Exploring Episodic Future Thinking (EFT) for Behavior Change: NLP and Few-Shot In-Context Learning for Health Promotion

Sareh Ahmadi
Department of Computer Science
Virginia Tech
Blacksburg, VA, 24061, USA
email:saraahmadi@vt.edu

Edward A. Fox
Department of Computer Science
Virginia Tech
Blacksburg, VA, 24061, USA
email:fox@vt.edu, ORCiD 0000-0003-1447-6870

***Abstract*—Maladaptive health behaviors are closely linked to lifestyle-related diseases, such as obesity and type 2 diabetes. One significant factor contributing to maladaptive behavior is delay discounting, the tendency to prioritize immediate rewards over delayed ones. Episodic Future Thinking (EFT) is an intervention to reduce delay discounting and promote behavior change. EFT involves mentally simulating future events in a vivid manner, influencing decision-making and emotional well-being. Studies show EFT's effectiveness in reducing delay discounting and its potential for improving various health behaviors, including exercise and medication adherence. However, EFT's mechanisms of action and the conditions that impact its efficacy are unknown. This paper describes a study of EFT 'cue texts' to determine what makes them effective. It explains a new and efficient method to classify such texts with a few data, which can be used for further analysis to identify what characteristics of the texts lead to positive health outcomes. Classification framework is built using the FLAN-T5 large language model, with good results from zero-shot, and better results from few-shot in-context learning. This approach may be extended to address other behavioral health, wellness informatics, and technology-related approaches to global health challenges.**

Keywords*— *Episodic Future Thinking (EFT); delay discounting; maladaptive health behavior; Natural Language Processing (NLP); zero-shot learning; few-shot in-context learning.

I. INTRODUCTION

Maladaptive health behavior refers to actions or habits that are detrimental to a person's physical or mental well-being. Smoking and excessive alcohol consumption are examples of these behaviors. Maladaptive health behaviors are closely linked to lifestyle-related diseases, e.g., obesity, type 2 diabetes (T2D), cardiovascular diseases, certain types of cancer, and respiratory conditions. Promoting healthy lifestyle choices – regular exercise, a balanced diet, stress management, and avoidance of harmful substances – can significantly reduce the risk of developing these diseases. Interventions focusing on behavior change, support, and education can help individuals adopt healthier habits.

One contributing factor to maladaptive health behavior and lifestyle-related diseases is Delay Discounting (DD), which refers to the tendency to devalue delayed rewards in favor of immediate gratification [1]. Many unhealthy behaviors, and diseases related to lifestyle, are connected to DD [2]. Excessive discounting of delayed rewards is observed not only in substance-dependent individuals but also in individuals with behavioral disorders, such as pathological gambling, overeating, obesity, and Attention Deficit Hyperactivity Disorder (ADHD). Interventions could help with treating addiction and disorders linked to excessive discounting.

An increasing number of health behaviors and populations have been targeted by Episodic Future Thinking (EFT) [3] as an intervention for behavior change, aiming to decrease DD. EFT is a cognitive process that involves mentally simulating or envisioning future events in a detailed and vivid manner. It allows individuals to project themselves into the future and imagine specific situations, actions, and outcomes. EFT has garnered significant attention in recent years due to its potential role in influencing behavior, decision-making, and emotional well-being [3].

When investigating the impact of EFT on both DD and health behavior, participants commonly produce written or spoken depictions of personally significant future events. These descriptions are subsequently utilized as prompts/cues to facilitate EFT during decision-making tasks conducted in a laboratory setting or in real-world environments [4].

One study [5] focused on the association between DD and glycemic regulation, medication adherence, and eating and exercise behaviors in adults with prediabetes. It suggests that DD is a significant predictor of glycemic control and health behaviors in adults with prediabetes. Modifying DD can improve glycemic control and prevent the progression from prediabetes to T2D; interventions such as EFT may be beneficial. Another study [6] examined the effects of EFT on medication adherence in individuals with T2D, and potential mechanisms underlying these effects, such as improvements in prospective memory and DD. EFT had a positive impact on medication adherence among participants with T2D. Further research [7] focused on the long-term effects of EFT training on DD in individuals with prediabetes, as well as its impact on weight, HbA1c levels, and physical activity. Results indicate that EFT training can lead to sustained changes in DD and that

 ISBN: 978-1-68558-112-1

a combination of EFT and a low carbohydrate, low glycemic index diet can be effective for weight loss and glycemic control in individuals with prediabetes.

