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

Authors: Sareh Ahmadi, Edward A. Fox

Presenter: Sareh Ahmadi, Virginia Tech, Email: saraahmadi@vt.edu
Blacksburg, VA 24061 USA

The Twelfth International Conference on Global Health Challenges, GLOBAL HEALTH 2023, September 25-29, 2023 – Porto, Portugal

Short Resume of Presenter

Education:

PhD student, Department of Computer Science, Virginia Tech August 2020 - Present

Skills:

- Machine Learning, Deep Learning, Natural Language Processing (NLP), Computer Vision.
- Proficient in Python (Pytorch, LangChain), Java, C++.

Work Experience:

Graduate Research Assistant, Virginia Tech 2020-present

- Research on large language models and conversational agents (AI assistant and chatbot systems).
- Design and implement machine learning algorithms for NLP tasks including semantic search, classification, and dialog generation.
- Analyze and improve model performance through experiments and evaluation.

Summer Graduate Student Developer, University libraries at Virginia Tech May-August 2022

- Machine learning operations (MLOps) for a biomedical image segmentation research project.

Research Highlights:

- Pioneer effort to build an AI Assistant aimed at promoting behavior change and health.
- Contributed to an NIH-funded project for analyzing medical texts and adapting large language models for classifications.
- Worked on neural retrieval systems, by leveraging transformer models for semantic-query search tasks.

Research Topics

- With a dedicated role as a Graduate Research Assistant at Virginia Tech, I have immersed myself in cutting-edge research on large language models and conversational agents.
- My primary focus is on the development of advanced AI assistants and chatbot systems tailored to cater to diverse domains and user needs.
- I have gained hands-on experience in designing and implementing machine learning algorithms specifically for Natural Language Processing (NLP) tasks. These tasks range from semantic search and classification to sophisticated dialog generation.
- My role demands a rigorous evaluation of model performance. Through systematic experimentation, I constantly endeavor to refine and enhance the AI's capabilities.
- My ultimate goal is to build an AI assistant that not only communicates effectively but also empowers its users across various domains.

Outline

- Background
    - Introduction, Research Question, Goals
    - Large Language Models, Zero-Shot and Few-Shot Learning
    - Foundation Model, Prompt Engineering
- Approach
    - Dataset, Categories
    - Method
- Results
- Conclusions

Introduction

- Maladaptive health behavior:
  - Actions/habits detrimental to physical / mental well-being
  - Resulting from lifestyle choices such as:
  - Smoking, excessive alcohol consumption.
  - Linked to type 2 diabetes (T2D), cancer, and cardiovascular / respiratory diseases.
- Promoting healthy lifestyle choices
  - Like regular exercise, balanced diet, stress management, and avoidance of harmful substances
  - Can reduce the risk of developing such diseases.
- Interventions focusing on behavior change, support, and education can help individuals adopt healthier habits.

Introduction cont’d

- One contributing factor to maladaptive health behavior is delay discounting (DD).
- DD = tendency to devalue delayed rewards in favor of immediate gratification.
- Episodic Future Thinking (EFT) can reduce delay discounting.
- EFT is a cognitive process: mentally simulating or envisioning future events in a detailed and vivid manner.
- The resulting “cues” are texts about personally significant future events.
- Studies confirm that EFT is an effective intervention to reduce DD and promote healthy behavior.

Research Question

- EFT mechanisms of action and the conditions that impact its efficacy are unknown.
- There are significant variations in the content characteristics of EFT cues, e.g.,
    - extent to which they form a coherent narrative or
    - describe different personal goals such as weight loss or family milestones.
- Cues differ in terms of imagery, vividness, emotional valence, and level of detail.
- **What makes EFT cues effective in improving health outcomes?**

Goals

- Gain a better understanding of how EFT works, and the factors that influence its efficacy.
- Enhance the effectiveness of EFT in preventing and treating T2D.
- Build natural language processing (NLP) classifiers to predict EFT content characteristics.
- Reduce the cost of annotating participant data for classification.
- Apply Large Language Models (LLMs) to health improvement methods.
- Validate application of zero-shot, few-shot, and in-context learning techniques within the emerging field of instruction-tuned models.

Large Language Models

- Fine-tuning was a dominant approach for building classifiers by further training the model on task-specific labeled data.
- Fine-tuning heavily relies on large amounts of data.
- Annotation is expensive, time-consuming, and infeasible without adequate data.
- Prompt tuning emerged as a response to address the limitations of fine-tuning.
- Prompt tuning leverages the pre-trained language model’s ability to generate text, by providing input prompts.
- Prompts, including task-specific descriptions or instructions, allow language models to perform new tasks without extensive labeled data.
- Constructing appropriate prompts (prompt engineering) enables prompt tuning.

