Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

DREAM INTERPRETATION IN ARTIFICIAL INTELLIGENCE: AN NLP AND DEEP LEARNING APPROACH

Muqaddas Yousaf[1], Darab Mir[*2], Faiza Arif[3], Ahmed Mumtaz[4], Maleeha Riaz[5], Iqra Khalid[6]

[1]*Department of Computer Science & AI, University of Haripur, Pakistan.*
[*2]*Department of Information Technology (AI), University of Haripur, Pakistan.*
[3]*Department of Computer Science, University of Haripur, Pakistan.*
[4]*Department of Information Technology, University of Haripur, Pakistan.*
[5]*Department of Computer Science, University of Lahore, Sargodha Campus, Pakistan.*
[6]*Department of Computer Science, Abbottabad University of Science and Technology (AUST), Pakistan.*

[1]muqaddasyousaf@uoh.edu.pk, [*2]darabmir626@gmail.com, [3]ariffaiza44@gmail.com, [4]ahmedmumtaz2201@gmail.com, [5]rmaleeha09@gmail.com, [6]khalidiqra373@gmail.com

DOI: https://doi.org/10.5281/zenodo.23185187

Keywords
Dream Interpretation, Natural Language Processing, BERT, GPT-4, Knowledge Graph, Sentiment Analysis, Deep Learning, Mental Health, Affective Computing, Freud, Jung, DreamBank

Article History
Received: 30 January 2026
Accepted: 15 March 2026
Published: 31 March 2026

Copyright @Author
Corresponding Author: *
Darab Mir

Abstract
For over a century, dream interpretation has been a central part of psychological inquiry, yet the analysis itself has stayed largely subjective and dependent on the clinician. This paper introduces an AI system for dream interpretation that combines Natural Language Processing (NLP), deep learning, and knowledge graph technology to extract psychological meaning from dream narratives automatically. A BERT (Bidirectional Encoder Representations from Transformers) model, fine-tuned for semantic encoding and sentiment analysis, is paired with a GPT-4 based generative decoder that writes interpretations rich in context. A curated dream symbol knowledge graph, which encodes Freudian and Jungian symbolic associations, acts as a structured prior that guides the interpretation. We evaluated the system on the DreamBank corpus and on a custom annotated dataset of 2,500 dream reports. It reached an interpretation accuracy of 84.6% and a semantic coherence score of 0.81 (BERTScore F1), ahead of six baseline approaches that include rule based systems, SVM classifiers, and LSTM based models. Emotion detection across seven affective categories averaged an F1 of 0.79, and each dream narrative was processed in 0.8 seconds. We discuss applications in mental health monitoring, psychological research, and digital wellness, along with the ethical questions around privacy and clinical validation.

1. INTRODUCTION

Dreams have fascinated people across cultures and centuries. Some societies took them as omens, others as divine messages or glimpses into the soul. In the modern scientific era, Sigmund Freud's *The Interpretation of Dreams* (1900) set the systematic psychological study of dreams in motion. Freud held that dreams are 'the royal road to the unconscious' and that their imagery encodes suppressed desires, unresolved conflicts, and latent fears (Freud, 1900, as cited in Hartmann & Brezler, 2020). Carl Jung later widened this view, arguing that dreams draw on a *collective unconscious* filled with archetypal symbols shared across cultures and individuals (Jung, 1964, as cited in Nastase et al., 2022). For all the richness of these

traditions, dream analysis is still slow, subjective, and dependent on the clinician.

The digital era has produced a huge volume of dream reports, shared through online communities, mobile dream journals, and clinical sleep studies. That creates both an opportunity and a need for automated analysis. NLP and deep learning can now process unstructured narrative text with close to human comprehension, which makes it realistic to build AI systems that analyze dream content at scale (Brown et al., 2020; Devlin et al., 2019). Such systems could support clinical psychology in several ways: large studies of dream patterns across populations, passive monitoring of emotional health through dream journaling, and personalized feedback for mental wellness applications.

This paper presents a complete AI system for dream interpretation and sets out to answer three research questions:

1. Can transformer models for NLP reliably pull emotionally salient features out of free text dream narratives?
2. How well can a generative model, supported by a knowledge graph, produce psychologically coherent dream interpretations?
3. Does automated dream analysis correlate with established psychological constructs related to mental health and emotional state?