It has been shown in several studies that EFT is an effective intervention that can be scaled up to reduce DD and promote healthier behaviors. This study aims to enhance the effectiveness of EFT in preventing and treating T2D by gaining a better understanding of how EFT works and the factors that influence its efficacy. Previous and ongoing studies have shown significant variations in the content characteristics of EFT cues. The structure of the cues also varies, such as the extent to which they form a coherent narrative or describe achievement of health and personal goals, such as weight loss or financial planning. Additionally, cues differ in terms of imagery, vividness of events, emotional tone, and level of detail provided. We aim to build Natural Language Processing (NLP) classifiers to predict EFT content characteristics. In addition, our research aims to reduce the cost of annotating participant data for classification, and leverage more efficient and adaptable Large Language Models (LLMs). By utilizing techniques that rely on pre-trained LLMs and their ability to generalize to new tasks, we hope to pave the way for more accessible and scalable methods in NLP research. We are particularly interested in exploring the application of few-shot and zero-shot in-context learning techniques within the emerging field of instruction-tuned models.

In Section II, we discuss the application of NLP in the health domain, as well as the methodologies employed in constructing NLP classifiers. Section III delves into the zero-shot and few-shot classification of EFT data. The results are presented in Section IV, and we draw our conclusions in Section V.

II. Related Work

In recent years, NLP has been shown to help with global health challenges. In this domain, by leveraging corpora and learning approaches, NLP has demonstrated strong performance in various tasks, e.g., text mining [8], classification [9], sentiment mining [10], and information extraction [11]. In particular, NLP techniques may offer multiple perspectives in mental health research and in mental health clinical practice [12]. For instance, a study [13] explored the feasibility of automatically extracting schemas from thought records. A method for identifying the use of evidence-based psychotherapy for post-traumatic stress disorder was developed by applying NLP methods to clinical notes [14]. NLP also can address the knowledge gap in utilizing lifestyle modification data, including diet, exercise, and tobacco cessation, from Electronic Health Records (EHRs) for research purposes [15].

Furthermore, as the research in diabetes care is growing, and a considerable portion of real-world data exists in narrative form, NLP technology presents a viable solution for effectively analyzing narrative electronic data [16]. Given the success of NLP approaches, several studies have been dedicated to diabetes care and diseases [17] [18]. A high-performance NLP system [19] was developed for automatically detecting hypoglycemic events from EHR notes of diabetes patients. It can be utilized for EHR-based hypoglycemia surveillance and population studies to improve patient care and enhance research in diabetes management. A thorough thematic analysis was performed [20] to identify 12 themes of vulnerability related to the health and well-being of T2D patients by leveraging language models with high test accuracy. To understand the information needs of diabetics a classification schema for diabetes-related questions was developed by analyzing questions collected from a health website [21]. An investigation of the relationship between lifestyle counseling in primary care settings and clinical outcomes in patients with diabetes applied NLP to electronic notes [22] to retrieve and classify lifestyle modification assessments and advice.

Transformers such as BERT [23] offer a promising approach for building text classifiers, but one significant challenge lies in the amount and quality of data they require. Annotating data for training classifiers can be costly, time-consuming, and requires domain-specific knowledge. Manual annotation involves experts meticulously labeling large volumes of data, which is a resource-intensive and time-consuming task. Likewise, since many text datasets are imbalanced – with few instances of the minority category relative to those in the majority category – special care and techniques are required, as shown in our prior studies [24] [25].