Zero-Shot and Few-Shot Learning

- Zero-shot learning involves a language model performing a task without training.
- Few-shot (k-shot) in-context learning adapts a language model to a new task with a few labeled examples.
- K = number of examples (demonstrations).
- Instruction tuning improves the zero-shot and few-shot learning abilities of LLMs.
- Instruction-tuned LLMs can follow text instructions (with inputs and correct outputs).
- This fine-tuning can enhance performance on previously unseen tasks.
- Studies have shown instruction-tuned language models outperformed other models in zero-shot and few-shot frameworks.

Foundation Model

- Given the superiority of the instruction-tuned language models, we need to choose a pre-trained instruction-tuned model.
- Publicly released FLAN-T5 is the instruction-tuned version of the T5 encoder-decoder model that has undergone fine-tuning across a variety of tasks to follow instructions.
- It performs zero-shot NLP tasks, as well as few-shot in-context learning tasks.
- It excelled on difficult tasks in the BIG-Bench dataset.
- The FLAN-T5 11B model (11 billion parameters) outperforms larger language models (e.g., PaLM 62B).

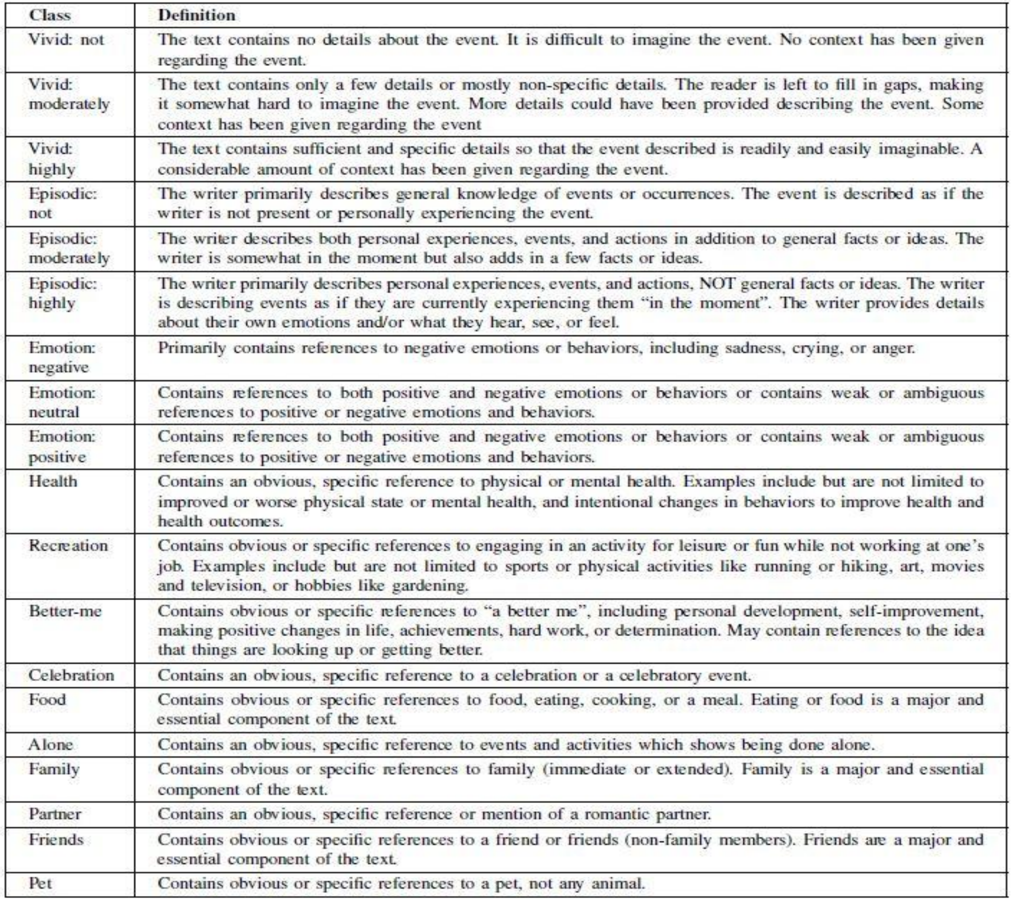
Prompt Engineering

- Instructions and context provided to a language model are within prompts.
- Instruction-tuned models require prompt engineering: the input to the model contains well-crafted prompts, ensuring meaningful guidance.
- Components that constitute a prompt are:
  - **Instruction**: Instruct/guide model on desired actions and use of external information, and outline the construction of the output.
  - **Context**: Gives supplementary knowledge for the model.
  - **Input Data:** Provided by a human user (i.e., the user input or query).
  - **Output Indicator**: Denotes the starting point of the to-be generated text.

