entity recognition (NER), relation extraction (RE), and event extraction (EE). Prior to the popularity of deep learning, information extraction systems heavily relied on handcrafted rules and patterns that required domain expertise and significant human effort and labour to develop. The adoption of deep learning technologies and more recently large language models (e.g., GPT-4 [1], Llama [45], Flan-T5 [9]) further propelled IE research due to their advanced capability to understand, generate, and generalize text.

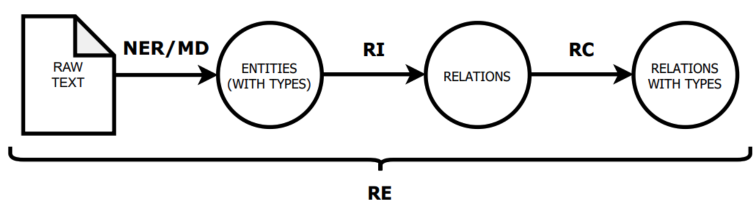

Figure 4.2: Relation extraction pipeline from Bassignana and Plank [5].

The primary task concerned with the extraction of triplets in the format of (*subject*, *relation*, *object*) is called relation extraction. As shown in Figure 4.2, relation extraction is usually performed as a series of sub-tasks that starts with identifying named entities or nominals, i.e., group of words that function as nouns, from the input text either by NER or mention detection (MD). This step is followed by relation identification (RI) which entails finding pairs of entities that potentially have a semantic relationship through binary classification. Upon obtaining paired entities, they are assigned a relation r usually through multi-class classification. This final step corresponds to the relation classification (RC) sub-task shown in the diagram. As highlighted by Bassignana and Plank [5], there is no definitive approach for extracting relations. Some studies work with the assumption that a relation exists between every pair of entities, thus reducing the task to only RC. While other studies introduce a `no_relation`/`no-rel` label, merging RI and RC into a single step.

Previous approaches that we explored for persona attribute extraction such as topic modeling using BERTopic [21] and keyword extraction using YAKE [8] struggled in effectively identifying topics/keywords/entities given the length and noisy structure of the Older Adults Interview Dataset. The lack of dependable entities hindered our advancement through the pipeline or other downstream tasks such as topic-focused summarization. Therefore, considering these factors and findings from our preliminary experiments, we find that reversing the typical RE pipeline i.e., predicting all possible relations without pre-identifying entities, more suitable for our specific use case.

Figure 4.3: Triplet extraction as template infilling sample from Kim et al. [29].

We adopt the framework proposed by Kim et al. [29] for this task by reformulating entity extraction as a template infilling task; therefore, aligning the task objective with the pre-training objective of the language model. In their original work, this approach is implemented using T5 [39], a model that employs a text-to-text framework such that both the input and output is a sequence of text. Figure 4.3 illustrates a sample input to the model. For each relation r, a relation template t_r is constructed in the format "`<X>` *relation* `<Y>`" where `<X>` is a placeholder for the subject and `<Y>` is for the object. For example, the template for the relation *dislike* is "`<X>` does not like `<Y>`". The input $\boldsymbol{x}$ to the model is constructed by concatenating the context c and t_r. In our study, c would be the dialogue consisting of sequence of utterances. The T5 model produces `<X>` and `<Y>` given $\boldsymbol{x}$. It is fine-tuned such that it maximizes the likelihood of $P(\texttt{<X>}, \texttt{<Y>} | c, t_r)$ using the combined Persona-Chat-Dialogue-NLI dataset. Because `<X>` and `<Y>` can span multiple tokens, T5 auto-regressively generates a series of tokens for `<X>` and `<Y>`.

Chapter 5

Experimental Design

5.1 Data Preprocessing

We begin with preprocessing DIALOGUE-NLI by combining relations that have a similar semantic meaning such as `favourite_activity` and `like_activity` or `favourite_hobby` and `has_hobby` to eliminate ambiguity. We exclude relations such as `gender`, `has_age`, and `nationality` as they are not relevant to our intended use case. We provide the relation type mappings in Table 5.1 and the list of relation types used in this study in Table 5.2. We preprocess PERSONA-CHAT by merging the initial 95/5 train and validation split then redividing it to create the 80/10/10 split for the train, validation, and test sets resulting to 5,539 samples for train and 1,187 samples each for validation and test. We can observe from Figure 5.1 that there is a severe imbalance between the relation types. From a multi-label classification perspective, there is also more negative samples than positive ones for each relation type. Afterwards, we map the triplet annotations form DIALOGUE-NLI to the dialogues in PERSONA-CHAT. We discard triplets that are missing subject and object annotations. Table 5.3 shows the descriptive statistics of the combined PERSONA-CHAT-DIALOGUE-NLI dataset. We can observe that these dialogues are quite short in length such that the discussion between the two speakers are rather straightforward as shown in an earlier example in Table 4.1. It is important to note that the characteristics of this dataset is different from what might be observed in a dialogue with older adults or persons with dementia. Additionally, we only retain utterances from Speaker 2 as we are only interested in the responses to the interview questions in Table 4.4. Finally, we format the datasets for each subtask by converting triplets annotations into multi-hot vectors for multi-label classification and performing the necessary tokenization for the entity extraction subtask.

	Original relation	Mapped relation
1	favorite_activity	like_activity
2	favorite_hobby	has_hobby
3	favorite_animal	like_animal
4	favorite_book	like_read
5	favorite_drink	like_drink
6	favorite_food	like_food
7	favorite_movie	like_watching
8	like_movie	like_watching
9	favorite_show	like_watching
10	favorite_music	like_music
11	favorite_music_artist	like_music
12	favorite_sport	like_sports
13	have_children	have_family
14	have_sibling	have_family
15	employed_by_general	has_employment
16	employed_by_company	has_employment
17	teach	has_employment
18	previous_profession	has_employment
19	has_profession	has_employment
20	want_do	want
21	has_degree	attend_school
22	live_in_general	live_in_citystatecountry

Table 5.1: Relation type mappings

	Relation types
1	attend_school
2	dislike
3	favorite_color
4	favorite_place
5	favorite_season
6	has_ability
7	has_employment
8	has_hobby
9	have_family
10	have_pet
11	have_vehicle
12	job_status
13	like_activity
14	like_animal
15	like_drink
16	like_food
17	like_general
18	like_goto
19	like_music
20	like_read
21	like_sports
22	like_watching
23	live_in_citystatecountry
24	marital_status
25	own
26	place_origin
27	want
28	want_job