Initially, fine-tuning was a dominant approach where these pre-trained language models were adapted to specific downstream tasks by further training on task-specific labeled data. While fine-tuning yields impressive results, it heavily relies on large amounts of task-specific labeled data, which could be limiting when labeled data is scarce or non-existent for certain tasks or domains. Prompt tuning [26] emerged as a response to address the limitations of fine-tuning. Instead of relying solely on task-specific labeled data, prompt tuning leverages the pre-trained language model's ability to generate text by providing input prompts. By constructing appropriate prompts, including task-specific information or instructions, language models can be fine-tuned on new tasks without the need for extensive labeled data. Zero-shot learning takes the concept of prompt tuning further by allowing language models to generalize their understanding to unseen tasks or categories. With zero-shot learning, a pre-trained language model is capable of performing tasks for which it has never been explicitly trained [27]. By leveraging auxiliary information or prompts, such as textual descriptions or instructions, zero-shot learning enables language models to classify or generate text for new categories or tasks. Few-shot in-context learning [28] [29] builds upon the zero-shot learning paradigm and focuses on adapting language models to new tasks or categories with only a small number of labeled examples (also referred to as demonstrations). In few-shot in-context learning, the language model leverages a few labeled examples to quickly learn task-specific patterns or characteristics. The training examples are concatenated and provided as a single input to the model, which suits the k-shot learning scenario. GPT3 [28] showed emergent few-shot learning by simply pre-pending examples of the task as the input to the model. During testing, the model is assessed on a new target task with k-training examples. This approach significantly reduces both computation costs and the data requirements for adaptation to new tasks, making it particularly useful when labeled data is scarce or expensive to acquire.

In the context of few-shot in-contex learning, the term "k-shot" refers to the number of labeled examples available for each task. For example, if a model is trained on a 5-shot for a classification task, it means each task is provided with only five labeled examples or demonstrations. The model then uses this limited data to make predictions when presented with new tasks during testing. The value of k can vary depending on the specific few-shot learning scenario and the available data. In this context, k=0 represents zero-shot learning, k=1 corresponds to one-shot learning, and k>1 indicates few-shot learning. It is expected that the model's performance improves with a larger k because it can learn from more examples. Additionally, the inclusion of a prompt provides additional context, which enhances the model's accuracy, particularly when k is small.

The capacity of LLMs to adjust to specific tasks using few-shot demonstrations (in-context learning) has been observed [30]. As the size of LLMs increases, emergent capabilities have become more apparent [31] [32]. LLMs have demonstrated the capacity to generalize to unfamiliar tasks through instruction-based learning. Instruction tuning is a novel approach in NLP that utilizes natural language instructions to enable zero-shot and few-shot performance of language models on previously unseen tasks. Instruction-tuned LLMs are fine-tuned with inputs and outputs that are instructions, using techniques such as Reinforcement Learning from Human Feedback (RLFH) [33], or instruction-tuned based on supervised fine-tuning which involves the process of refining a pre-trained language model using public benchmarks and datasets which have instruction template formats. To enhance the fine-tuning process, additional instructions are introduced, either manually created or automatically generated, to augment the training data [34] [35]. These approaches can improve the LLMs' ability to follow instructions and safely adapt to new tasks.

III. METHOD

Given the superiority of the instruction-tuned language models [30], as the first step, we need to choose a pre-trained instruction-tuned model. The FLAN-T5 11B model (11 billion parameters, FLAN-T5 XXL) [36] outperforms the PaLM 62B model (62 billion parameters) [36], a novel transformer language model trained using the Pathways ML system [32] which is a recently developed machine learning system that allows for highly effective training. Moreover, FLAN-T5 excelled on difficult tasks in the BIG-Bench dataset [36]. Given the exceptional performance of the FLAN-T5 model and the public release of it, this model was selected. It is the instruction-tuned version of the T5 encoder-decoder model [37] that has undergone fine-tuning across a variety of tasks to follow instructions. It is able to perform zero-shot NLP tasks, as well as few-shot in-context learning tasks.

