Automatic Reminiscence Therapy for Dementia

Mariona Caros
Universitat Politecnica de Catalunya

Maite Garolera
Consorci Sanitari de Terrassa
mgarolera@cst.cat

Petia Radeva
Universitat de Barcelona
petia.ivanova@ub.edu

Xavier Giro-i-Nieto
Universitat Politecnica de Catalunya
xavier.giro@upc.edu

Abstract

With people living longer than ever, the number of cases with dementia such as Alzheimer's disease increases steadily. It affects more than 46 million people worldwide, and it is estimated that in 2050 more than 100 million will be affected. While there are not effective treatments for these terminal diseases, therapies such as reminiscence, that stimulate memories from the past are recommended. Currently, reminiscence therapy takes place in care homes and is guided by a therapist or a carer. In this work, we present an AI-based solution to automatize the reminiscence therapy, which consists in a dialogue system that uses photos as input to generate questions. We run a usability case study with patients diagnosed of mild cognitive impairment that shows they found the system very entertaining and challenging. Overall, this paper presents how reminiscence therapy can be automatized by using machine learning, and deployed to smartphones and laptops, making the therapy more accessible to every person affected by dementia.

1 Introduction

Increases in life expectancy in the last century have resulted in a large number of people living to old ages and will result in a double number of dementia cases by the middle of the century [8][16]. The most common form of dementia is Alzheimer disease which contributes to 60–70% of cases [11]. Research focused on identifying treatments to slow down the evolution of Alzheimer's disease is a very active pursuit, but it has been only successful in terms of developing therapies that eases the symptoms without addressing the cause [1][20]. Besides, people with dementia might have some barriers to access to the therapies, such as cost, availability and displacement to the care home or hospital, where the therapy takes place. We believe that Artificial Intelligence (AI) can contribute in innovative systems to give accessibility and offer new solutions to the patients needs, as well as help relatives and caregivers to understand the illness of their family member or patient and monitor the progress of the dementia.

Therapies such as reminiscence, that stimulate memories of the patient's past, have well documented benefits on social, mental and emotional well-being [22][9], making them a very desirable practice, especially for older adults. Reminiscence therapy in particular involves the discussion of events and past experiences using tangible prompts such as pictures or music to evoke memories and stimulate conversation [28]. With this aim, we explore multi-modal deep learning architectures to be used to develop an intuitive, easy to use, and robust dialogue system to automatize the reminiscence therapy for people affected by mild cognitive impairment or at early stages of Alzheimer's disease.

We propose a conversational agent that simulates a reminiscence therapist by asking questions about the patient's experiences. Questions are generated from pictures provided by the patient, which

contain significant moments or important people in user's life. Moreover, to engage the user in the conversation we propose a second model which generates comments on user's answers. A chatbot model trained with a dataset containing simple conversations between different people. The activity pretends to be challenging for the patient, as the questions may require the user to exercise the memory. Our contributions include:

- Automation of the Reminiscence therapy by using a multi-modal approach that generates questions from pictures, without using a reminiscence therapy dataset.

- An end-to-end deep learning approach which do not require hand-crafted rules and it is ready to be used by mild cognitive impairment patients. The system is designed to be intuitive and easy to use for the users and could be reached by any smartphone with internet connection.

2   Related Work

The origin of chatbots goes back to 1966 with the creation of ELIZA [27] by Joseph Weizenbaum at MIT. Its implementation consisted in pattern matching and substitution methodology. Recently, data driven approaches have drawn significant attention. Existing work along this line includes retrieval-based methods [10][29] and generation-based methods[18][19]. In this work we focus on generative models, where sequence-to-sequence algorithm that uses RNNs to encode and decode inputs into responses is a current best practice.

Our conversational agent uses two architectures to simulate a specialized reminiscence therapist. The block in charge of generating questions is based on the work *Show, Attend and Tell* [30]. This work generates descriptions from pictures, also known as image captioning. In our case, we focus on generating questions from pictures. Our second architecture is inspired by *Neural Conversational Model* from [26] where the author presents an end-to-end approach to generate simple conversations. Building an open-domain conversational agent is a challenging problem. As addressed in [32] and [6], the lack of a consistent personality and lack of long-term memory which produces some meaningless responses in these models are still unresolved problems.

