Dream Interpretation through Machine Learning and NLP: A Research Proposal

Tanvee Proag (2422814)

University of Mauritius

Abstract. Dreams are a part of human beings' subconscious mind. The subconscious mind is the brain's ability to perceive and process information. This paper is a research proposal on the interpretation of dreams through the use of natural language processing and machine learning techniques. The aim is to have a better understanding of people's dreams, find patterns in dreams and observe correlation between dreams and mental health. The research question that will guide this study is *How can AI models be trained to interpret dream narratives?* It will do so by trying to identify emotional as well as symbolic patterns and the connection between dream content and the psychological states. This research adds to both AI advancements and psychological studies to understand the subconscious mind as it can help with mental health self-assessment. The results of this research could have several practical applications such as assisting therapists in understanding patients' subconscious thoughts and helping individuals in their self-reflection and in their personal growth.

Keywords: Dreams, Dream Interpretation, Mental Health, Psychotherapy, Artificial Intelligence, Machine Learning, NLP

1 Introduction

Human beings' ability to dream has been a source of intrigue for many centuries. The Babylonians and Sumerians attempted to unravel this mystery by categorizing dreams into *good* and *bad*, attributing them to their mythical gods and demons respectively. In ancient Greece, figures such as Antiphon believed that dreams are the gateway to prophecies and could serve as guidance. Meanwhile, the Chinese held that dreams involved the soul journeying outside the body during sleep [Wiki 2019].

In the modern era, dreams remain fascinating. Sigmund Freud's *The Interpretation of Dreams* posited that dreams reflect unconscious desires and unresolved conflicts [Freud 1932]. Carl Jung's *Structure and Dynamics of the Psyche* expanded on Freud's work, focusing on the collective unconscious and personal

symbolism in dreams [Jung 1981]. Today, dream interpretation continues to grow with diverse psychological and cultural approaches. It serves as an aid for increased personal awareness and also a therapeutic technique in psychotherapy.

Dreams are considered windows into the subconscious mind because they reflect unconscious thoughts, emotions and memories. Both historical and contemporary theories agree on the correlation between the nature of dreams and their connection to the subconscious. Today, neuroscientific research supports the notion that dreams are linked to the subconscious. During REM sleep, brain regions associated with emotion, memory and self-processing are active which suggests that dreams help integrate and consolidate emotional experiences into long-term memory. [Scalabrini et al. 2021] Dreams often reflect unconscious memories from early childhood and attachment-related experiences which are stored implicitly in memory. Thus, they can serve as a means of emotional self-state expression, helping to process and reorganize emotional experiences. Moreover, they facilitate communication between the conscious and subconscious mind, allowing for findings on thoughts and emotions. [Hinckley 2023]

Artificial Intelligence has the ability to revolutionize the field of dream interpretation by tapping into the subconscious mind and retrieving patterns. This integration of AI with traditional dream analysis techniques looks promising as it can improve scientists' understanding of dreams thus providing new tools dedicated to personal growth and self-awareness.

AI models have the ability to quickly analyze vast amounts of data, identifying recurring themes, emotions and symbols in dreams that may not be immediately apparent to human interpreters. This significantly reduces personal biases and offers a clear understanding of dream symbolism and significance. There exists AI tools that can process and interpret dreams quickly, making dream analysis more accessible and user-friendly. Platforms like *Dreamology* and *Temenos* provide free or low-cost access to AI-driven dream interpretation. [DemoWP 2024, Temenosdreams 2025]

2 Problem Statement

This research proposes to address the question: *Can AI models be trained to interpret dream narratives by categorizing them into emotional and symbolic patterns to figure out a dreamer's mental state?* Both explicit emotional expressions and implicit symbolic references need to be observed within dream reports.

The interpretation of dreams is a subjective process as it often relies on human intuition and cultural beliefs. This results in dream analysis being inconsistent and inaccessible to people who do not have expert knowledge. Despite rapid advances in the fields of artificial intelligence and natural language processing, there is currently a lack of automated systems that can accurately examine and evaluate dream narratives from user input with both symbolic and emotional complexity that represents the human psyche.