In our dataset, the participants were asked to write short texts (cues) about the events for different time frames, ranging from one month to ten years. Participants generated detailed and vivid descriptions of these events. The data used in this study originated from 18 different research studies conducted by medical research teams at two universities (Virginia Tech and the University of Buffalo). These studies specifically investigated the impact of EFT on diabetes and other relevant health-related outcomes. In total, the dataset comprises approximately 11,000 cues, each a few sentences in length. An example of selected data from one participant is as follows:

- In about 1 month, I am playing golf with my friends. We are having a great time and enjoying the company and competition. We laugh and have a great time.
- In about 3 months, I am picking my daughter up from college. I am excited she is done with school and we go to lunch at our favorite sushi restaurant and enjoy each other's company.
- In about 6 months, I am fishing in the bay with friends. We are on a charter boat and excited to catch some nice fish. We bet on who will catch the biggest fish.

The goal is to build a classification framework to predict the topic of the cues, as well as a level/value for three categories related to imagery, featuring variation in: event vividness, episodicity, and emotional valence. Table I shows the definitions for the 14 categories.

A subset of the data is used for manual labeling in Amazon Mechanical Turk. Each text was labeled by three different annotators to ensure the quality of the labeling process. The annotators are given the definition, and an example, for each category (Table I). Overall there are 400 labeled texts. For binary categories, we randomly sample 10 labeled texts belonging to a category (positive examples) and 10 labeled texts not belonging to that category (negative examples), i.e., 20-shots, and the 380 remaining are used for testing the few-shot and zero-shot settings. For 10-shot, 5 of the positive and 5 of the negative sampled labeled texts are used. The test data is the same (i.e., for zero-shot, 10-shot, and 20-shot). For the three-class classification, we randomly sample 10 examples from each class (30 labeled data for few-shot learning) and the 370 remaining are used for testing. For the 15-shot case, we use half of the labeled examples.

For an instruction-tuned model, the instructions and context provided to a language model are encompassed within prompts. Therefore it requires prompt engineering such that the input to the model contains well-crafted prompts, ensuring meaningful guidance, rather than blindly inputting everything without context. Prompt engineering involves the process of designing and refining prompts to effectively utilize language models for various applications [38]. Typically, the components that constitute a prompt are as follows:

- Instruction: Instruct the model on the desired actions,

 ISBN: 978-1-68558-112-1

TABLE I
CATEGORY DEFINITION

Class	Definition
Vivid: not	The text contains no details about the event. It is difficult to imagine the event. No context has been given regarding the event.
Vivid: moderately	The text contains only a few details or mostly non-specific details. The reader is left to fill in gaps, making it somewhat hard to imagine the event. More details could have been provided describing the event. Some context has been given regarding the event
Vivid: highly	The text contains sufficient and specific details so that the event described is readily and easily imaginable. A considerable amount of context has been given regarding the event.
Episodic: not	The writer primarily describes general knowledge of events or occurrences. The event is described as if the writer is not present or personally experiencing the event.
Episodic: moderately	The writer describes both personal experiences, events, and actions in addition to general facts or ideas. The writer is somewhat in the moment but also adds in a few facts or ideas.
Episodic: highly	The writer primarily describes personal experiences, events, and actions, NOT general facts or ideas. The writer is describing events as if they are currently experiencing them "in the moment". The writer provides details about their own emotions and/or what they hear, see, or feel.
Emotion: negative	Primarily contains references to negative emotions or behaviors, including sadness, crying, or anger.
Emotion: neutral	Contains references to both positive and negative emotions or behaviors or contains weak or ambiguous references to positive or negative emotions and behaviors.
Emotion: positive	Contains references to both positive and negative emotions or behaviors or contains weak or ambiguous references to positive or negative emotions and behaviors.
Health	Contains an obvious, specific reference to physical or mental health. Examples include but are not limited to improved or worse physical state or mental health, and intentional changes in behaviors to improve health and health outcomes.
Recreation	Contains obvious or specific references to engaging in an activity for leisure or fun while not working at one's job. Examples include but are not limited to sports or physical activities like running or hiking, art, movies and television, or hobbies like gardening.
Better-me	Contains obvious or specific references to "a better me", including personal development, self-improvement, making positive changes in life, achievements, hard work, or determination. May contain references to the idea that things are looking up or getting better.
Celebration	Contains an obvious, specific reference to a celebration or a celebratory event.
Food	Contains obvious or specific references to food, eating, cooking, or a meal. Eating or food is a major and essential component of the text.
Alone	Contains an obvious, specific reference to events and activities which shows being done alone.
Family	Contains obvious or specific references to family (immediate or extended). Family is a major and essential component of the text.
Partner	Contains an obvious, specific reference or mention of a romantic partner.
Friends	Contains obvious or specific references to a friend or friends (non-family members). Friends are a major and essential component of the text.
Pet	Contains obvious or specific references to a pet, not any animal.

guide its utilization of external information (if available), and outline the construction of the output.