Some works have proposed conversational agents for older adults with a variety of uses, such as stimulate conversation [31] , palliative care [25] or daily assistance. An example of them is 'Billie' reported in [13] which is a virtual agent that uses facial expression for a more natural behavior and is focused on managing user's calendar, or 'Mary' [24] that assists the users by organizing their tasks offering reminders and guidance with household activities. Both of the works perform well on its specific tasks, but report difficulties to maintain a casual conversation. Other works focus on the content used in Reminiscence therapy. Like [3] where the authors propose a system that recommends multimedia content to be used in therapy, or Visual Dialog [5] where the conversational agent is the one that has to answer the questions about the image.

3   Methodology

In this section we explain the main two components of our model, as well as how the interaction with the model works. We named it *Elisabot* and its goal is to mantain a dialog with the patient about her user's life experiences.

Before starting the conversation, the user must introduce photos that should contain significant moments for him/her. The system randomly chooses one of these pictures and analyses the content. Then, Elisabot shows the selected picture and starts the conversation by asking a question about the picture. The user should give an answer, even though he does not know it, and Elisabot makes a relevant comment on it. The cycle starts again by asking another relevant question about the image and the flow is repeated for 4 to 6 times until the picture is changed. The Figure 1 summarizes the workflow of our system.

Elisabot is composed of two models: the model in charge of asking questions about the image which we will refer to it as VQG model, and the Chatbot model which tries to make the dialogue more engaging by giving feedback to the user's answers.

2

3.1 VQG model

The algorithm behind VQG consists in an Encoder-Decoder architecture with attention. The Encoder takes as input one of the given photos I from the user and learns its information using a CNN. CNNs have been widely studied for computer vision tasks. The CNN provides the image's learned features to the Decoder which generates the question y word by word by using an attention mechanism with a Long Short-Term Memory (LSTM). The model is trained to maximize the likelihood $p(y|I)$ of producing a target sequence of words:

$$\mathbf{y} = \{y_1, ..., y_c\}, y_i \in \mathbb{R}^K \tag{1}$$

where K is the size of the vocabulary and C is the length of the caption.

Since there are already Convolutional Neural Networks (CNNs) trained on large datasets to represent images with an outstanding performance, we make use of transfer learning to integrate a pre-trained model into our algorithm. In particular, we use a *ResNet-101* [7] model trained on ImageNet. We discard the last 2 layers, since these layers classify the image into categories and we only need to extract its features.

3.2 Chatbot network

The core of our chatbot model is a sequence-to-sequence [23]. This architecture uses a Recurrent Neural Network (RNN) to encode a variable-length sequence to obtain a large fixed dimensional vector representation and another RNN to decode the vector into a variable-length sequence.

The encoder iterates through the input sentence one word at each time step producing an output vector and a hidden state vector. The hidden state vector is passed to the next time step, while the output vector is stored. We use a bidirectional Gated Recurrent Unit (GRU), meaning we use two GRUs one fed in sequential order and another one fed in reverse order. The outputs of both networks are summed at each time step, so we encode past and future context.

The final hidden state h_t^{enc} is fed into the decoder as the initial state h_0^{dec}. By using an attention mechanism, the decoder uses the encoder's context vectors, and internal hidden states to generate the next word in the sequence. It continues generating words until it outputs an <end> token, representing the end of the sentence. We use an attention layer to multiply attention weights to encoder's outputs to focus on the relevant information when decoding the sequence. This approach have shown better performance on sequence-to-sequence models [2].

4 Datasets

One of the first requirements to develop an architecture using a machine learning approach is a training dataset. The lack of open-source datasets containing dialogues from reminiscence therapy lead as to use a dataset with content similar to the one used in the therapy. In particular, we use two types of datasets to train our models: A dataset that maps pictures with questions, and an open-domain conversation dataset. The details of the two datasets are as follows.
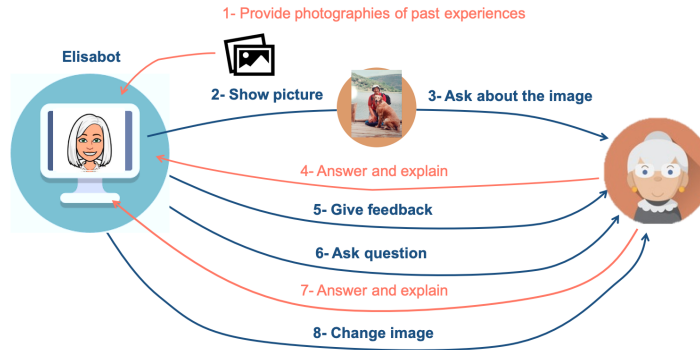