Dataset

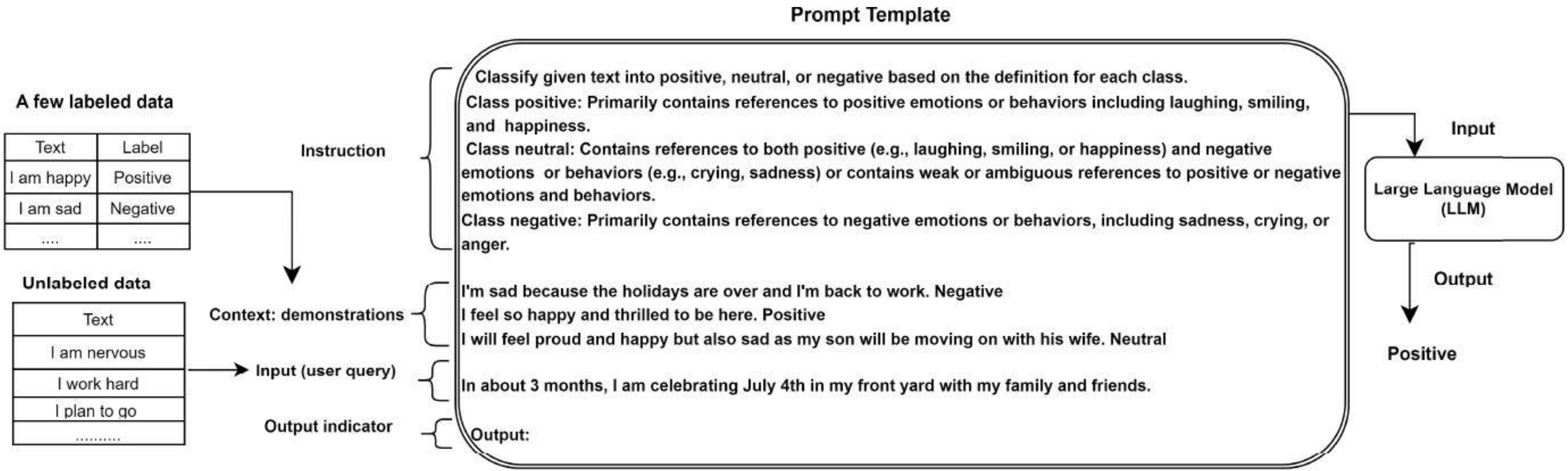
- In our dataset, participants write cues about the events for different time frames, ranging from one month to ten years. For example:
    - In about 1 month, I am playing golf with my friends. We are having a great time and enjoying the company and competition. We laugh and have a great time.
    - 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.
- A subset of the data is used for manual labeling in Amazon Mechanical Turk.
- Each text is labeled by three different annotators.
- The annotators are given the definition and an example for each category.

Categories



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.

Method



A few labeled data

Text	Label
I am happy	Positive
I am sad	Negative
....	....

Unlabeled data

Text
I am nervous
I work hard
I plan to go
....

Prompt Template

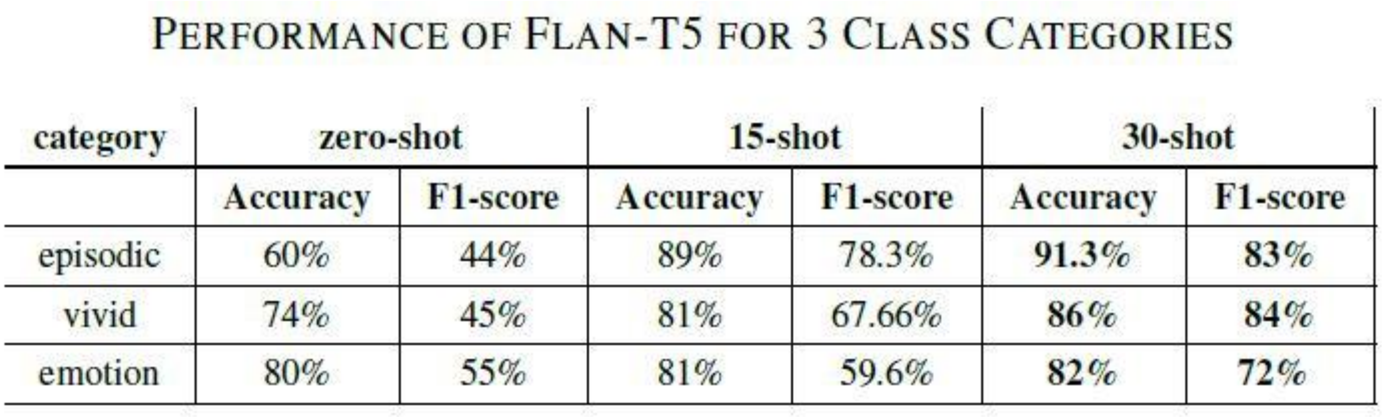
Classify given text into positive, neutral, or negative based on the definition for each class.
Class positive: Primarily contains references to positive emotions or behaviors including laughing, smiling, and happiness.
Class neutral: Contains references to both positive (e.g., laughing, smiling, or happiness) and negative emotions or behaviors (e.g., crying, sadness) or contains weak or ambiguous references to positive or negative emotions and behaviors.
Class negative: Primarily contains references to negative emotions or behaviors, including sadness, crying, or anger.