- Context: Serves as supplementary knowledge for the model, providing additional information. They can be manually included within the prompt, obtained through a vector database using retrieval augmentation, or acquired through alternative methods.
- Input Data: Refers to the input provided by a human user (i.e., the user input or query)
- Output indicator: Denotes the starting point of the to-be-generated text

Although not all prompts incorporate these elements, a well-crafted prompt frequently incorporates two or more of them. To adapt the model to our dataset, the instruction for the model is set as the category definition followed by some demonstrations (examples) for few-shot in-context learning. For classification, a demonstration involves an input x and its corresponding ground-truth label y. The model is provided with a sequence of such demonstrations, followed by a test input. The objective is for the language model to predict the label of this final data point. The demonstrations are sampled randomly for the model. Below is an example of an input instruction template for the 3-shot setting. Fig. 1 also depicts the framework for building a classifier. For the zero-shot setting, the input prompt contains only the instructions, query, and output indicator. The block below is an example.

”””Classify the given text into three categories: not episodic, moderately episodic, and highly episodic based on the definition for each category.

- Highly episodic definition: The writer primarily describes personal experiences, events, and actions, NOT general facts or ideas. The writer is describing events as if they are currently experiencing them at the moment. The writer provides details about their own emotions and/or what they hear, see, or feel.
- Moderately episodic definition: The writer describes both personal experiences, events, and actions in addition to general facts or ideas. The writer is somewhat in the moment but also adds in a few facts or ideas.
- Not episodic definition: The writer primarily describes general knowledge of events or occurrences. The event is described as if the writer is not present or personally experiencing the event.

In about 10 years, I am blowing out the candles on my birthday cake and feel pleased that my family has gathered because they love me so much. I am smiling because my adult children are making fun of how many candles are on my cake this year. I hear one of them jokingly say it's time to call the fire department and everyone laughs with love in their voices.
Output: Highly episodic

##

Input: In 6 months, I am visiting my mother-in-law. She is down visiting from Indiana. She is staying with us for a week. We have waited for her to come down for a long time because we have not seen her in years. We are ecstatic she is here and is visiting. The kids are excited that they finally get to see her. We are going shopping and do other fun things while she is here. We are looking forward to all the excitement that happens while she is visiting.
Output: Moderately episodic

##

In about 5 years, my car is paid off. Sweet! Who doesn't love any more car payments?
Output: Not episodic

##

Input: In about 4 years, I am on vacation at the beach relaxing in the sun under a big umbrella. I am with my husband and my daughter and we are finally taking a family vacation. My husband is playing with my daughter out in the ocean, jumping waves, and helping her on a boogie board. I am back in my chair on the sand, my sunglasses are on and I am enjoying the quiet and the warmth. The sand is warm between my toes and the weather is perfect. I am feeling at peace.
Output: ””””

For few-shot learning, given that the maximum sequence length for the model is 2048 tokens, we can provide up to 30-shots i.e., we randomly sample 10 examples per class for the three-class classification. To be consistent, for the binary classes, we also randomly sampled 10 positive and 10 negative examples for each category. Therefore, we can provide the model with 20-shots for each binary category