Table 5.2: Relation types used in this study

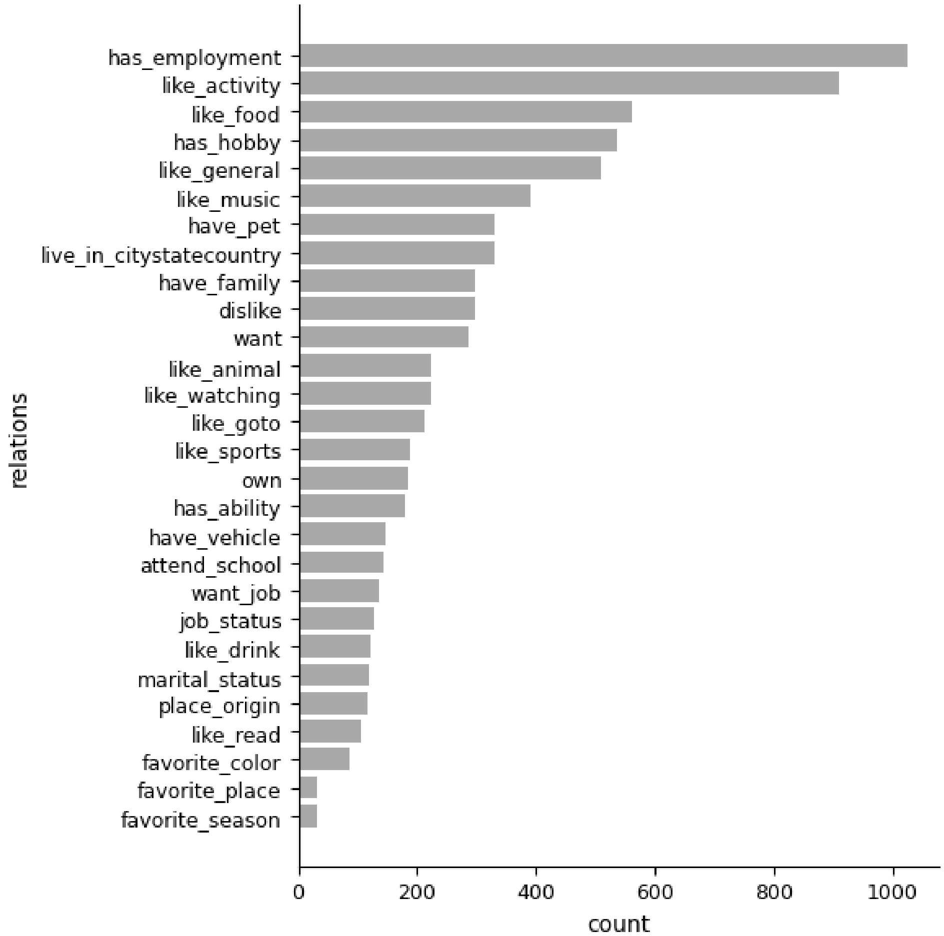

Figure 5.1: Distribution of relations in the train set.

	Train	**Val**	**Test**
Avg number of relations	1.42	1.42	1.41
Avg number of words	80.21	79.58	80.77

Table 5.3: Statistics from the combined Persona-Chat-Dialogue-NLI dataset. Reported word counts were measured per dialogue.

5.2 Experiments

5.2.1 Relation Predictor

In this section, we present a multi-label classification approach for predicting relations within a dialogue context. As a baseline, we fine-tune a pretrained BERT model for multi-label sequence classification, specifically `bert-base-uncased`[1] from HuggingFace, on the combined PERSONA-CHAT–DIALOGUE-NLI dataset. This entails introducing an additional fully connected layer to the original BERT model with a sigmoid activation over the final hidden state corresponding to the `[CLS]` token. In simpler terms, the summary representation represented by the `[CLS]` token is used as input to the fully connected layer to make a classification prediction. We fine-tune the BERT model for 20 epochs with a training batch size of 32 and a learning rate of 2×10^{-4}. Although memory-intensive, fine-tuning is still the prevalent technique for adapting transformer-based models like BERT for downstream tasks. This approach allows for efficient transfer learning, enabling the model to generalize to new tasks with minimal task-specific labeled data.

We present a variety of models designed for this task and detail implementation specifics below. Our experimental setup consists of two primary types of models: a standard feedforward neural network and a feedforward neural network with residual connections, as depicted in 2.2, inspired by the ResNet framework proposed by He et al. [25]. The final layer of each network utilizes a sigmoid activation function, as our task involves multi-label classification. Henceforth, we refer to these models as NN_{plain} and $\text{NN}_{\text{residual}}$ respectively.

For the loss function, we conduct experiments with both binary cross entropy (BCE) and BCE with class weights. In classification tasks, class weights are used to assign importance to different classes to help the model handle imbalanced datasets. We compute class weights by normalizing the inverse of the frequency of each class in the train dataset. We also evaluate the effects of dropout regularization with dropout rates 0.2 and 0.5. We use the Adam optimizer [30] with its default learning rate of 1×10^{-3}. All models were trained for 1,000 epochs with a batch size of 32 and a learning rate of 2×10^{-4} across 5 different seeds. Finally, we evaluate the performance of the models using F_1 score and select the best model based on the highest F_1 validation score.

All models use a ReLU activation function for the hidden layers and a sigmoid activation function in the final layer. A summary of the model variants used in the experiments is provided below.

[1]BERT base model (uncased) – https://huggingface.co/google-bert/bert-base-uncased

- $\text{NN}_{\text{plain2}}$: a feedforward network composed of 2 hidden layers.
- $\text{NN}_{\text{plain4}}$: a feedforward network composed 4 hidden layers.
- $\text{NN}_{\text{residual3}}$: a feedfoward network composed of 4 hidden layers and 3 residual connections between every other layer.
- $\text{NN}_{\text{residual5}}$: a feedfoward network composed 6 hidden layers and 5 residual connections between every other layer.
- $\text{NN}_{\text{residual3\_dropout}}$: a feedfoward network composed of 4 hidden layers and 3 residual connections between every other layer, with dropout applied at each residual connection.
- $\text{NN}_{\text{residual5\_dropout}}$: a feedfoward network composed 6 hidden layers and a 5 residual connections between every other layer, with dropout applied at each residual connection.

5.2.2 Entity Extractor

Implementation Details. Following the framework proposed by Kim et al. [29], we create a relation template for each of the 28 relations in our dataset as shown in Table 5.4. We use a pretrained T5 model, specifically `t5-base`[2] from HuggingFace. We also conducted experiments using `t5-small`[3] which is roughly 73% smaller than `t5-base` with only 60 million parameters. We fine-tune the model for 3 epochs with a training batch size of 64, and a learning rate of 3×10^{-5} and tune the hyperparameters on the validation set across 5 different seeds. We also experiment with {128, 256} for the maximum input length given the difference in dialogue length between our datasets. At inference, the output tokens are limited to tokens from the input sentence. We use a beam size of 4 to generate a maximum of 4 entity pairs for a given relation each with a score computed by $P_{T5}(x|y)$ where the input x is the utterance and template of the predicted relation concatenated together and y is the output sequence consist of the subject and the object. This score is used to rank the generated triplets.