The main problem is the absence of AI models that can classify and interpret dream content into meaningful psychological categories. Dreams usually

present a complex combination of literal events, symbolic imagery and emotional undertones. As a consequence, their analysis becomes far more complex than conventional text-based emotion detection. Sentiment analysis models that exist currently are adept at identifying vague emotional sentiments. However, they struggle to account for what is abstract and dream narratives tend to be surreal sometimes. Symbols in dreams such as *water*, *animals* and *movement* carry meanings that are specific to cultures and to the individual as well. They require in-depth interpretation that is beyond the scope of simple sentiment analysis.

Moreover, dreams are fragmented and have non-linear structures which add another layer of complexity. Unlike traditional text data, dreams frequently lack coherent structure. This presents disjointed events that can defy existing NLP pipelines. Dreams, being non-linear, require advanced contextual modeling to connect elements into a cohesive interpretation. In addition, dream content have an emotional weight that can fluctuate rapidly. This is challenging as it requires models to capture multiple overlapping emotions within the same narrative.

The connection between dream content and the dreamer's psychological well-being is another gap. Psychological studies suggest that recurring dream patterns, nightmares and symbolic imagery can indicate some hidden mental health conditions [Heidt 2025]. Usually, they are related to anxiety, depression or trauma. However, current AI systems lack the ability to link dream reports with certain psychological concerns. This limits their ability to assess mental health. Developing an AI model that could solve this issue could outline an individual's emotional and mental state. It could be used to detect mental health issues early.

Basically, sentiment analysis models that exist now can detect general emotions in text but they do not account for the symbolic nature of dreams. This would require deep contextual understanding and the ability to decipher metaphoric language. Also, dreams frequently contain a mixture of emotions, symbolic imagery and fragmented narratives. This results in challenges for traditional NLP models to process. There is also limited research that explores how AI can close the gap between dream content and the hidden emotional and mental health states. This leaves a shortfall in knowledge in understanding the psychological implications of dreams.

3 Aims and Objectives

This study tries to create an AI model that combines emotional sentiment analysis with symbolic recognition. If this study is successful, it might transform the field of dream analysis as it would act as a useful asset to assess mental health. The research can contribute to both AI advancements and psychological studies to understand the subconscious mind.

Aims - to develop an AI-based system that can interpret dream narratives by identifying emotional as well as symbolic patterns and to explore the connection between dream content and the psychological states. This can help individuals to understand their subconscious thoughts.

Objectives - to collect and pre-process a dataset of dream reports and design and implement NLP-based machine learning models. These classify dream narratives into different emotional and symbolic categories. Then, by integrating sentiment analysis and symbolic recognition techniques, dream interpretation becomes comprehensive. The system's performance shall be compared against human interpretations and psychological benchmarks. The potential of the AI model in detecting early signs of mental health conditions through dream analysis are to be studied.

Specific - Machine learning algorithms that can identify both emotional and symbolic elements in dream narratives are to be implemented.
Measurable - The system's performance will be evaluated through precision, recall and accuracy metrics by comparing the AI-generated interpretations with expert human assessments.
Achievable - The model will use available NLP libraries, datasets and computational resources within the project's time frame.
Realistic - Pre-existing NLP technologies and publicly available dream datasets will be used to ensure that the system's development aligns with the available resources including time, technical skills and computational power.
Time Constrained - The project will be completed in well-defined stages including data collection, model development, testing and evaluation within the project duration and will accommodate for buffer time for debugging and unexpected delays

4 Rationale and Significance of the Study

There is a lack of consistent methodologies for analyzing dream reports. Therefore, there is a pressing necessity to close the gap between subjective dream interpretation and objective analysis using AI. Currently, approaches are dependent solely on human expertise and this can induce bias and limit accessibility. An AI model that is capable of dream interpretation would create a more uniform method for understanding subconscious thoughts.