 ISBN: 978-1-68558-112-1

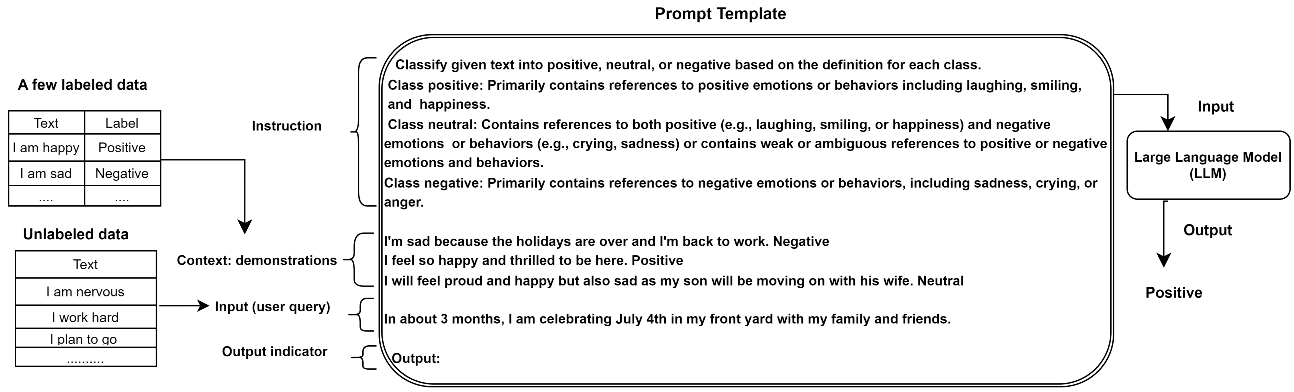

Fig. 1. Classification framework for few-shot in-context learning

(topical categories) and 30 shots for the 3-class categories (episodic, vivid, and emotion).

IV. RESULTS

We present the results for model performance for zero-shot and few-shot in-context learning. To evaluate the classification, we measure the accuracy and macro F1-score. Each experiment is run for 3 trials and the average result is reported in Tables II and III. Table II shows the classifier performance for the three class categories while Table III shows the performance for topical (binary) categories. Performance with few-shot examples is better than the zero-shot learning approach, for all the classes, allowing the model to better comprehend and distinguish for those classes. The improvement shown in Table II through few-shot learning is impressive, i.e., for the three classes of episodicity, vividness, and emotion. Increasing the number of demonstrations from 5 to 10 per class also helped the model to improve performance. As a result, the few-shot learning paradigm demonstrates superior performance, effectively capturing underlying patterns in the data, surpassing the limitations of zero-shot learning, which relies solely on generalizing to entirely unseen classes without any labeled training samples. Our model demonstrates strong performance in the binary category for zero-shot setting, showcasing its capability to handle classification tasks effectively. However, its true capability becomes evident when faced with few-shot examples. Even with limited training data, the model exhibits superior adaptability and displays enhanced performance. These findings highlight the importance of learning from examples to enhance the classification capabilities of the model and showcase its potential for real-world applications with limited labeled data. In this study, we employed a single model for classification and observed notable advantages with few-shot learning. Overall, prompting one model provides a more efficient approach that can lead to enhanced performance and easier management of machine learning tasks. It is particularly advantageous when dealing with resource constraints, like those arising from the expense of manual labeling, especially when datasets are imbalanced and have very few positive/minority examples.

TABLE II
PERFORMANCE OF FLAN-T5 FOR 3 CLASS CATEGORIES

category	zero-shot		15-shot		30-shot	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
episodic	60%	44%	89%	78.3%	**91.3%**	**83%**
vivid	74%	45%	81%	67.66%	**86%**	**84%**
emotion	80%	55%	81%	59.6%	**82%**	**72%**

TABLE III
PERFORMANCE OF FLAN-T5 FOR BINARY CATEGORIES

category	zero-shot		10-shot		20-shot	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
health	94%	93%	94%	93%	**94.3%**	**94.6%**
better-me	79%	77%	80.3	77	**81%**	**78%**
recreation	83%	80%	83%	81%	**85%**	**83%**
family	79%	79%	83.6%	83.6%	**85.3%**	**84%**
friend	85%	79%	89.3%	82.33%	**89.6%**	**83%**
future	96%	95%	99.6%	99.6%	**100%**	**100%**
food	55%	50%	95.6%	94.6%	**96%**	**96%**
pet	95%	84%	96.6%	87%	**98%**	**89.3%**
alone	91%	83%	91 %	83.3%	**92%**	**84%**
celebration	71%	69%	82%	79%	**82.3%**	**79.6%**
partner	84%	81%	94%	92%	**95%**	**94%**