[2]Google T5-base model: https://huggingface.co/google-t5/t5-base
[3]Google T5-small model: https://huggingface.co/google-t5/t5-small

	Relation	Template
1	attend_school	<X> attend school in <Y>
2	dislike	<X> doesn't like <Y>
3	favorite_color	<X> favorite color is <Y>
4	favorite_place	<X> favorite place is <Y>
5	favorite_season	<X> favorite season is <Y>
6	has_ability	<X> can or has the ability to <Y>
7	has_employment	<X> is working at or for <Y>
8	has_hobby	<X> hobby is <Y>
9	have_family	<X> family member is <Y>
10	has_pet	<X> has a pet <Y>
11	have_vehicle	<X> own vehicle <Y>
12	job_status	<X> job status is <Y>
13	like_activity	<X> likes to do <Y>
14	like_animal	<X> likes the animal <Y>
15	like_drink	<X> likes to drink <Y>
16	like_food	<X> likes to eat <Y>
17	like_general	<X> likes <Y>
18	like_goto	<X> likes to go to <Y>
19	like_music	<X> likes to listen to <Y>
20	like_read	<X> likes to read <Y>
21	like_sports	<X> likes playing the sport <Y>
22	like_watching	<X> likes to watch the movie or show <Y>
23	live_in_citycountrystate	<X> resides in <Y>
24	marital_status	<X> marital status is <Y>
25	own	<X> own <Y>
26	place_origin	<X> is originally from <Y>
27	want	<X> wants <Y>
28	want_job	<X> want to work as or at <Y>

Table 5.4: Examples of relations and templates

Evaluation Metrics. For automatic evaluation, we report the accuracy of the extracted triplets compared to the ground truth triplets. The extracted triplet is considered a true positive if and only if the subject, relation, and object is an exact match of the ground truth. However, given that the annotated triplets from DIALOGUE-NLI may contain inaccuracies, i.e., the subject or object may not be found in the dialogue, this error may propagate into the creation of the test dataset through distant supervision. Referencing an example from Table 4.2, the sentence "*I play the violin*" was annotated with the triplet (i, `has_ability`, play instrument). The object "play instrument" cannot be found from the sentence therefore, data points with out-of-vocabulary tokens are ignored.

5.3 Results and Analysis

In this section, we present results on evaluating the relation predictor and entity extractor as separate models, followed by an assessment of the models combined as a two-stage attribute extractor for persona attribute extraction.

5.3.1 Relation Predictor

For the relation predictor subtask, we employ micro-averaging to compute the precision, recall, and F_1 scores. In micro-averaging, we aggregate the total counts of true positives (TP), false positives (FP), and false negatives (FN) across all classes (i.e., relation types). Then, we calculate the overall precision, recall, and F_1 scores using the combined counts across all classes. This approach ensures that each prediction and true label receive equal weight, thus effectively evaluating each data point equally. Micro-averaged precision, recall, and F_1 score are calculated as follows:

$$\text{Precision}_{micro} = \frac{TP_{total}}{TP_{total} + FP_{total}}$$

$$\text{Recall}_{micro} = \frac{TP_{total}}{TP_{total} + FN_{total}}$$

$$\text{F1}_{micro} = \frac{2 \times \text{Precision}_{micro} \times \text{Recall}_{micro}}{\text{Precision}_{micro} + \text{Recall}_{micro}}$$

We use the micro-averaging method included in scikit-learn[4] when computing for these metrics to account for the imbalance in the dataset.

As a baseline, we consider a random classifier that predicts the presence of relation r_i with probability p_i, where p_i represents the fraction of instances in the dataset where the relation r_i is present. For example, based on Figure 5.1, 1,026 out of 5,539 train instances have the `has_employment` relation, then $p_{\text{has\_emplyoment}} = \frac{1026}{5539} \approx 0.185$. Similarly, we can compute $p_{\text{favorite\_season}} = \frac{31}{5539} \approx 0.006$. Under micro-averaging, the expected precision for this random classifier can be computed by

$$\text{precision} = \frac{\Sigma_{i=1}^{28} p_i^2}{\Sigma_{i=1}^{28} p_i^2 + (1 - p_i)p_i}$$

We can compute the recall using the same formula because the number of false positive equals the number of false negatives. Since precision is the same as the recall, the F_1 score is also the same since it is the harmonic mean of the two metrics.

We show the results of the relation prediction task without the utilization of class weights in Table 5.5. We can observe that the models generally achieves higher recall compared to the precision which implies that the models capture positive instances from the dataset at the expense of capturing false positives. The BERT$_{finetuned}$ model, as expected, is underfitting to the data due to the limited complexity of the model such that there is only one linear layer on top of the pre-trained architecture. Therefore, experimenting with dropout on a fine-tuned model is not possible.

Observing the effects of dropout, there is no strong relationship between $p_{dropout}$ and the F_1 scores in general. However, consider these two models $\text{NN}_{\text{residual3\_dropout}}$ with $p_{dropout}$ = 0.2 and $\text{NN}_{\text{residual3}}$ with $p_{dropout} = 0.5$. We define the *effective dropout rate* as the net information loss as a result of all dropout layers from the input layer to the final output layer. The effective dropout rate for $\text{NN}_{\text{residual3\_dropout}}$ with $p_{dropout} = 0.2$ can be computed as $1 - (1 - 0.2)^3 = 0.488$, while for $\text{NN}_{\text{residual3}}$ with $p_{dropout} = 0.5$, it is trivially 0.5 since there is only a single dropout layer. While both models have roughly similar effective dropout rate, the models achieved different F_1 scores: 0.1810 and 0.1467 respectively. This suggests that decomposing dropout layers might improve performance because the model is not subjected to a sudden loss of information and yet promotes better model generalization.

Table 5.6 summarizes the results for the relation prediction task where we incorporate class weights into the loss function. We continue to observe the same trend between precision

[4]Scikit-learn `precision_recall_fscore_support`: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html

Model	$p_{dropout}$	**Prec**	**Rec**	F_1
Random classifier	-	0.0864	0.0864	0.0864
Embedding Similarity	-	0.0517	0.0517	0.0517
BERT$_{finetuned}$	-	0.0725	0.0517	0.0603
$\text{NN}_{\text{plain2}}$	0.2	0.1433	0.3311	0.2000
$\text{NN}_{\text{plain4}}$	0.2	0.1226	0.3474	0.1813
$\text{NN}_{\text{residual3}}$	0.2	0.1427	0.1791	0.1588
$\text{NN}_{\text{residual5}}$	0.2	0.1257	0.2212	0.1603
$\text{NN}_{\text{residual3\_dropout}}$	0.2	0.1305	0.2951	0.1810
$\text{NN}_{\text{residual5\_dropout}}$	0.2	0.1114	0.3480	0.1689
$\text{NN}_{\text{plain2}}$	0.5	0.1419	0.3846	0.2074
$\text{NN}_{\text{plain4}}$	0.5	0.0944	0.3425	0.1480
$\text{NN}_{\text{residual3}}$	0.5	0.1266	0.1743	0.1467
$\text{NN}_{\text{residual5}}$	0.5	0.1433	0.1863	0.1620
$\text{NN}_{\text{residual3\_dropout}}$	0.5	0.1321	0.4303	0.2021
$\text{NN}_{\text{residual5\_dropout}}$	0.5	0.1000	0.4441	0.1632

Table 5.5: Results for the relation predictor on the combined Persona-Chat-Dialogue-NLI test set without class weights.