Figure 1: Scheme of the interaction with Elisabot

4.1 MS-COCO, Bing and Flickr datasets

We use MS COCO, Bing and Flickr datasets from [15] to train the model that generates questions. These datasets contain natural questions about images with the purpose of knowing more about the picture. As can be seen in the Figure 2, questions cannot be answered by only looking at the image. Each source contains 5,000 images with 5 questions per image, adding a total of 15,000 images with 75,000 questions. COCO dataset includes images of complex everyday scenes containing common objects in their natural context, but it is limited in terms of the concepts it covers. Bing dataset contains more event related questions and has a wider range of questions longitudes (between 3 and 20 words), while Flickr questions are shorter (less than 6 words) and the images appear to be more casual.

(a)	(b)	(c)
Q: Do these people go boating a lot?	*Q: What kind of chair is the cat in?*	*Q: What year did they graduate?*
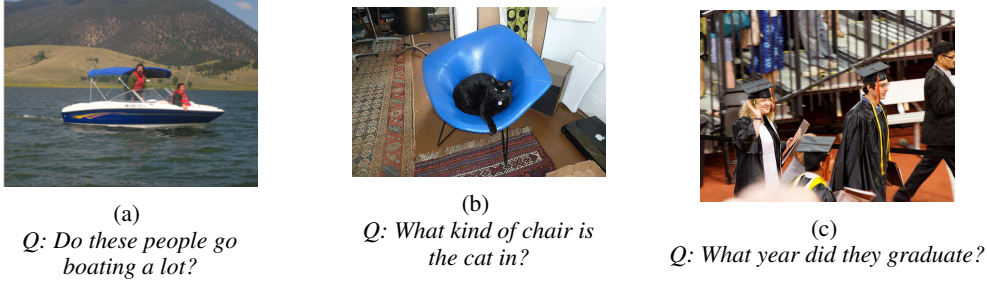
Figure 2: Samples from Bing 2a), Coco 2b) and Flickr 2c) datasets

4.2 Persona-chat and Cornell-movie corpus

We use two datasets to train our chatbot model. The first one is the Persona-chat [32] which contains dialogues between two people with different profiles that are trying to know each other. It is complemented by the Cornell-movie dialogues dataset [4], which contains a collection of fictional conversations extracted from raw movie scripts. Persona-chat's sentences have a maximum of 15 words, making it easier to learn for machines and a total of 162,064 utterances over 10,907 dialogues. While Cornell-movie dataset contains 304,713 utterances over 220,579 conversational exchanges between 10,292 pairs of movie characters.

5 Validation

An important aspect of dialogue response generation systems is how to evaluate the quality of the generated response. This section presents the training procedure and the quantitative evaluation of the model, together with some qualitative results.

5.1 Implementation

Both models are trained using Stochastic Gradient Descent with ADAM optimization [12] and a learning rate of 1e-4. Besides, we use dropout regularization [21] which prevents from over-fitting by dropping some units of the network.

The VQG encoder is composed of 2048 neuron cells, while the VQG decoder has an attention layer of 512 followed by an embedding layer of 512 and a LSTM with the same size. We use a dropout of 50% and a beam search of 7 for decoding, which let as obtain up to 5 output questions. The vocabulary we use consists of all words seen 3 or more times in the training set, which amounts to 11.214 unique tokens. Unknown words are mapped to an <unk> token during training, but we do not allow the decoder to produce this token at test time. We also set a maximum sequence length of 6 words as we want simple questions easy to understand and easy to learn by the model.

In the Chatbot model we use a hidden size of 500 and Dropout regularization of 25%. For decoding we use greedy search, which consists in making the optimal token choice at each step. We first train it with Persona-chat and then fine-tune it with Cornell dataset. The vocabulary we use consists of all words seen 3 or more times in Persona-chat dataset and we set a maximum sequence length of 12 words. For the hyperparameter setting, we use a batch size of 64.

5.2 Quantitative evaluation

We use the BLEU [17] metric on the validation set for the VQG model training. BLEU is a measure of similitude between generated and target sequences of words, widely used in natural language processing. It assumes that valid generated responses have significant word overlap with the ground truth responses. We use it because in this case we have five different references for each of the generated questions. We obtain a BLEU score of 2.07.

Our chatbot model instead, only have one reference ground truth in training when generating a sequence of words. We considered that it was not a good metric to apply as in some occasions responses have the same meaning, but do not share any words in common. Thus, we save several models with different hyperparameters and at different number of training iterations and compare them using human evaluation, to chose the model that performs better in a conversation.