I'm sad because the holidays are over and I'm back to work. Negative
I feel so happy and thrilled to be here. Positive
I will feel proud and happy but also sad as my son will be moving on with his wife. Neutral

In about 3 months, I am celebrating July 4th in my front yard with my family and friends.

Output:

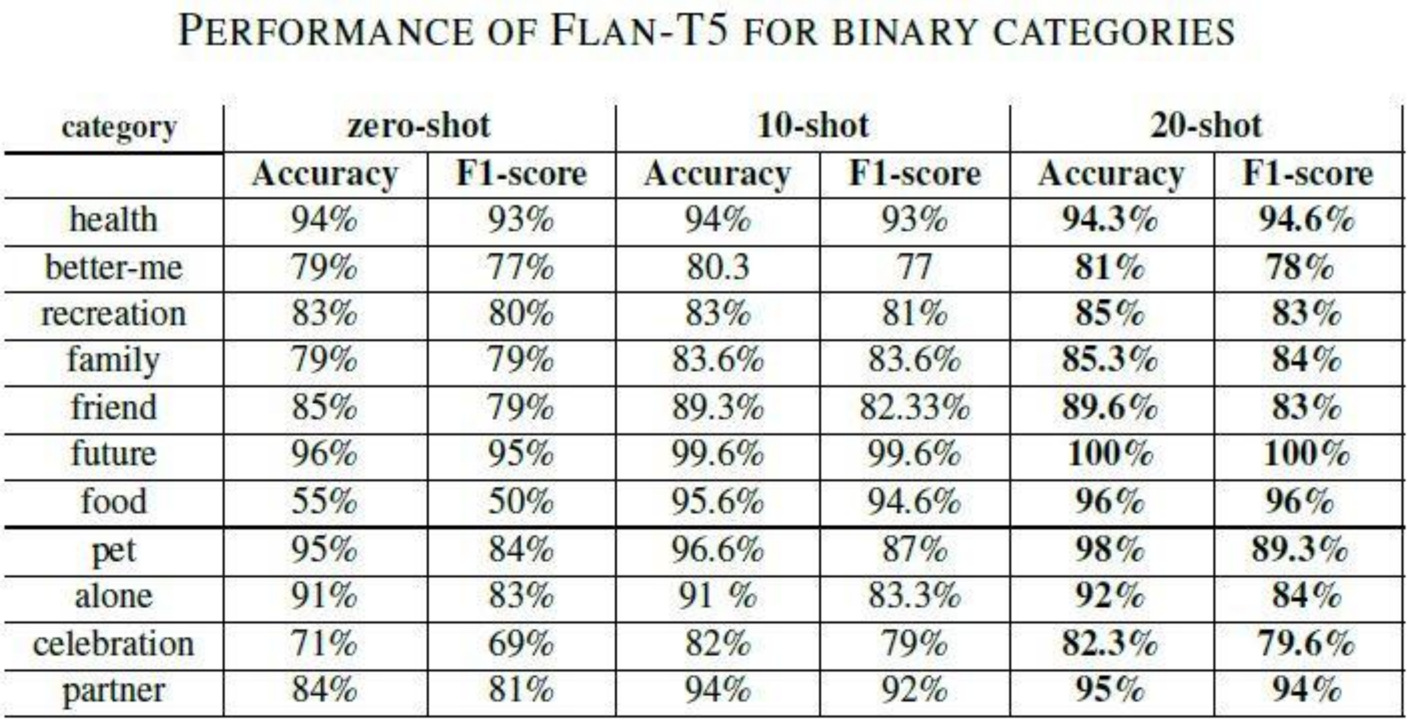
Results



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%**

Results cont’d



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%**

Conclusions

- We use a pre-trained instruction-tuned model, and few-shot in-context learning, for text classification.
- This avoids the need for a large amount of labeled data of traditional fine-tuning methods.
- We make good predictions regarding characteristics of cues.
- The method can be used when annotation is expensive, or when there is limited labeled data.
- The classifiers can aid in-depth analysis regarding what cue text features contribute to positive health outcomes.