V. CONCLUSION

This research examines content characteristics of EFT data generated by people who suffer from conditions such as diabetes. Little is known about how these content characteristics influence the effectiveness of EFT in promoting behavior change. We proposed to utilize a pre-trained instruction-tuned model and apply zero-shot and few-shot in-context learning for classification to predict the content and characteristics of the generated EFTs. This can then be used for in-depth analysis to pinpoint which text features contribute to positive health results. The proposed method serves as a powerful tool that addresses the barriers posed by traditional fine-tuning methods, which typically demand a large amount of labeled data. Unlike fine-tuning, few-shot in-context learning significantly reduces the data requirements, making it more accessible and applicable in scenarios where labeled data is scarce or costly to obtain. By utilizing a single pre-trained model for each classification task and adapting it to new tasks with only a few examples, this approach avoids the need for maintaining separate models for every specific classification task. This efficiency not only saves computational resources but also opens up opportunities for practical implementations across various domains and industries. In essence, few-shot in-context learning represents a new and effective NLP technique that bridges the gap between data scarcity and high-performance artificial intelligence, offering a promising pathway for further advances in health-related domains. Future work includes using other instruction-tuned LLMs that can handle longer input sequence lengths, as well as more generalization capabilities for different NLP tasks.

ACKNOWLEDGEMENT

We would like to express our gratitude for the support provided by NIH NIDDK 3R01DK129567-02S1.