Model	$p_{dropout}$	**Prec**	**Rec**	F_1
BERT$_{finetuned}$	-	-	-	-
$\text{NN}_{\text{plain2}}$	0.2	0.1387	0.3498	0.1986
$\text{NN}_{\text{plain4}}$	0.2	0.1080	0.4044	0.1705
$\text{NN}_{\text{residual3}}$	0.2	0.1349	0.1899	0.1577
$\text{NN}_{\text{residual5}}$	0.2	0.1432	0.1809	0.1599
$\text{NN}_{\text{residual3\_dropout}}$	0.2	0.1316	0.2644	0.1758
$\text{NN}_{\text{residual5\_dropout}}$	0.2	0.1247	0.3161	0.1788
$\text{NN}_{\text{plain2}}$	0.5	0.1502	0.3810	0.2155
$\text{NN}_{\text{plain4}}$	0.5	0.0925	0.3401	0.1454
$\text{NN}_{\text{residual3}}$	0.5	0.1413	0.1857	0.1605
$\text{NN}_{\text{residual5}}$	0.5	0.1277	0.1911	0.1531
$\text{NN}_{\text{residual3\_dropout}}$	0.5	0.1180	0.4213	0.1844
$\text{NN}_{\text{residual5\_dropout}}$	0.5	0.1247	0.4393	0.1942

Table 5.6: Results for the relation predictor on the combined Persona-Chat-Dialogue-NLI test set incorporating class weights.

	BCE		BCE w/ class weights	
	$p_{dropout} = 0.2$	$p_{dropout} = 0.5$	$p_{dropout} = 0.2$	$p_{dropout} = 0.5$
$\text{NN}_{\text{plain2}}$	0.2000	0.2074	0.1986	0.2155
$\text{NN}_{\text{plain4}}$	0.1813	0.1480	0.1705	0.1454
$\text{NN}_{\text{residual3}}$	0.1588	0.1467	0.1577	0.1605
$\text{NN}_{\text{residual3\_dropout}}$	0.1810	0.2021	0.1758	0.1844
$\text{NN}_{\text{residual5}}$	0.1603	0.1620	0.1599	0.1531
$\text{NN}_{\text{residual5\_dropout}}$	0.1688	0.1632	0.1788	0.1942

Table 5.7: Relation predictor F_1 scores on the test dataset.

and recall from Table 5.5. In terms of the effect of dropout, the improvements are more evident compared to the earlier table except for models $\text{NN}_{\text{plain2}}$ and $\text{NN}_{\text{residual5}}$.

In Table 5.7, we present a comparison of the effects achieved by incorporating class weights into the loss function for mitigating data imbalance. We can observe that incorporating the class weights by normalizing by the inverse of the frequency does not always lead to an improved F_1 score. This suggests that there may be other factors influencing the models' performance, apart from data imbalance. In terms of model architecture, residual learning does not improve model performance. This suggests that NN_{plain} does not suffer from vanishing gradients which is not surprising because there is not many hidden layers. As a result, in introducing residual connections potentially incurs redundancies from the addition with identity functions.

5.3.2 Entity Extractor

Table 5.8 shows the results for evaluating the entity extractor given ground truth relations as input. For input sequences of 128 tokens, it is evident that `t5-base` outperforms `t5-small`, as expected due to it having more parameters and thus, its ability to capture complex patterns and relationships in the data. However, we encounter a somewhat unexpected outcome for input sequences of 256 tokens, where `t5-small` achieves better results compared to `t5-base`. Given that the dialogues in the combined Persona-Chat-Dialogue-NLI are short, as shown in Table 5.3, a shorter sequence length is likely adequate. We hypothesize that a model capable of processing longer sequences may be more suitable for the Older Adults Interview Dataset, given the lengths of each utterance and dialogue as illustrated in Table 4.3.

Model	Input seq length	Acc
`t5-small`	128	0.3379
	256	0.1351
`t5-base`	128	0.4400
	256	0.0452

Table 5.8: Entity extractor F_1 scores on the test dataset.

5.3.3 Pipeline

We select the best model from the relation predictor and entity extractor experiments from earlier sections and present the results below. First, we discard incorrect predictions produced by the first model in the pipeline, i.e., relation predictor, since there is no ground truth triplet for such predictions. Essentially, only correct predictions are forwarded to the entity extractor to complete the persona attribute extraction pipeline for evaluation sake. From the results in Table 5.9, we can observe the same trend as before where the more complex `t5-base` achieves better results compared to `t5-small`.

Model	Input seq length	Acc
$\mathrm{NN}_{\mathrm{plain2}}$ ($p_{dropout=0.5}$) + `t5-base`	128	0.5124
$\mathrm{NN}_{\mathrm{plain2}}$ ($p_{dropout=0.5}$) + `t5-small`	128	0.3632

Table 5.9: F_1 scores of the complete persona attribute extractor pipeline on the test dataset.

Chapter 6

Case Study: Testing Methodology on Older Adults

6.1 Experiment with Older Adults Interview Dataset

With very limited data, it is difficult to train models that will directly be applicable for a dementia use case. Balancing the trade-off between the quantity and quality of data is crucial, as it directly impacts the effectiveness of solving the problem at hand. For instance, while large language models are typically trained on vast amounts of data, not all of it may be relevant to the problem we aim to address. Therefore, fine-tuning the model with the appropriate data is crucial in improving model performance and suitability for the desired task. As a step towards adapting our model for a dementia use case, we train and fine-tune our models on datasets that more closely reflects the process of extracting persona attributes from interviews with persons with dementia. In this chapter, we conduct further experiments with our methodology using the Older Adults Interview Dataset.

As discussed in Section 4.2, Persona-Chat and Dialogue-NLI are datasets created for an entirely different task that is not persona attribute extraction. Unlike human interactions, which often feature spontaneity, depth, and nuanced communication, the dialogues in Persona-Chat lack these essential qualities found in genuine human conversations. The Older Adults Interview Dataset captures these fundamental characteristics of human interaction. As discussed in the introduction, conducting prototype testing with persons with dementia poses significant challenges. Thus, using this dataset is an essential first step in that direction as this allows us to evaluate our model on data that more closely reflects the linguistic complexities observed in persons with dementia. We examine these linguistic characteristics and their implications on our methodology in detail in Section 6.2.