5.3 Qualitative results

Our first goal was to generate meaningful questions from the provided pictures. Table 1 contains some generated questions by our VQG model for each of the proposed images. These images were not in the training set as they were taken from photo albums, so the model was the first time that had these images as input. By looking at the questions, we can easily realise that the model is not only able to generate questions grammatically correct, but to understand the content of the image and ask about it. Furthermore, we not only wanted to generate only a question per image, but to bring out up to five questions for a deeper conversation.

We used different kind of pictures as we wanted a robust model able to work with all kind of images. In the Table 1 we can see both, old gray and colorful photos, both with vertical and horizontal shapes. Most of the pictures we selected for testing were showing people, because these are the type of pictures used in reminiscence therapy, as they evoke memories from the relationships with that persons. We also tested pictures showing pets, cars and motorbikes as some people are very attached to them, triggering strong emotions.

Most of the generated questions could be asked by a person. Nonetheless, we found some of the questions such as *How did the car get damaged?* that do not correspond with the image's content, or are not well formulated like *What body of dog is that?*, but can be easily understood.

About the chatbot module, we test it manually by having a conversation with it. In the following lines, we can see two samples of a conversation, where each sentence starting with ">" is an input to our seq2seq model. In this first sample, the chatbot model is clearly overfitted as it repeats the same sentence twice. While in the second sample, we can see that the model is able to follow the conversation.

Overfitted chatbot model:

```
> hi
Elisabot:  hello !  how are you ?
> fine and you?
Elisabot :  i am good do you have any hobbies ?
> yes i like to ski
Elisabot:  i love to play sports .  it is my favorite
> I have a sister
Elisabot :  i have a dog .  i love them .
> I have a dog too
Elisabot:  i have a dog .  i love them
> ok
Elisabot :  i am not really a good person
> why?
Elisabot :  i love to play sports .  it is my favorite
```

Final chatbot model:

```
> hello
```

Table 1: Generated questions

Photo	Generated questions
	• How old is the dog? • What body of dog is that? • What beach is that? • Where is this? • Where is the dog going?
	• What kind of cake is that? • Who made the cake? • Is this a birthday cake? • How old is the birthday person?
	• What kind of car is that? • What year was this photo taken? • How old is that car? • How did the car get damaged?

```
Elisabot:  hello how are you ?
> fine and you?
Elisabot :  i m great !
> I have a sister, do you have any brother or sister?
Elisabot :  i have a iphone but i have no idea .
> ok
Elisabot :  i am a pro fan of thrones
> me too
Elisabot :  i like to eat healthy
> have you seen the last chapter?
Elisabot :  no i have not .
> where do you live?
Elisabot :  i live in the mountains
```

6   Usability study

As most of the metrics correlate very weakly with human judgements in the non-technical domain [14] we decide to evaluate our system with a simple user study with two patients. We present the user interface built and the feedback obtained from the patients.

6.1   User interface

We developed a user interface for Elisabot with Telegram, an instant messaging application available for smartphones or computers. We select it because it is easy to use and it offers an API for developers to connect bots to the Telegram system. It enables to create special accounts for bots which do not require a phone number to set up.

(a) /Start   (b) /Change
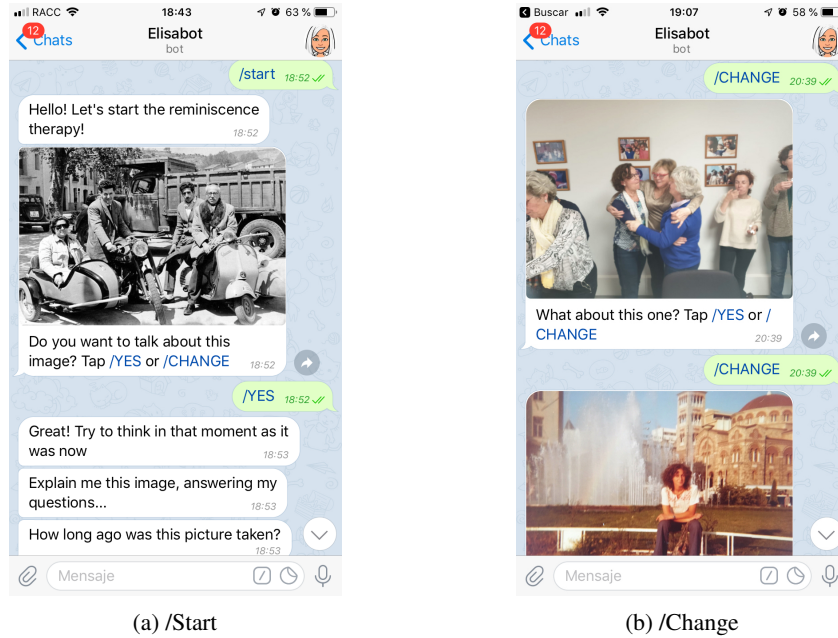
Figure 3: Elisabot running on Telegram application