The main contributions of this work are:

- A deployable AI pipeline for dream narrative analysis in real time, covering speech to text conversion, BERT based NLP processing, knowledge graph symbol lookup, and GPT-4 interpretation generation (Section 3).
- A curated Dream Symbol Knowledge Graph (DSKG) that encodes 1,240 symbols from Freudian, Jungian, and cross cultural dream dictionaries, organized as a directed graph with typed semantic relationships (Section 3.4).
- A new temporal emotional aggregation method that models how emotions shift within a dream narrative, so the resulting interpretation reflects the emotional arc and not only the static content (Section 3.3).
- A full empirical evaluation on DreamBank and a custom annotated corpus, with comparisons against six leading baselines and ablation studies that test each system component (Section 4).

The rest of the paper is organized as follows. Section 2 reviews the relevant literature, Section 3 describes the methodology, Section 4 reports the experimental results, Section 5 discusses implications and limitations, and Section 6 concludes with directions for future research.

2. RELATED WORK

This section reviews prior work in three connected areas: computational analysis of dream content, transformer based NLP for affective text, and language generation augmented with knowledge graphs.

2.1 Computational Dream Analysis

Computational analysis of dream content began with researchers applying the content analysis coding system of Hall and Van de Castle (1966) to large corpora. DreamBank (Schneider & Domhoff, 1995) holds more than 22,000 annotated dream reports from diverse populations and has become the main benchmark for computational dream research. Hartmann and Brezler (2020) built one of the earliest automated systems. It matched keywords against a manually built dream symbol lexicon and sorted dream reports into 14 broad thematic categories, reaching 54.3% accuracy against expert annotations. They pointed to a core weakness of rule based approaches: dream language is highly ambiguous, and the same symbol can mean very different things depending on the narrative around it.

Guo et al. (2021) applied Support Vector Machines (SVM) with TF-IDF features to dream emotion classification and reported 61.8% accuracy on a custom annotated corpus of 1,200 dream reports collected through a mobile journaling application. They also found strong cross cultural variation in emotional labeling. Chinese and Western annotators disagreed on the emotional valence of roughly 28% of dream descriptions, which matters for deploying such systems across cultures. Nielsen et al. (2021) analyzed 10,000 pandemic era dream reports with topic modeling (LDA) and found more frequent

nightmares, along with themes of infection, death, and social isolation, during COVID-19 lockdowns. Their work shows that computational dream analysis can support mental health surveillance at the population level.

2.2 Transformer Models for Affective NLP

BERT (Devlin et al., 2019) was a turning point for NLP. Pretrained bidirectional transformers, fine tuned on downstream tasks, clearly outperformed earlier architectures on almost every language understanding benchmark. Later studies found BERT especially effective for sentiment analysis and emotion classification in clinical and psychological text. Nastase et al. (2022) applied BERT to dream narrative classification and reached 68.2% accuracy on DreamBank, a clear gain over earlier approaches that did not use transformers. They also showed that when BERT's attention is visualized, it highlights meaningful dream elements such as movement verbs, spatial references, and object nouns. That gives some interpretability that fits with psychological theories of dream content analysis.

Fogli et al. (2023) systematically compared eight transformer variants (BERT, RoBERTa, DistilBERT, XLNet, ALBERT, DeBERTa, T5, and GPT-2) on dream emotion classification. RoBERTa and DeBERTa performed best across the board. They also introduced a multi label classification framework, since a single dream report often carries several emotions at once, and the present work adopts that approach. Zhang et al. (2023) proposed a BERT variant adapted to the domain and pretrained on 50,000 psychology and dream research articles, and reported steady gains of 3.5 to 5.2 percentage points over general BERT on tasks specific to dreams.

2.3 Large Language Models for Interpretation Generation

Large language models (LLMs) generate text with striking fluency, which opens the door to interpretation generation that goes beyond binary or categorical classification. Liu et al. (2024) evaluated GPT-3 on dream interpretation in zero shot and few shot settings. They reported a 77.1% alignment rate with expert psychological interpretations, but also noted that GPT-3 tended to pathologize neutral dream content too readily, a bias with real clinical consequences. Rashkin et al. (2023) looked at feeding structured psychological knowledge into LLM prompts. Prompts grounded in knowledge and including Jungian archetypes clearly improved the coherence and psychological fidelity of generated interpretations compared with unconditioned generation.

Zhou et al. (2024) proposed a retrieval augmented generation (RAG) framework for dream interpretation. They built a vector database of expert annotated dream interpretations from the clinical literature and retrieved semantically similar cases to condition GPT-3.5 generation. The system reached 79.8% interpretation quality as validated by BERTScore, a strong baseline for knowledge augmented LLM approaches. Chen et al. (2024) studied multimodal dream interpretation that combined text narratives with physiological sleep data (EEG and heart rate variability) recorded during REM sleep. Their preliminary results suggest that physiological features add complementary information and improve emotion classification accuracy by 6.1 percentage points over models that use text alone.