For the remainder of this section, we will discuss our experiments on the Older Adults Interview Dataset. We begin by matching the annotations provided by our annotators with the appropriate relations from Dialogue-NLI to obtain ground truth triplets. In Table 6.1, we show examples of persona attributes extracted for Mrs. Rechenmacher. We can observe that the model was able to extract the persona attributes <i, `has_hobby`, knitting> and <i, `like_food`, tapioca pudding> that are inline with the annotations. We can also identify two errors made by the model. Due to the length of the utterance being more than the maximum allowed input for T5, the model failed to extract the triplet corresponding to Mrs. Rechenmacher's sister. In the third example, the relation predictor fails to correctly identify the appropriate relation and this error propagates to the entity extractor. Recall from Figure 5.1 that there is severe class imbalance in our dataset. Specifically, the relation `like_activity` has significantly more positive examples compared to the relation `dislike` therefore, it is possible that the model is more biased to predicting the former relation. We hypothesize that the class weights may required further calibration.

In Table 6.2, we can observe that the model struggles more in less coherent utterances. Analyzing the extracted triplet <i, `have_family`, mother>, the entity extractor identifies "mother" as the object entity as it is the best choice in the context of the utterance and the predicted relation. Given that "wife" is not found in the utterance, the model is unable to extract it as part of the triplet. This shows the limitations of the approach where the model is unable to extract implicit or abstract persona attributes. For the last example in the table, the model extracts a persona attribute that does not exactly match the ground truth triplet, but, however, is still considerably accurate in the context of Mr. Mao's response. Therefore, we suggest that evaluation metrics (e.g., BLEU score, human evaluation) other that F_1 score should be considered in the future.

Table 6.3 shows examples of persona attributes extracted for Mrs. Mean. In the first example, the model predicted `attend_school` because there are multiple words in the utterance associated with schooling. This prediction is incorrect and this error is propagated to the entity extractor which identifies "Sycrause" as the object. The model was able to correctly predict the triplet <i, `has_hobby`, reading> but does not capture the second triplet (e.g., hanging out with my kids). Based on previous examples, we hypothesize this is caused by disfluencies in the utterance (e.g., "hanging out with my kids" is not a complete thought or sentence). In the last example, the model ultimately fails to predict any triplet for the same reason. Moreover, we see the effect of the combined Persona-Chat-Dialogue-NLI dataset on the model training. Our model can capture clearer sentences like "I love tapioca pudding" but fails when dealing with incomplete clauses.

Annotator Summary	
Mrs. Rechenmacher is a passionate and positive person – a mother and homemaker who values a large family. She grew up in the Santa Clara valley and enjoys all the outdoor activities available in the region. She also values travel and adventure, helping others, and education. Her life is shaped by WW2 and she classifies all events of her life as pre ,during or post war. She admires the work of F.D. Roosevelt, fondly remembers a wooden playhouse she had as a child and later passed on to her own children and is very proud of a little red sweater she knit for her sister when she was a child.	
Human Annotation / Code	**Extracted Persona Attribute**
`has_hobby` : knitting `have_family` : sister	<i, `has_hobby`, knitting>
Q: Was there anything that you wanted to do but never did? And why did you never do said things? Mrs. R: Well, I like to make things. Right now I'm knitting and I find that when it's cold or when you can't go outside, it's nice to have something, has some handwork, something, whatever you like to do at embroidery or knit or read or write or anything to fill the cracks in the day. (...) But I was only, I had a sister. When I was nine years old, my mother had a little girl, my sister, and when she was three years old and I was 12 years old, I knit her a sweater, a little red sweater. I'll never forget it, and I'm so proud of that in my own heart. So I think the things we can create are just very, very important. Yes.	
Human Annotation / Code	**Extracted Persona Attribute**
`like_food` : tapioca pudding	<i, `like_food`, tapioca pudding>
Q: It's always good to keep busy. Mm-hmm. Keeps our spirits up. Right. Do you have a go-to comfort food? How has that changed throughout your life? Mrs. R: A comfort food. Well, you know what? I love tapioca pudding and I haven't made it for a long time. And, and you put later on, at the end of it, you fold in the egg whites and, it's wonderful. I love tapioca, pudding. It's really good. I like it. Yeah, good for you too. (...) So what was the other question?	
Human Annotation / Code	**Extracted Persona Attribute**
`dislike` : Facebook	<i, `like_activity`, Facebook>
Q: Is there something about today's world that you never expected to become this way? Mrs. R: Well, I think the computer industry, there's good things about it, but there's a lot about it that I don't like. I don't, I don't like Facebook. I won't have any part of Facebook and cause people don't, young people especially don't know how to manage it. They tell too many of their secrets that they let people frighten them and get them depressed. You know, that wasn't the intention to start.	

Table 6.1: Example of extracted persona attributes for Mrs. Rechenmacher. Underlined phrases denote persona attributes based on the human annotations. Extracted persona attributes in red denote incorrect predictions by the model.

Annotator Summary
Mr. Mao is a positve, forward thinking, yet emotionally reserved person who values knowledge and education above all else. He worked in the field of Aeronautics. He has always had an interst in new techonology from a young age and is proud of his academic accomplishments, including writing and reading, but also teaching his class as a young student. He is active and values good health and well being and considers personal connection a part of this.

Human Annotation	**Extracted Persona Attribute**
`like_drink` : coffee	<i, `like_drink`, coffee>

Q: Do you have a go-to comfort food and how has that changed throughout your life? Uh, so like a food that is like a lot of meaning to you?

Mr. M: Uh, food. I would say drink. I would say drinking coffee would be very meaningful to me.

Human Annotation	**Extracted Persona Attribute**
`have_family` : wife	<i, `have_family`, mother>

Q: Uh, how did you meet your spouse and what do you remember most about your wedding?

Mr. M: I met your mom, mother in a gathering in one child's, uh, uh, home, and, uh, when I was young. This is the first time I met your mother.

Human Annotation	**Extracted Persona Attribute**
`like_general` : science, engineering, helping students	<i, `has_ability`, research>

Q: What is something that you really enjoy doing or something that brings you like joy in life?

Mr. M: Oh yeah. I found as a research or through the research discovery new, uh, things. That could be on the science or engineering, or this is make me happy. Another one is, again, most important helping students. I see the development, even though, even for me right now, is for looking at my kids. Development, the progressing of the kids, a progressing of the students I found is enjoyable and uh, is very important for me.

Table 6.2: Example of extracted persona attributes for Mr. Mao. Underlined phrases denote persona attributes based on the human annotations. Extracted persona attributes in red denote incorrect predictions by the model.

Annotator Summary
Mrs. Mean is a divorced grandmother who was a stay at home mom who enjoys spending time with her family. She expressed a high value on living your authentic life, respecting your elders, reading, movies, and her unrealized aspiration to be a teacher. Much of her interview was expressed in regrets due to her belief that her husband was the wrong person for her and the fact that she was not able to pursue her love of teaching.