This study will not address the broader philosophical aspects of dreams such as their potential connection to spiritual experiences. Instead, it will focus on the psychological and emotional dimensions of dreams using machine learning to find out patterns that may relate to mental health.

The overall aim of this study is to create a reliable AI-based system that automatically interprets dream narratives. It does so by offering understanding of the dreamer's emotional and mental state. This shall enable individuals to understand their dreams more effectively.

The research is important because of its benefits in mental health. As mental health issues are becoming increasingly more common over time, early detection and assessment tools are more important than ever. According to WHO, in 2019, approximately 970 million people worldwide were living with a mental disorder

[WHO 2024]. A study showed that in the UK, mental health-related issues led to 18.5 million lost working days in 2022 impacting the world on an economic level [Priory 2023]. An AI model that analyzes dream narratives could identify emotional distress or psychological disorders at an early stage.

This study adds value to to both AI research and psychological sciences. From an AI perspective, it will advance the use of NLP in interpreting abstract symbolic text. From a psychological perspective, it will count as research into the connection between dreams and mental health. This has the potential to lead to further interdisciplinary research.

The outcomes of this research could have several practical applications such as assisting therapists in understanding patients' subconscious thoughts, helping individuals in their self-reflection and in their personal growth, developing programs that provide personal dream analysis and adding to broader studies on the role of dreams in mental health and in emotional well-being. By simultaneously addressing both technical and psychological aspects, this study has the potential to help us understand the subconscious mind.

5 Research Questions

The research questions that will guide this study are ***How can AI models be trained to identify and categorize emotional and symbolic patterns in dream narratives?***, ***What is the relationship between dream content and the dreamer's emotional or mental state?*** and ***Can AI-based dream interpretation systems achieve accuracy comparable to expert human interpretations?***.

These three research questions act as the guides to the phases of data collection, model development and evaluation. They help ensure that the study focuses on its core aims and objectives. They also help in deciding the time frame of the study period based on its feasibility.

6 Literature Review

For data collection, there exist multiple data banks online.

Domhoff and Bulkeley established a dreambank (dreambank.net) comprising 20,000 dream reports and a spreadsheet program [Bulkeley et al. 2010]. These help in calculating the dream content ratios that were derived from the major Hall-Van de Castle categories. Even though the dreambank.net spreadsheet proved helpful in the study of dream content, there was a steep learning curve as researchers had to learn a detailed scoring system that was associated with the Hall-Van de Castle categories. Then, they had to code all dream reports prior to the calculation of normative ratios in the spreadsheet manually.

A study on dream content analysis used Dreamboard for its rich dataset of dream content [Das et al. 2022]. In the Dreamboard internet forum that started in 2012, users record and track their dreams online over time. After narrating

their post in text format, they then themselves categorize or make comments on their dreams. The entries they post comprise a mixture of standardized fields and open text fields to capture their personal dream narratives and themes. For confidentiality, identification codes are assigned once the data goes into the database. In 2017, the website made publicly available a database of almost 38,000 dreams of the last three years. The limitation of this study was that it only focused on dreams posted in English.

Reddit's 44,213 dream reports that consist of 217 topics and grouped into 22 larger themes can be used [Das et al, 2022]. It is believed to be the most extensive collection of dream topics currently - A crowd-sourced dataset of dream self-reports from the r/Dreams community on Reddit. Unlike traditional lab studies, dream experiences shared on r/Dreams are reported voluntarily and spontaneously, enabling the collection of a large set of dream reports and conducting an ecological study. Over 44K dream reports from more than 34K Reddit users were collected over the past five years.

Another study that followed the previous one used Reddit dream reports and justified that it was the best option because unlike prior dream research that relied heavily on traditional laboratory, survey and diary methods or dreams that were recorded by waking up people from REM sleep and asking for their input so as to increase dream recall, Reddit dream reports are not influenced by laboratory setting [Das et al. 2024]. Reddit dream reports unlike survey responses do not suffer from memory distortion since the dreamers write about their dreams immediately instead of waiting for a period of time. Reddit dream reports unlike diary studies are not as ecological with minimal retrospective bias but they are not limited by small sample size from the burden of participation. Another benefit of using Reddit is that it is a social media platform that is popular which means that its usage is growing in numbers and it might continue to be a source of dream content for analysis.