Telegram is only the interface for the code running in the server. The bot is executed via an HTTP-request to the API. Users can start a conversation with Elisabot by typing `@TherapistElisabot` in the searcher and executing the command `/start`, as can be seen in the Figure 3. Messages, commands and requests sent by users are passed to the software running on the server. We add `/change`, `/yes` and `/exit` commands to enable more functionalities. `/Change` gives the opportunity to the user to change the image in case the user does not want to talk about it, `/yes` accepts the image which is going to talk about and `/exit` finishes the dialogue with Elisabot. The commands can be executed either by tapping on the linked text or typing them.

7   Feedback from patients

We designed a usability study where users with and without mild cognitive impairment interacted with the system with the help of a doctor and one of the authors. The purpose was to study the acceptability and feasibility of the system with patients of mild cognitive impairment. The users were all older than 60 years old. The sessions lasted 30 minutes and were carried out by using a laptop computer connected to Telegram. As Elisabot's language is English we translated the questions to the users and the answers to Elisabot.

Figure 5 is a sample of the session we did with mild cognitive impairment patients from *anonymized institution and location*. The picture provided by the patient (Figure 4 is blurred for user's privacy rights. In this experiment all the generated questions were right according to the image content, but the feedback was wrong for some of the answers. We can see that it was the last picture of the session as when Elisabot asks if the user wants to continue or leave, and he decides to continue, Elisabot finishes the session as there are no more pictures remaining to talk about.

At the end of the session, we administrated a survey to ask participants the following questions about their assessment of Elisabot:

- Did you like it?
- Did you find it engaging?
- How difficult have you found it?

Responses were given on a five-point scale ranging from *strongly disagree* (1) to *strongly agree* (5) and *very easy* (1) to *very difficult* (5). The results were 4.6 for amusing and engaging and 2.6 for

7

Figure 4: Photo provided by the user and selected by Elisabot

Figure 5: Sample of the session study with mild cognitive impairment patient

difficulty. Healthy users found it very easy to use (1/5) and even a bit silly, because of some of the generated questions and comments. Nevertheless, users with mild cognitive impairment found it engaging (5/5) and challenging (4/5), because of the effort they had to make to remember the answers for some of the generated questions. All the users had in common that they enjoyed doing the therapy with Elisabot.

8 Conclusions

We presented a dialogue system for handling sessions of 30 minutes of reminiscence therapy. Elisabot, our conversational agent leads the therapy by showing a picture and generating some questions. The goal of the system is to improve users mood and stimulate their memory and communication skills. Two models were proposed to generate the dialogue system for the reminiscence therapy. A visual question generator composed of a CNN and a LSTM with attention and a sequence-to-sequence model to generate feedback on the user's answers. We realize that fine-tuning our chatbot model with another dataset improved the generated dialogue.

The manual evaluation shows that our model can generate questions and feedback well formulated grammatically, but in some occasions not appropriate in content. As expected, it has tendency to produce non-specific answers and to loss its consistency in the comments with respect to what it has said before. However, the overall usability evaluation of the system by users with mild cognitive impairment shows that they found the session very entertaining and challenging. They had to make an effort to remember the answers for some of the questions, but they were very satisfied when they achieved it. Though, we see that for the proper performance of the therapy is essential a person to support the user to help remember the experiences that are being asked.

This project has many possible future lines. In our future work, we suggest to train the model including the Reddit dataset which could improve the chatbot model, as it has many open-domain conversations. Moreover, we would like to include speech recognition and generation, as well as real-time text translation, to make Elisabot more autonomous and open to older adults with reading

and writing difficulties. Furthermore, the lack of consistency in the dialogue might be avoided by improving the architecture including information about passed conversation into the model. We also think it would be a good idea to recognize feelings from the user's answers and give a feedback according to them.