Human Annotation	**Extracted Persona Attribute**
`want_job` : teach English, teach children	<i, `attend_school`, Sycrause>

Q: What career would you have pursued if you, if there wasn't any concern with money or practicality?

Mrs. M: I would have taught English. And Syracuse was the best teachers college of its time. Oh yeah. I would've loved to. Help children. Enjoy reading. And speak properly. That would have been my dream.

Human Annotation	**Extracted Persona Attribute**
`like_activity` : reading, hanging out with my kids	<i, `has_hobby`, reading>

Q: What's something that you really enjoy doing something that brings you a lot of joy in your life.

Mrs. M: Oh, I love sitting on my porch and reading. I could do that all day. That that, that, and hanging out with my kids. My greatest joy. If I could be with you guys, you know, every day. I would. But of course, you know, No.

Human Annotation	**Extracted Persona Attribute**
`like_activity` : watch TV, talk to my brother and sister, reading, hanging out with my dog `have_pet` : dog `like_watching` : movies	`None`

Q: Whare are some of the little things that you do to relax or take your mind off things?

Mrs. M: Um, I watch TV. I talk to my brother and sister. Hang out with my dog. Little things, pick up a book or get on my ipad watch a movie. I love movies.

Table 6.3: Example of extracted persona attributes for Mrs. Mean. Underlined phrases denote persona attributes based on the human annotations. Extracted persona attributes in red denote incorrect predictions by the model.

6.2 Aging, Cognition, and Language

In the previous chapters, we have discussed dementia and it's implications on identity, as well as various factors to consider in the development of technology for dementia. For this chapter, we shift our focus to exploring the effects of dementia, alongside the general effects of aging, on cognitive and linguistic processes.

Research in cognitive aging shows that older adults in the range of 60 to 80 years old, relative to younger adults around the age of 18 to 30 years old, exhibit a decrease in mental processing speed, have lower working memory spans, and diminished cognitive reasoning [11]. This similarly applies to language processing wherein development progresses rather quickly during infancy and childhood, is relatively stable in adolescence and adulthood, and undergoes a gradual decline in older age. Understanding these age-related changes in cognitive and linguistic functions is crucial for the development of NLP technologies that cater to the needs of aging population.

Coherence is an important aspect for effective communication. A coherent discourse enables the listener to maintain mental representation and understand how these representations are interconnected within the broader context of the discourse. Coherence can be measured locally between two utterances, like as discussed in Section 4.2 about natural language inference and Dialogue-NLI, or globally between an utterance and the overarching discourse theme using 4-point Likert scales or error analysis. For older adults, there is strong evidence that local and global coherence declines with age. To measure this, older adults were assigned to complete a variety of discourse tasks such as describing pictures [33] and recounting family and work experiences [19]. Glosser and Deser [19] reached the conclusion that the differences observed between local and global coherence are partially attributed to dissociation in the the cognitive systems that underlie these forms of coherence. Specifically, since global coherence in the context of this task relied on long-term memory of personal information, it was expected that the global coherence rate would be low for older adults. On the other hand, low local coherence rate is often attributed to older adults' difficulty in finding or recalling certain words.

We see these observations emphasized in the context of dementia. In the early stages of Alzheimer's disease, affected individuals primarily exhibit with naming objects and finding appropriate words, but their communication is as expected for healthy aging. It is in the severe to moderate stages of Alzheimer's disease that communication becomes significantly fluent and memory is affected. A study by Dijkstra et al. [13] shows that general knowledge and vocabulary is more compromised for persons with dementia causing memory retrieval failures that result to aborted phrases, empty words/phrases, repetitions, and disruptive topic shifts during conversations.

6.3 Interdisciplinary Insights

Research plays an important role in reinforcing or challenging particular notions of identity. Computational research done with little to no scrutiny, particularly in the context of older adults and dementia, risks perpetuating societal notions of identity that may overlook the unique experiences and needs of these populations. Therefore, it is imperative to approach such research problems with careful consideration and ensure that it accurately reflects the diverse realities of persons with dementia. To guide our study in developing the methodology, we asked two of our collaborators for this project to annotate the OLDER ADULTS INTERVIEW DATASET with identity cues. Both annotators, affiliated with the English Language and Literature Department from the Faculty of Arts at the University of Waterloo, are knowledgeable about identity and dementia studies from their involvement and collaboration with us in this project. As the concept of identity is rather dynamic, we opted not to provide explicit directions on how to annotate the interview transcripts. Instead, by giving them the autonomy, we aimed for diversity in interpretation that would provide us with a comprehensive understanding of the concept of identity. This process, commonly known as *coding*, is a fundamental step in thematic analysis. It involves identifying categories and concepts within raw text e.g., interview recordings and transcripts, and assigning descriptive text representations, known as *codes*, to each sentence or clause. Additionally, we requested our annotators to provide a biographical summary, approximately 3-5 sentences, capturing the interviewee's identity.

The annotation process revealed both interesting and significant insights about how identity can be inferred from text and how this process can be conceptualized computationally. By closely analyzing expressions and contextual references within the interview transcripts, our annotators identified subtle nuances that hinted at individuals' identities. For example, the annotators deemed the highlighted clause in Figure 6.1 a significant aspect of Mrs. Rechenmacher's identity due to the emotional attachment conveyed when mentioning the sweater. The mention of the sweater, although seemingly unrelated to the question, naturally prompted Mrs. Rechenmacher to remember it, suggesting its inherent significance in her narrative. The sweater was also only mentioned once during the interview. Translating this process of inferring abstract and emotionally charged connections into an automated NLP task is not a straightforward task, especially with limited data.

Considering the excerpt in Figure 6.2 from the interview with Mr. Mao, we can observe not only more of the speech characteristics inherent in older adults, but also disfluencies caused by communicating in their non-native language. This introduces additional complexities for the model as there are utterances that do not constitute a complete thought or sentence.

Q : What is something you really enjoy doing? Something that brings you much joy?

Mrs. R : Well, I like to make things. Right now I'm knitting and I find that when it's cold or when you can't go outside, it's nice to have something, has some handwork, something, whatever you like to do at embroidery or knit or read or write or anything to fill the cracks in the day. (...) When I was nine years old, my mother had a little girl, my sister, and when she was three years old and I was 12 years old, I knit her a sweater, a little red sweater. I will never forget it, and I am so proud of that in my own heart.

Figure 6.1: Annotated transcript excerpt from Mrs. Rechenmacher.

Q : What is something you really enjoy doing? Something that brings you much joy?

Mr. M : Oh yeah. I found as a research or through the research discovery new, uh, things. That could be on the science or engineering, or this is make me happy. Another one is, again, most important helping students. I see the development, even though, even for me right now, is for looking at my kids. Development, the progressing of the kids, a progressing of the students I found is enjoyable and uh, is very important for me.

Figure 6.2: Transcript excerpt from Mr. Mao.