The first to carry out a study about dream analysis was Hall and Van de Castle who developed a classification system based on the scientific criteria that is objective, generalised and reproducible [Hall et al. 1966]. Topics can be validated by comparing them to the Hall and Van de Castle scale. The usual technique to gain insights from dream topics used to be by sifting through many reports manually. The Hall and Van de Castle method offers a supervised approach for content analysis by using predetermined categories. However, these are usually biased towards existing knowledge on dreams. In reality, dreams are weird in nature and they consist of impossible or unlikely events that are not common in the daily life of a real person. This means that dreams may not be found in the predetermined categories that already exist. That is why current approaches are likely to miss some aspects of dreams [McNamara et al. 2019]. Unsupervised methods may prove better suited for dream analysis than supervised methods for a better understanding of dreams to address this gap in knowledge.

7 Methodology

The data will be collected, preprocessed and cleaned before the modeling using machine learning techniques and sentiment analysis. Finally, the system will be evaluated using a set of metrics.

Data collection - The public dream database Reddit r/Dreams posts is to be used as it is extensive and keeps on growing. One advantage worth noting is that since the dreamers write down their dreams on the social forum, they can do so immediately after waking up which avoids bias caused by forgetting or memorizing not vividly enough. On top of that, surveys can be sent out to participants willing to note down their dreams. This might be beneficial for contextual symbolism as people from different cultures and backgrounds might assign different meanings to different symbols. This survey can be done through Google Forms and the participants can be kept anonymous. The data can be classified into different categories: *Nightmares*, *Happy dreams*, *Anxiety dreams*, *Symbolic dreams*, *Recurring dreams*.

Pre-processing - The text is to be cleaned by removing unnecessary elements such as stop words (e.g., 'and', 'the'), punctuation marks and special characters to simplify the text and to reduce computational complexity. Then, the text will be split into smaller units i.e. tokens, which are typically words or phrases that form the basis to analyze further. The NLTK (Natural Language Toolkit) library can be used to identify and label common dream symbols like *water*, *animals*, *death* and *flying*. It is important that these symbols are recognized now to link the dream content to their psychological meaning later. VADER can be used for sentiment analysis - it can classify the emotional tone of the dream narratives. This helps to identify the emotions associated with the dream content.

Feature Extraction - The raw text has to be transformed into a structured format that machine learning models can process. Bag of Words (BoW) can be applied to represent dream content numerically by counting word frequencies and then focusing on significant terms compared to common ones. AffectNet and NRC Emotion Lexicon are two pre-trained lexicons that are to be used to assign emotional labels to words. This allows the model to detect the hidden feelings within the dreams. For semantic similarity detection, BERT can be used to detect symbolic patterns in the text. For example, words like *falling* could be associated with anxiety or fear. This improves the model's ability to interpret.

Modeling - Machine learning and deep learning techniques can be used to model the system for dream analysis. The sentiment analysis model will use VADER to classify the overall emotional tone of the dreams. For example, *joy*, *sadness*, *fear* etc. The LSTM (Long Short-Term Memory) model will analyze sequences in the dream narratives to capture patterns and shifts in the dream sequences. A fine-tuned BERT model that is pre-trained will be used on the dream dataset to detect meaning related to symbols in the texts. For example, it may recognize that *falling* in a dream symbolizes anxiety.

Expected Output - The model should categorize dream narratives into different emotional categories e.g. *anxiety*, *fear*, *sadness*, *happiness* etc. The model

should provide explanations of dreams based on patterns it has learned and by symbolic associations. Common dream elements should be identified by the system e.g. *Death* may represent the fear of loss, *Flying* may represent the desire for freedom, *Water* may represent an emotional depth. By analyzing recurring patterns in dream narratives, the model may be able to predict and point out early signs of anxiety or depression.