Acknowledgements

References

[1] Association Alzheimer's. 2015 alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 11(3):332, 2015.

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

[3] Adam Bermingham, Julia O'Rourke, Cathal Gurrin, Ronan Collins, Kate Irving, and Alan F Smeaton. Automatically recommending multimedia content for use in group reminiscence therap. In *Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare*, pages 49–58. ACM, 2013.

[4] Cristian Danescu-Niculescu-Mizil and Lillian Lee. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*, 2011.

[5] Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José MF Moura, Devi Parikh, and Dhruv Batra. Visual dialog. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 326–335, 2017.

[6] Asma Ghandeharioun, Judy Hanwen Shen, Natasha Jaques, Craig Ferguson, Noah Jones, Agata Lapedriza, and Rosalind Picard. Approximating interactive human evaluation with self-play for open-domain dialog systems. *arXiv preprint arXiv:1906.09308*, 2019.

[7] 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.

[8] Richard A Hickman, Arline Faustin, and Thomas Wisniewski. Alzheimer disease and its growing epidemic: risk factors, biomarkers, and the urgent need for therapeutics. *Neurologic clinics*, 34(4):941–953, 2016.

[9] Alina Huldtgren, Anja Vormann, and Christian Geiger. Reminiscence map: Insights to design for people with dementia from a tangible prototype. 2015.

[10] Zongcheng Ji, Zhengdong Lu, and Hang Li. An information retrieval approach to short text conversation. *arXiv preprint arXiv:1408.6988*, 2014.

[11] Harashish Jindal, Bhumika Bhatt, Shashikantha Sk, and Jagbir Singh Malik. Alzheimer disease immunotherapeutics: then and now. *Human vaccines & immunotherapeutics*, 10(9):2741–2743, 2014.

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

[13] Stefan Kopp, Mara Brandt, Hendrik Buschmeier, Katharina Cyra, Farina Freigang, Nicole Krämer, Franz Kummert, Christiane Opfermann, Karola Pitsch, Lars Schillingmann, et al. Conversational assistants for elderly users–the importance of socially cooperative dialogue. In *AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications*, 2018.

[14] Chia-Wei Liu, Ryan Lowe, Iulian V Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023*, 2016.

[15] Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Larry Zitnick, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende. Generating natural questions about an image. *CoRR*, abs/1603.06059, 2016.

[16] Olawale Olanrewaju, Linda Clare, Linda Barnes, and Carol Brayne. A multimodal approach to dementia prevention: a report from the cambridge institute of public health. *Alzheimer's & Dementia: translational research & clinical interventions*, 1(3):151–156, 2015.

[17] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

[18] Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[19] Iulian Vlad Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron Courville, and Yoshua Bengio. A hierarchical latent variable encoder-decoder model for generating dialogues. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[20] Maite Solas, Elena Puerta, and Maria J Ramirez. Treatment options in alzheimer s disease: The gaba story. *Current pharmaceutical design*, 21(34):4960–4971, 2015.

[21] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[22] Ponnusamy Subramaniam and Bob Woods. The impact of individual reminiscence therapy for people with dementia: systematic review. *Expert Review of Neurotherapeutics*, 12(5):545–555, 2012.

[23] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[24] Christiana Tsiourti, Maher Ben Moussa, João Quintas, Ben Loke, Inge Jochem, Joana Albuquerque Lopes, and Dimitri Konstantas. A virtual assistive companion for older adults: design implications for a real-world application. In *Proceedings of SAI Intelligent Systems Conference*, pages 1014–1033. Springer, 2016.

[25] Dina Utami, Timothy Bickmore, Asimina Nikolopoulou, and Michael Paasche-Orlow. Talk about death: End of life planning with a virtual agent. In *International Conference on Intelligent Virtual Agents*, pages 441–450. Springer, 2017.

[26] Oriol Vinyals and Quoc Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[27] Joseph Weizenbaum et al. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[28] Bob Woods, Laura O'Philbin, Emma M Farrell, Aimee E Spector, and Martin Orrell. Reminiscence therapy for dementia. *Cochrane database of systematic reviews*, (3), 2018.

[29] Yu Wu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li. Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. *arXiv preprint arXiv:1612.01627*, 2016.

[30] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057, 2015.

[31] Kiyoshi Yasuda, Jun-ichi Aoe, and Masao Fuketa. Development of an agent system for conversing with individuals with dementia. pages 3C1IOS1b2–3C1IOS1b2, 2013.

[32] 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.