In the name of God

RESEARCH METHODS - HW2

Author:
Marzieh Alidadi
810101236

Contents

1 Question 1.

The introduction of an article typically consists of five main parts:
1. Highlighting the importance and centrality of the topic
2. Reviewing previous works on the subject
3. Identifying any gaps in current knowledge
4. Stating the aim or objective of the research
5. Outlining the structure of the article

1.1 Importance and centrality of the topic:

- The related part in the article: "Generative AI (GenAI) is a class of machine learning (ML) algorithms that can learn from content such as text, images, and audio in order to generate new content. In contrast to discriminative ML algorithms, which learn decision boundaries, GenAI models produce artifacts as output, which can have a wide range of variety and complexity. One major recent development of GenAI is the introduction of OpenAI's GPT-3 [13] model, which can generate human-like language output and has striking versatility. Other generative language models have emerged that focus on specific domains such as software engineering, implementing use cases of auto-completing code [15, 39], translating code from one programming language to another [88], and converting natural language to code [23]. The industry has begun to use these models to support software engineering practices, with the most prominent example being GitHub CoPilot [27], a GenAI-based co-programming tool."
- The introduction begins by discussing the concept of Generative AI and its significance in machine learning algorithms. It also mentions recent developments in this field, such as OpenAI's GPT-3 model, which generates human-like language output.

1.2 Previous works review:

- The related part in the article: "This question of what do users need to understand about AI systems is core to the nascent field of Human-Centered Explainable AI (HCXAI) [21, 22, 52], which is a subset of the fields of human centered AI and human centered data science [6, 7, 25, 41, 65–67]. Our work is informed by a few key lessons from recent work in HCXAI, mostly conducted in the context of discriminative ML (e.g., for decision-support systems). First, explainability needs should be considered broadly as any means of helping users achieve a better understanding of the AI system. Liao et al. [48] proposed to define user's explainablity needs by what questions they ask to understand the AI [48] and developed a framework of common questions. This framework demonstrates that users are interested in a broad range of explanatory information about an AI model, including its overall logic, how it reasons to produce a particular output, the training data, and its performance and range of output. However, user needs regarding generative models were not explored in that work."
- The article reviews recent works in Human-Centered Explainable AI (HCXAI) and discriminative ML to lay the foundation for their research on explainability needs for GenAI in the context of generative code models.

1.3 Knowledge gaps:

- The related part in the article: "As a novel technology applied to novel domains, there are many open questions to be answered for how to make GenAI more capable and user-friendly. One open question is how to enable explainability—allowing users to understand and have a better mental model—of GenAI. Recent works by Goodfellow et al. [28] and Ross et al. [86] have explored developing more interpretable GenAI models that follow more human-understandable processes. However, a more comprehensive view of explainability for GenAI is lacking: what do users need to understand about a GenAI model in order to effectively achieve their goals when working with it? In this paper, we build a foundational understanding of explainability needs for GenAI in the context of generative code models."
- The author identifies an open question related to GenAI's explainability, specifically how to enable users to understand and have a better mental model of GenAI. The author notes that while recent works have explored developing more interpretable GenAI models, a more comprehensive view of explainability for GenAI is lacking.

1.4 Research objective:

- The related part in the article: "This question of what do users need to understand about AI systems is core to the nascent field of Human-Centered Explainable AI (HCXAI) [21, 22, 52], which is a subset of the fields of human centered AI and human centered data science [6, 7, 25, 41, 65–67]. Our work is informed by a few key lessons from recent work in HCXAI, mostly conducted in the context of discriminative ML (e.g., for decision-support systems). First, explainability needs should be considered broadly as any means of helping users achieve a better understanding of the AI system. Liao et al. [48] proposed to define user's explainablity needs by what questions they ask to understand the AI [48] and developed a framework of common questions. This framework demonstrates that users are interested in a broad range of explanatory information about an AI model, including its overall logic, how it reasons to produce a particular output, the training data, and its performance and range of output. However, user needs regarding generative models were not explored in that work. Second, XAI solutions that address explainability needs should not be limited to algorithmic explanations or showing model internals. Depending on user needs, it may be more critical to provide transparent information about a model's capabilities, limitations (e.g. uncertainty [10]) or provenance [8]. Moreover, users may need additional information beyond algorithmic explanations to fill in gaps of understanding. For example, Ehsan et al. [19] proposed that social transparency – making visible the socio-organizational factors that govern the use of AI – can help users form a sociallysituated understanding of an AI system and take more effective actions with it. Finally, perhaps the most important lesson from HCXAI is that users' explainability needs emerge in a usage context, guided by their goals and shaped by their backgrounds, expectations, as well as socio, organizational and cultural contexts [20, 48, 52]. It is thus necessary to follow a user-centered approach to understand explainability needs by involving target users and leveraging HCI methods that allow inquiry within the context of usage."
- The introduction states the aim of the research as building a foundational understanding of explainability needs for GenAI in the context of generative code models by involving target users and leveraging HCI methods.