Evaluation Metrics - The accuracy of the system can be assessed by using a few techniques. The F1-Score is a standard classification metric that can help to evaluate how well the model categorizes dream emotions when compared to a labeled dataset. A sentiment consistency score can measure how well the model's sentiment classifications align with dream interpretation theories. Furthermore, the model can be evaluated by certified psychologists. Experts can review the dream interpretations that are generated by artificial intelligence and compare them with their own interpretations. This will be an assessment of the system's credibility.

Limitations - There might be some challenges during the implementation phase. Dream interpretation might be too subjective as different people assign different meanings to similar dream symbols. This can result in inconsistencies between the AI model's predictions and human interpretations. Moreover, there is a lack of large annotated dream datasets which could limit the model's ability to make generalizations across a large number of dream narratives. Symbolic representations in dreams can be ambiguous in nature. A single symbol, like *water*, might have multiple interpretations depending on the context of the dream and the individual's personal experiences. In addition, integrating mental health predictions based on dream content without clinical diagnosis can create misdiagnosis. Therefore, it must be approached with caution.

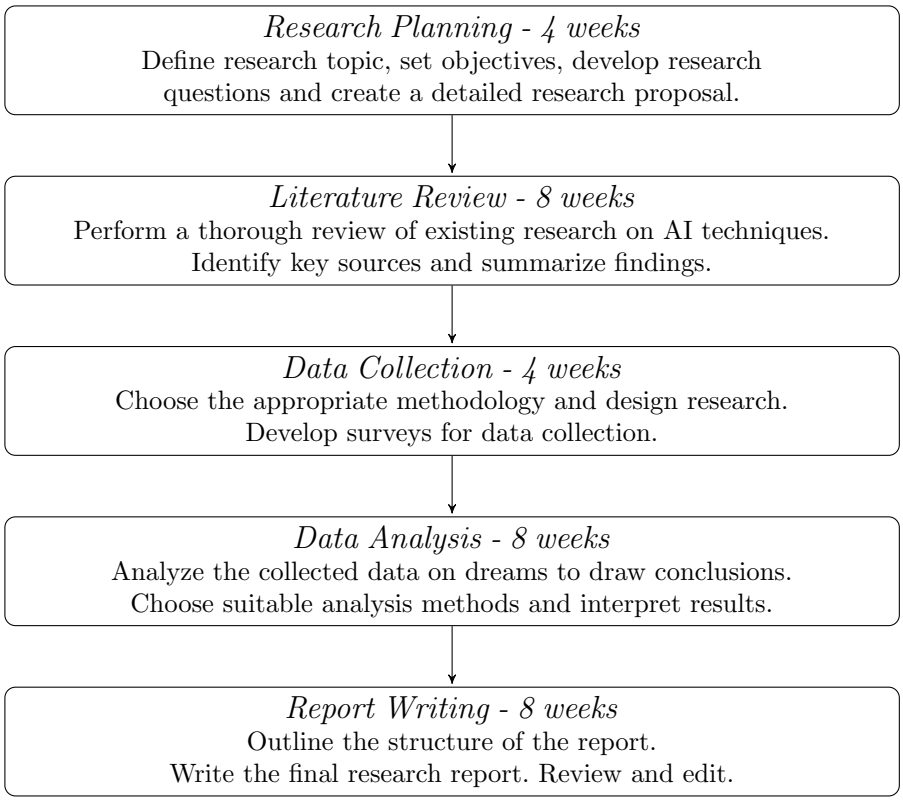
Justification - Even though these challenges exist, the potential benefits of this research outweigh the risks. The development of an AI-based dream interpretation system could add to emotional well-being of people and help with mental health self-assessment [Jenni et al. 2024]. The system is automated. This makes dream interpretation accessible and consistent for people who need it.