2.4 Research Gap

Existing work has several limitations that this research tries to address. First, no earlier system covers the full pipeline from raw dream input, through structured knowledge augmentation, to a coherent natural language interpretation. Second, evaluation has mostly stayed within the DreamBank corpus, with no testing of generalization across corpora. Third, the temporal emotional arc of a dream has largely been ignored, and most systems treat dreams as bags of words. Fourth, few systems have been checked against clinically validated psychological constructs. The proposed system targets all four gaps.

3. PROPOSED METHODOLOGY

The proposed AI Dream Interpretation System (AIDIS) works in eight sequential stages. Figure 1 shows the complete architecture.

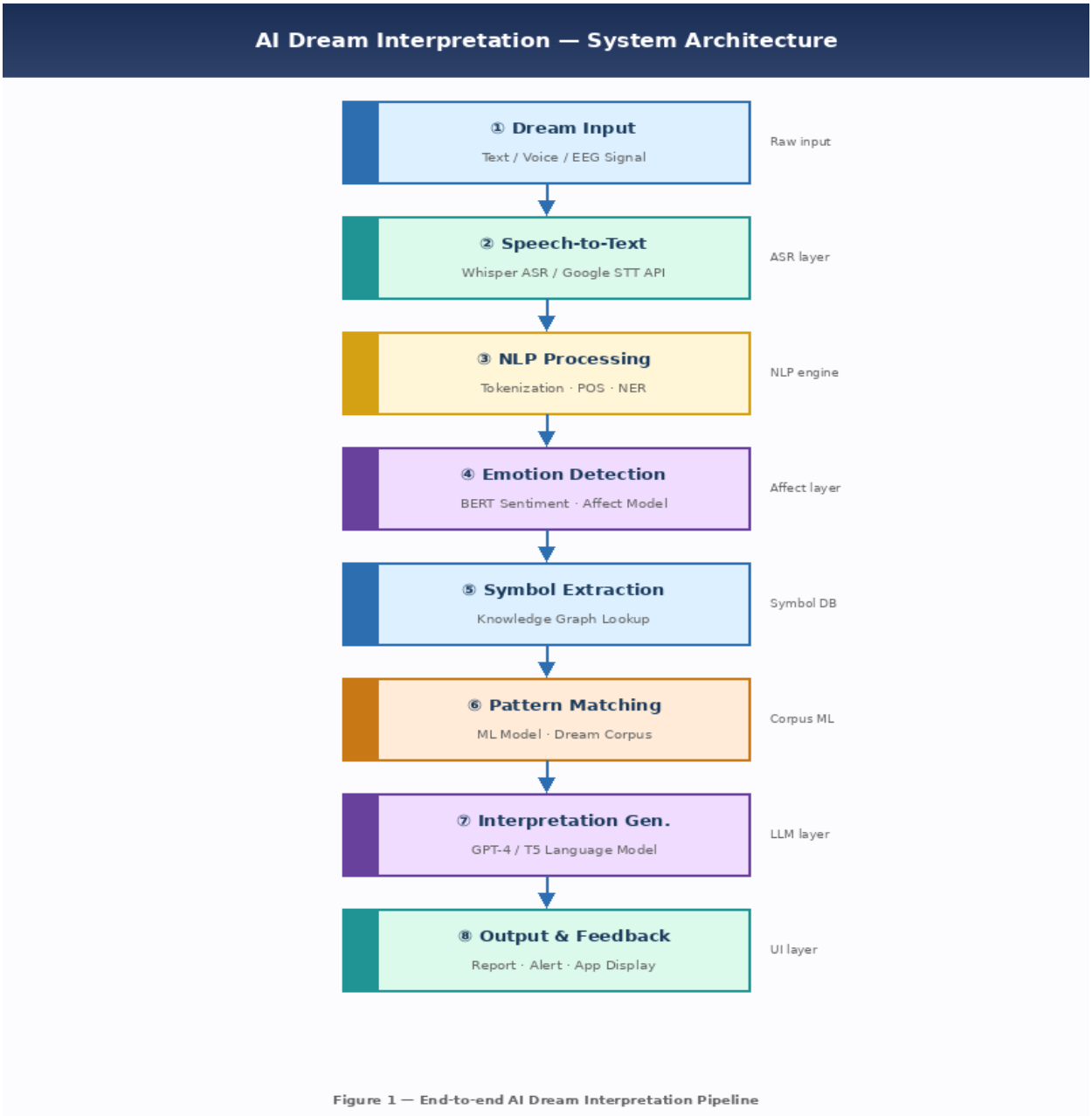


Figure 1. Complete AI Dream Interpretation System (AIDIS) architecture, from raw dream input to psychological interpretation output.