REFERENCES

[1] G. J. Madden and W. K. Bickel, *Impulsivity: The behavioral and neurological science of discounting.* Amer. Psychological Assoc., 2010.
[2] W. K. Bickel, D. P. Jarmolowicz, E. T. Mueller, M. N. Koffarnus, and K. M. Gatchalian, "Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence," *Pharmacology & therapeutics*, vol. 134, no. 3, pp. 287–297, 2012.
[3] C. M. Atance and D. K. O'Neill, "Episodic future thinking," *Trends in cognitive sciences*, vol. 5, no. 12, pp. 533–539, 2001.
[4] J. M. Brown and J. S. Stein, "Putting prospection into practice: Methodological considerations in the use of episodic future thinking to reduce delay discounting and maladaptive health behaviors," *Frontiers in Public Health*, vol. 10, p. 1020171, 2022.
[5] L. H. Epstein *et al.*, "Delay discounting, glycemic regulation and health behaviors in adults with prediabetes," *Behavioral Medicine*, vol. 47, no. 3, pp. 194–204, 2021.
[6] L. H. Epstein, T. Jimenez-Knight, A. M. Honan, R. A. Paluch, and W. K. Bickel, "Imagine to remember: an episodic future thinking intervention to improve medication adherence in patients with type 2 diabetes," *Patient preference and adherence*, pp. 95–104, 2022.
[7] L. H. Epstein *et al.*, "Effects of 6-month episodic future thinking training on delay discounting, weight loss and HbA1c changes in individuals with prediabetes," *Journal of Behavioral Medicine*, vol. 45, no. 2, pp. 227–239, 2022.
[8] A. Abbe, C. Grouin, P. Zweigenbaum, and B. Falissard, "Text mining applications in psychiatry: a systematic literature review," *Int'l j. of methods in psychiatric research*, vol. 25, no. 2, pp. 86–100, 2016.
[9] W.-H. Weng, K. B. Wagholikar, A. T. McCray, P. Szolovits, and H. C. Chueh, "Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach," *BMC medical informatics and decision making*, vol. 17, no. 1, pp. 1–13, 2017.
[10] K. Denecke and D. Reichenpfader, "Sentiment analysis of clinical narratives: A scoping review," *J, Biomedical Informatics*, vol. 140, p. 104336, 2023.
[11] K. Kreimeyer *et al.*, "Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review," *Journal of biomedical informatics*, vol. 73, pp. 14–29, 2017.
[12] A. Le Glaz *et al.*, "Machine learning and natural language processing in mental health: systematic review," *J. of Medical Internet Research*, vol. 23, no. 5, p. e15708, 2021.
[13] F. Burger, M. A. Neerincx, and W.-P. Brinkman, "Natural language processing for cognitive therapy: extracting schemas from thought records," *PloS one*, vol. 16, no. 10, p. e0257832, 2021.
[14] S. Maguen *et al.*, "Measuring use of evidence based psychotherapy for posttraumatic stress disorder in a large national healthcare system," *Administration and Policy in evidence Health and Mental Health Services Research*, vol. 45, pp. 519–529, 2018.
[15] K. Shoenbill, Y. Song, L. Gress, H. Johnson, M. Smith, and E. A. Mendonca, "Natural language processing of lifestyle modification documentation," *Health Informatics J.*, vol. 26, no. 1, pp. 388–405, 2020.
[16] A. Turchin and L. F. Florez Builes, "Using natural language processing to measure and improve quality of diabetes care: a systematic review," *J. of Diabetes Science and Technology*, vol. 15, no. 3, pp. 553–560, 2021.
[17] D. H. Smith *et al.*, "Lower visual acuity predicts worse utility values among patients with type 2 diabetes," *Quality of life research*, vol. 17, pp. 1277–1284, 2008.
[18] Y. Zheng *et al.*, "Identifying patients with hypoglycemia using natural language processing: systematic literature review," *JMIR diabetes*, vol. 7, no. 2, p. e34681, 2022.
[19] Y. Jin *et al.*, "Automatic detection of hypoglycemic events from the electronic health record notes of diabetes patients: empirical study," *JMIR medical informatics*, vol. 7, no. 4, p. e14340, 2019.
[20] S. Wang, F. Song, Q. Qiao, Y. Liu, J. Chen, and J. Ma, "A comparative study of natural language processing algorithms based on cities changing diabetes vulnerability data," in *Healthcare*, vol. 10, no. 6, 2022, p. 1119.
[21] X. Zhou, Y. Ni, G. Xie, W. Zhu, C. Chen, T. Wang, and Z. Pan, "Analysis of the health information needs of diabetics in China," in *MEDINFO 2019: Health and Wellbeing e-Networks for All*, 2019, pp. 487–491.
[22] H. Zhang, S. I. Goldberg, N. Hosomura, M. Shubina, D. C. Simonson, M. A. Testa, and A. Turchin, "Lifestyle counseling and long-term clinical outcomes in patients with diabetes," *Diabetes care*, vol. 42, no. 9, pp. 1833–1836, 2019.
[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.
[24] S. Ahmadi, A. Shah, and E. Fox, "Retrieval-based text selection for addressing class-imbalanced data in classification," *arXiv preprint arXiv:2307.14899*, 2023.
[25] A. A. Shah, "Leveraging Transformer Models and Elasticsearch to Help Prevent and Manage Diabetes through EFT Cues," MS thesis defended 5/4/2023, Virginia Tech CS, http://hdl.handle.net/10919/115452.
[26] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.
[27] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.
[28] T. Brown *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
[29] S. Min, M. Lewis, L. Zettlemoyer, and H. Hajishirzi, "MetaICL: Learning to Learn In Context," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 2791–2809.
[30] W. X. Zhao *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.
[31] J. Wei *et al.*, "Emergent abilities of large language models," *arXiv preprint arXiv:2206.07682*, 2022.
[32] A. Chowdhery *et al.*, "Palm: Scaling language modeling with pathways," *arXiv preprint arXiv:2204.02311*, 2022.
[33] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.
[34] V. Sanh *et al.*, "Multitask prompted training enables zero-shot task generalization," in *ICLR 2022-Tenth International Conference on Learning Representations*, 2022.
[35] J. Wei, M. Bosma, V. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," in *International Conference on Learning Representations*, 2022.
[36] H. W. Chung *et al.*, "Scaling Instruction-Finetuned Language Models," *arXiv preprint arXiv:2210.11416*, 2022.
[37] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.
[38] M. T. James Phoenix, *Prompt Engineering for Generative AI.* O'Reilly Media, Inc., 2024.