8 Ethical Statement

This research will adhere to high ethical standards in order to ensure the protection and well-being of participants. For confidentiality, all dream reports used in the study are to be made anonymous to protect the privacy of the participants. Any identifiable information must be removed before data processing. Participants will be fully informed about the purpose of the study, about how their data will be used and they will have the right to withdraw at any time. The study will be careful so as to avoid any discomfort caused by analyzing dream content that might be sensitive by providing support to participants who experience emotional distress. Secure encrypted servers that comply with data protection regulations including GDPR will be used to store the data collected. The study will ensure the transparency and integrity of the research process.

9 Time Frame Plan

The study can be completed within a 32-week time frame, with the project divided into phases:



References

1. Arxiv.org. (2022). Dream Content Discovery from Reddit with an Unsupervised Mixed-Method Approach. [online] Available at: `https://arxiv.org/html/2307.04167` [Accessed 28 Feb. 2025].
2. Bulkeley, K., & Domhoff, G. W. (2010). Detecting meaning in dream reports: An extension of a word search approach. Dreaming, 20, 77-95.
3. Das, A., Sanja Scepanovic, Aiello, L.M., Mallett, R., Barrett, D. and Quercia, D. (2024). Dream Content Discovery from Social Media Using AI. Research Square (Research Square). [online] doi:`https://doi.org/10.21203/rs.3.rs-4549419/v1`.
4. DemoWP. (2024). Dreamology Insights — Unlock the Hidden Meanings Your Dreams. [online] Available at: `https://dreamologyinsights.com/` [Accessed 1 Mar. 2025].
5. Freud, S. (1932). The interpretation of dreams. London: Allen & Unwin.

6. Hall, C. S., & Van De Castle, R. L. (1966). The content analysis of dreams. Appleton-Century-Crofts.
7. Heidt, A. (2025). Why Are Recurring Dreams Usually Nightmares? [online] Scientific American. Available at: https://www.scientificamerican.com/article/why-are-recurring-dreams-usually-nightmares/ [Accessed 1 Mar. 2025].
8. Hinckley, C. (n.d.). Looking into the window of your subconscious mind. [online] The Spellbinder. Available at: https://thespellbinder.net/8501/features/looking-into-the-window-of-your-subconscious-mind/ [Accessed 27 Feb. 2025].
9. Jenni, P. and Roesler, C. (2024). A New Way of Analysing Dreams on Its Profoundest Level: The Development of Motif Analysis and Phase Model (MAP) as an Extension of Structural Dream Analysis (SDA). Behavioral Sciences, [online] 14(8), pp.658–658. doi:https://doi.org/10.3390/bs14080658.
10. Jung, C.G. (1981). The structure and dynamics of the psyche. Princeton, N.J.: Princeton University Press.
11. McNamara, P., Duffy-Deno, K., Marsh, T. and Marsh, T.J. (2019). Dream content analysis using Artificial Intelligence. International Journal of Dream Research, [online] pp.42–52. doi:https://doi.org/10.11588/ijodr.2019.1.48744
12. Priory (2023). Mental Health Statistics. [online] Priory. Available at: https://www.priorygroup.com/mental-health/mental-health-statistics [Accessed 28 Feb. 2025].
13. Scalabrini, A., Mucci, C. and Esposito, R. (2021). Dreaming the unrepressed unconscious and beyond: repression vs dissociation in the oneiric functioning of severe patients. Research in Psychotherapy: Psychopathology, Process and Outcome, 24(2). doi:https://doi.org/10.4081/ripppo.2021.545.
14. Temenosdream.com. (2025). Temenos Dreams. [online] Available at: https://www.temenosdream.com/ [Accessed 1 Mar. 2025].
15. Wikipedia Contributors (2019). Dream. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Dream [Accessed 27 Feb. 2025].
16. World Health Organization (2024). Mental Health. [online] World Health Organization. Available at: https://www.who.int/health-topics/mental-health#tab=tab_1 [Accessed 28 Feb. 2025].