1.5 Article structure:

- The related part in the article: "Based on these lessons, we adopted a scenario-based design method by constructing realistic usage scenarios for three use cases of GenAI for code: code translation, code auto-completion, and natural language to code. We invited 43 software engineers to participate in 9 workshops to elicit their explainability needs and design ideas around these scenarios. We adapted the question-driven method of Liao et al. [48, 50] to elicit and comprehensively explore participants' explainability needs by what kind questions they would ask in the scenarios. We also gathered feedback and design ideas from participants for four kinds of XAI features that we propose for the uses cases of GenAI for code. Our work makes three main contributions to the IUI community: (1) We identify 11 categories of explainability needs in the context of Generative AI (GenAI) for code, for which we provide definitions and examples. We further contrast these categories with previous XAI techniques for discriminative ML and discuss explainability needs unique to GenAI and code generation use cases. We believe we are among the first to explore users' explainability needs in an application domain of GenAI. (2) We propose four kinds of XAI features to support users of GenAI for code, based on prior work and adapted to the domain of code generation. These features are: AI documentation, indications of model uncertainty, visualizations of model attention, and social transparency. Based on participants' responses, we provide concrete design recommendations to operationalize these features. (3) Our work makes methodological contributions by combining scenario-based design, participatory design workshops, and a question-driven approach to elicit explainability needs. We also reflect on the values and limitations of this method to inform future work that explores GenAI in new domains."
- The introduction provides an overview of the article's structure, highlighting the main contributions of the paper, which include identifying 11 categories of explainability needs in the context of GenAI for code, proposing four kinds of XAI features to support users of GenAI for code, and making methodological contributions by combining scenario-based design, participatory design workshops, and a question-driven approach to elicit explainability needs.

2 Question 2.

The article is using a **topic-centric** previous works review in the introduction section. The authors provide an overview of the development and applications of generative AI (GenAI), including OpenAI's GPT-3 model and other generative language models that focus on specific domains such as software engineering. The authors also discuss recent works on developing more interpretable GenAI models, but they highlight the lack of a comprehensive view of explainability for GenAI, particularly in the context of generative code models.

Overall, the authors provide a broad summary of the previous works related to GenAI and its applications, rather than focusing on a specific author or group of authors.

3 Question 3.

There are three methods or types of identifying gaps in previous research, which are:
a) Highlighting inconsistencies or gaps in previous studies that have been conducted on the topic
b) Asking specific questions about previous research to better understand any gaps or limitations
c) Building upon the work of other researchers in the field to expand upon their findings and contribute new insights

...

The introduction of this article identifies gaps in previous research by highlighting the limitations and inconsistencies in previous studies, as well as by building upon the work of other researchers to expand upon their findings. The authors also ask specific questions about previous research to better understand any gaps or limitations.

3.1 a) Highlighting inconsistencies or gaps in previous studies that have been conducted on the topic:

- The related part in the article: "Prediction of cryptocurrency prices is a challenging task mainly because they represent a relatively new phenomenon (currency or asset) with high volatility. While researchers use standard time series forecasting methods to predict cryptocurrency prices, this method has limited applicability, especially when the time-series exhibit high volatility."
- By highlighting the limitations of existing researches, the authors identify a gap in the literature that their own research aims to address.

3.2 b) Asking specific questions about previous research to better understand any gaps or limitations:

- The related part in the article: " Additional parameters, like the sentiment of related news or micro-blogs, can improve the prediction models. There are various models for obtaining sentiments [1] with different precision and applicability. One of the simplest and most frequently used sentiment calculation models is VADER [2], applied in [3] for forecasting Bitcoin prices based on the news and Tweets related to Bitcoin. A similar approach used in [4] employs Tweets and Google Trends to predict Bitcoin and Ethereum prices. The micro-blogs from Reddit and Tweets related to Bitcoin are applied in [5] to predict the bitcoin price, and the VADER sentiment model of bitcoin-related tweets is used in [6]. Additionally, sentiment based on the Wall Street Journal and Financial Times Bitcoin-related news, using the bag-of-words model and a dictionary-based approach, is used in [7]. The methodology in [8] is based on the Naive Bayes based model for calculating sentiments of tweets. The Vector error correction models (VECM) model and data from tweets is proposed in [9] to predict bitcoin price. In [10], data from Facebook posts, Google Trends, and Dow Jones News is proposed to predict Bitcoin price using different regression/classification models and different techniques supported by IBM Watson."
- By stating that additional parameters such as sentiment analysis can improve prediction

models, implying that previous research may have neglected the use of sentiment analysis in cryptocurrency price prediction, the authors identify a gap.

3.3 c) Building upon the work of other researchers in the field to expand upon their findings and contribute new insights:

- The related part in the article: ”The recent advances in deep-learning, transfer-learning, and NLP have significantly improved sentiment extraction from financial news and texts [11]–[15]. In [13], Yang et al. present an inductive transfer-learning method using ULMFiT [16] for sentiment classification in financial texts. Their results show the effectiveness of inductive transfer-learning methodologies compared to traditional transfer-learning approaches. In [15], Zhao et al. present the superior performance of recent NLP transformers, BERT, and RoBERTA in sentiment analysis compared to dictionary-based models. In [17], Mishev et al. make a thorough performance evaluation of the known algorithms for sentiment analysis applied to financial headlines. They start the evaluation with specific lexicons for sentiment analysis in finance and gradually build the study to include word and sentence encoders, up to the latest available NLP transformers. Their results show that contextual embeddings produced by NLP transformers show superior results compared to the other methodologies and algorithms. The sentiment inferred from the sentiment analysis models applied to financial texts can be used to forecast stock prices [18]–[20], foreign exchange, and global financial market trends [21], [22], corporate earnings [23], as well as to predict the cryptocurrencies prices. This paper proposes a transfer learning-based methodology for Bitcoin price prediction based on sentiment analysis of finance micro-blogs. We use deep learning-based NLP transformers like RoBERTa [24], to create a model for calculating financerelated sentiment of micro-blogs. We then build a model for Bitcoin price prediction based on sentiments from the news and Bitcoin historical price data using modern approaches for time series forecasting (FbProphet [25] and XGBoost [26]).”
- By referencing previous studies that used sentiment analysis to predict Bitcoin prices and building upon their findings by proposing a transfer learning-based methodology using RoBERTa for sentiment analysis, the article is expanding upon previous researches and contributing new insights to the field of cryptocurrency price prediction.