Comparing annotations from each annotator, we also discovered that code frequency is a determining factor when describing the identity of the interviewee, as higher frequencies of certain codes corresponded to the identity cues emphasized by the annotators. We show the top 5 code categories from each annotator in Table 6.4. It is important to note that these annotations may be of different granularities and codes/code categories were decided on by the annotators.

	Annotator 1	**Annotator 2**
1	*Food* fruit, tapioca pudding, meat, vegetables, polenta	*Social identity* social connections, politics, religion, gender roles
2	*Family* children, family, sons, daughter, mom	*Time period* childhood, WW2, Post-war America, marriage
3	*Feeling* wonderful, proud, freedom, happy, horrible	*People* Children, spouse, friends, siblings, parents
4	*Place* Santa Clara, ocean, beach, San Jose, outside	*Feeling* happiness, contentment, pride, gratitude, fun
5	*Activity* knitting, handwork, embroidery, camp, garden	*Activity* bicycling, sewing, canning, traveling, knitting

Table 6.4: Top 5 code categories for Mrs. Rechenmacher according to each annotator. *Italicized* are the categories each code is grouped to.

Chapter 7

Conclusion

In this study, we investigated extracting persona attributes from dialogues as a proxy method for inferring identity cues from text. We devised a methodology specifically adapted for future implementation in dementia care. Concretely, we developed a two-stage attribute extractor that consists of a relation predictor and an entity extractor which was then trained on a proxy dialogue dataset created by combining PERSONA-CHAT and DIALOGUE-NLI. In an effort to capture nuanced dynamics of human dialogue, we consider relation prediction as a multi-label classification task and triplet extraction as a template infilling task. Given limited data for model training, our approach deviates from traditional information extraction approaches that are often unconstrained (i.e., relation types are more general in scope) and may yield to results that may not be relevant to persona attribute extraction.

Considering the intended application of this study, we evaluate our methodology on data that more closely reflects the linguistic complexities observed in persons with dementia using the OLDER ADULTS INTERVIEW DATASET. It is important to note that directly evaluating our methods on individuals with dementia is infeasible due to ethical adoption concerns. We present a comprehensive discussion on the effects of dementia, and in general aging, on cognitive and linguistic processes that makes this for a challenging research problem. Our results demonstrate that our method is capable handling noise in the data, such as aborted phrases, filler words, repetitions, to some degree but exhibits limitations when dealing with exceptionally longer input sequences.

The author wishes to emphasize that while progress has been made, there still remains significant work to be done in line with the ethical adoption principles. This thesis merely establishes the foundation for exploring this problem to ensure that the inclusion of persons with dementia is conducted with utmost consideration for their rights and well-being. In the context of limited data, collecting more interviews with questions specifically targeted on identity for the Older Adults Interview Dataset would enable us to develop and train models more effectively. Moreover, incorporating human evaluation to assess the outputs of these models is crucial in establishing trust in large language models, particularly when considering adherence to ethical adoption principles. In due course, inclusive participatory design involving persons with dementia and caregivers should also be considered as the next step. This step would allow us to refine the relation types considered in this study, as well as evaluate the suitability of the language models for persons with dementia.

From a broader standpoint, the advancements presented in this thesis point to promising direction for incorporating interdisciplinary perspectives in practical applications of NLP research. By drawing upon insights from diverse fields such as linguistics, psychology, and computer science, these advancements offer a more comprehensive approach to addressing complex challenges in natural language processing. Building upon this line of work, numerous alternative solutions warrant further exploration. Future research can investigate the capabilities of LLMs in processing data derived from older adults, considering the notable distinctions between the linguistic patterns and communication styles prevalent in this demographic and the data typically used to train LLMs. Additionally, the lack of comprehensive research on persona attribute extraction within information extraction, despite its crucial role in various applications such as conversational agents and recommendation systems, calls for further research and effort.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[2] Alzheimer's Association et al. 2018 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 14(3):367–429, 2018.

[3] Arlene J Astell, Nicole Bouranis, Jesse Hoey, Allison Lindauer, Alex Mihailidis, Chris Nugent, Julie M Robillard, et al. Technology and dementia: The future is now. *Dementia and geriatric cognitive disorders*, 47(3):131–139, 2019.

[4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[5] Elisa Bassignana and Barbara Plank. What do you mean by relation extraction? a survey on datasets and study on scientific relation classification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 67–83, 2022.

[6] James Beauregard. Dementia and the regulation of gerontechnology. *Intelligent assistive technologies for dementia: Clinical, ethical, social, and regulatory implications*, page 265, 2019.

[7] Mary Bucholtz and Kira Hall. Locating identity in language. *Language and identities*, 18(1):18–28, 2010.

[8] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes, and Adam Jatowt. Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, 509:257–289, 2020.

[9] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*, 2022.

[10] Jiska Cohen-Mansfield, Aleksandra Parpura-Gill, and Hava Golander. Utilization of self-identity roles for designing interventions for persons with dementia. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 61(4):P202–P212, 2006.

[11] David E Copeland, Nicole J Bies-Hernandez, and Kris Gunawan. The role of cognition on age-related changes in language, memory, and mental models. *Language and Aging*, page 135, 2016.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[13] Katinka Dijkstra, Michelle S Bourgeois, Rebecca S Allen, and Louis D Burgio. Conversational coherence: Discourse analysis of older adults with and without dementia. *Journal of Neurolinguistics*, 17(4):263–283, 2004.

[14] Paul Drew and Traci Curl. Conversation analysis: Overview and new directions. *Advances in discourse studies*, pages 32–45, 2008.

[15] Teuntje R Elfrink, Christina Ullrich, Miriam Kunz, Sytse U Zuidema, and Gerben J Westerhof. The online life story book: A randomized controlled trial on the effects of a digital reminiscence intervention for people with (very) mild dementia and their informal caregivers. *PLoS One*, 16(9):e0256251, 2021.

[16] Joe Enright, Megan E O'Connell, Camille Branger, Andrew Kirk, and Debra Morgan. Identity, relationship quality, and subjective burden in caregivers of persons with dementia. *Dementia*, 19(6):1855–1871, 2020.

[17] Luciano Gamberini, Francesco Martino, Bruno Seraglia, Anna Spagnolli, Malena Fabregat, Francisco Ibanez, Mariano Alcaniz, and Javier Montesa Andrés. Eldergames project: An innovative mixed reality table-top solution to preserve cognitive functions in elderly people. In *2009 2nd conference on human system interactions*, pages 164–169. IEEE, 2009.

[18] Philip Gleason. Identifying identity: A semantic history. *The journal of American history*, 69(4):910–931, 1983.

[19] Guila Glosser and Toni Deser. A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journal of Gerontology*, 47(4):P266–P272, 1992.

[20] Manuel Gonçalves-Pereira, Steven H Zarit, Ana L Papoila, and Raimundo Mateos. Positive and negative experiences of caregiving in dementia: The role of sense of coherence. *International journal of geriatric psychiatry*, 36(2):360–367, 2021.

[21] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[22] Sergio Grueso and Raquel Viejo-Sobera. Machine learning methods for predicting progression from mild cognitive impairment to alzheimer's disease dementia: a systematic review. *Alzheimer's research & therapy*, 13:1–29, 2021.

[23] Jia-Chen Gu, Zhen-Hua Ling, Yu Wu, Quan Liu, Zhigang Chen, and Xiaodan Zhu. Detecting speaker personas from conversational texts. *arXiv preprint arXiv:2109.01330*, 2021.

[24] Donghoon Ham, Jeong-Gwan Lee, Youngsoo Jang, and Kee-Eung Kim. End-to-end neural pipeline for goal-oriented dialogue systems using gpt-2. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 583–592, 2020.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[26] Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. A simple language model for task-oriented dialogue. *Advances in Neural Information Processing Systems*, 33:20179–20191, 2020.

[27] Matti Hyvärinen and Ryoko Watanabe. Dementia, positioning and the narrative self. *Style*, 51(3):337–356, 2017.

[28] Jolanda Jetten, Catherine Haslam, Cara Pugliese, James Tonks, and S Alexander Haslam. Declining autobiographical memory and the loss of identity: Effects on well-being. *Journal of clinical and experimental neuropsychology*, 32(4):408–416, 2010.

[29] Bosung Kim, Hayate Iso, Nikita Bhutani, Estevam Hruschka, Ndapa Nakashole, and Tom Mitchell. Zero-shot triplet extraction by template infilling. *arXiv preprint arXiv:2212.10708*, 2022.

[30] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[31] Renee Leung. Persona-based target-guided dialogue systems: A use case in dementia care. Master's thesis, University of Waterloo, 2024.

[32] Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*, 2016.

[33] Andrea Marini, Anke Boewe, Carlo Caltagirone, and Sergio Carlomagno. Age-related differences in the production of textual descriptions. *Journal of psycholinguistic research*, 34:439–463, 2005.

[34] Sulantha Mathotaarachchi, Tharick A Pascoal, Monica Shin, Andrea L Benedet, Min Su Kang, Thomas Beaudry, Vladimir S Fonov, Serge Gauthier, Pedro Rosa-Neto, Alzheimer's Disease Neuroimaging Initiative, et al. Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiology of aging*, 59: 80–90, 2017.

[35] Alex Mihailidis, Jennifer N Boger, Tammy Craig, and Jesse Hoey. The coach prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC geriatrics*, 8:1–18, 2008.

[36] Ursula Naue and Thilo Kroll. 'the demented other': identity and difference in dementia. *Nursing Philosophy*, 10(1):26–33, 2009.

[37] Astrid Norberg. Sense of self among persons with advanced dementia. 2019.

[38] Olimpia Pino, Giuseppe Palestra, Rosalinda Trevino, and Berardina De Carolis. The humanoid robot nao as trainer in a memory program for elderly people with mild cognitive impairment. *International Journal of Social Robotics*, 12:21–33, 2020.

[39] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140): 1–67, 2020.

[40] Kumar B Rajan, Jennifer Weuve, Lisa L Barnes, Robert S Wilson, and Denis A Evans. Prevalence and incidence of clinically diagnosed alzheimer's disease dementia from 1994 to 2012 in a population study. *Alzheimer's & Dementia*, 15(1):1–7, 2019.

[41] Rui Ribeiro, Joao P Carvalho, and Luísa Coheur. Pgtask: Introducing the task of profile generation from dialogues. *arXiv preprint arXiv:2304.06634*, 2023.

[42] Julie M Robillard, Ian Cleland, Jesse Hoey, and Chris Nugent. Ethical adoption: A new imperative in the development of technology for dementia. *Alzheimer's & Dementia*, 14(9):1104–1113, 2018.

[43] Daniel Sollberger. On identity: from a philosophical point of view. *Child and adolescent psychiatry and mental health*, 7(1):1–10, 2013.

[44] Anna Tigunova, Andrew Yates, Paramita Mirza, and Gerhard Weikum. Listening between the lines: Learning personal attributes from conversations. In *The World Wide Web Conference*, pages 1818–1828, 2019.

[45] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[47] Aline Vézina, Line Robichaud, Philippe Voyer, and Daniel Pelletier. Identity cues and dementia in nursing home intervention. *Work*, 40(1):5–14, 2011.

[48] Sean Welleck, Jason Weston, Arthur Szlam, and Kyunghyun Cho. Dialogue natural language inference. *arXiv preprint arXiv:1811.00671*, 2018.

[49] Heather Harris Wright, Anthony D Koutsoftas, Gilson J Capilouto, and Gerasimos Fergadiotis. Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Aging, Neuropsychology, and Cognition*, 21(2):174–196, 2014.

[50] Chien-Sheng Wu, Andrea Madotto, Zhaojiang Lin, Peng Xu, and Pascale Fung. Getting to know you: User attribute extraction from dialogues. *arXiv preprint arXiv:1908.04621*, 2019.

[51] Judith Wuest, Penny King Ericson, and Phyllis Noerager Stern. Becoming strangers: the changing family caregiving relationship in alzheimer's disease. *Journal of advanced nursing*, 20(3):437–443, 1994.

[52] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*, 2018.

[53] Xueliang Zhao, Wei Wu, Can Xu, Chongyang Tao, Dongyan Zhao, and Rui Yan. Knowledge-grounded dialogue generation with pre-trained language models. *arXiv preprint arXiv:2010.08824*, 2020.

[54] Luyao Zhu, Wei Li, Rui Mao, Vlad Pandelea, and Erik Cambria. Paed: Zero-shot persona attribute extraction in dialogues. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9771–9787, 2023.

APPENDICES

Appendix A

DIALOGUE-NLI relation types

1. place_origin
2. live_in_citystatcountry
3. live_in_general
4. nationality
5. employed_by_company
6. employed_by_general
7. has_profession
8. previous_profession
9. job_status
10. teach
11. school_status
12. has_degree
13. attend_school
14. like_general
15. like_food
16. like_drink
17. like_animal
18. like_movie
19. like_music
20. like_read
21. like_sports
22. like_watching
23. like_activity
24. like_goto
25. dislike
26. has_hobby
27. has_ability
28. member_of
29. want_do
30. want_job
31. want
32. favorite_food
33. favorite_color
34. favorite_book
35. favorite_movie
36. favorite_music
37. favorite_music_artist
38. favorite_activity
39. favorite_drink
40. favorite_show
41. favorite_place
42. favorite_hobby
43. favorite_season
44. favorite_animal
45. favorite_sport
46. favorite
47. own
48. have
49. have_pet
50. have_sibling
51. have_children
52. have_family
53. have_vehicle
54. physical_attribute
55. misc_attribute
56. has_age
57. marital_status
58. gender
59. other