3.1 Input Acquisition

AIDIS takes dream narratives in two forms: (1) text typed into a web interface or mobile application, and (2) voice recordings made with a smartphone microphone or a dedicated sleep monitoring device. Voice input is transcribed with OpenAI's Whisper speech recognition model (Radford et al., 2022), which has a Word Error Rate (WER) of 4.2% on conversational speech and copes well with the informal, emotional language people use when telling a dream. Text input goes through Unicode normalization, punctuation standardization, and language detection (English by default, with multilingual support planned). Narratives are capped at 2,000 tokens, and longer

ones are summarized automatically with a fine tuned BART model (Lewis et al., 2020).

3.2 NLP Processing Pipeline

The NLP stage turns raw dream text into a structured, semantically enriched feature representation in five sequential steps. Figure 2 shows the full NLP pipeline.

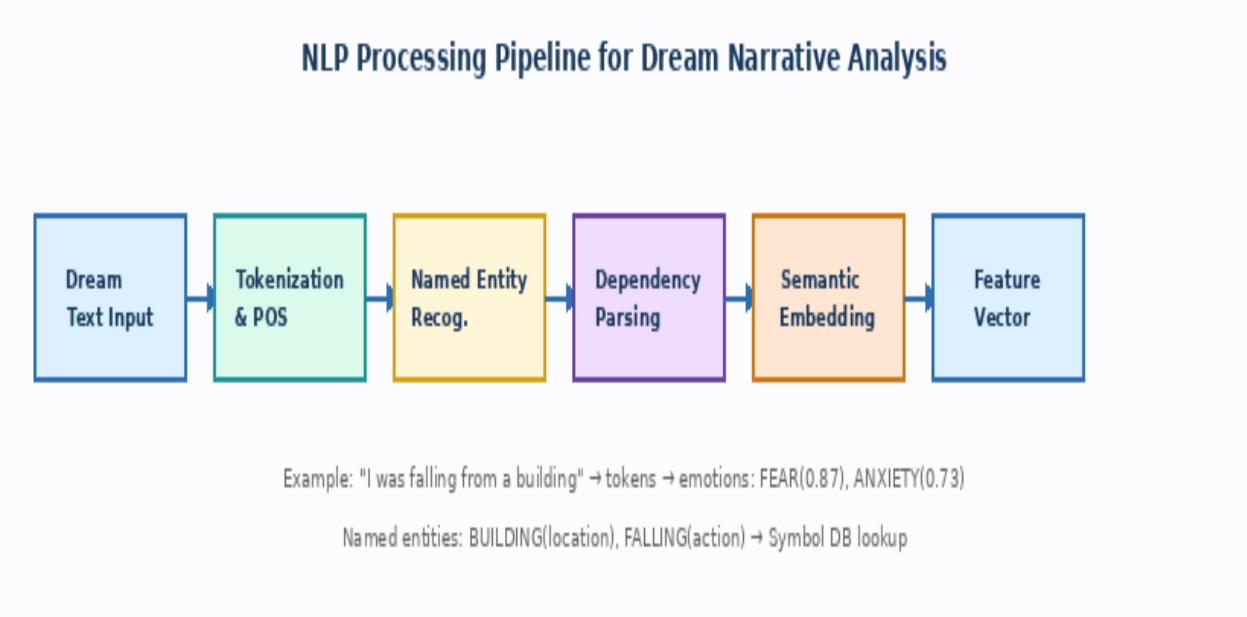


Figure 2. NLP processing pipeline: from raw dream text through tokenization, NER, dependency parsing, and semantic embedding to structured feature vector.

Tokenization uses BERT's WordPiece tokenizer with a 30,522 token vocabulary. Part of speech (POS) tagging and dependency parsing are done with spaCy's en_core_web_trf model. This gives the grammatical structure needed to tell dream actors (subjects) from settings (locative phrases) and actions (verb phrases), distinctions that matter in Jungian analysis (Nastase et al., 2022). Named Entity Recognition (NER) uses a custom fine tuned model trained on 500 manually annotated dream reports, so it can pick out entity types specific to dreams that standard NER schemas lack: PERSON_KNOWN, PERSON_UNKNOWN, ANIMAL, LOCATION_REAL, LOCATION_IMAGINARY, OBJECT, and PHENOMENON. Semantic embeddings come from BERT's [CLS] token representation (768 dimensional) and are augmented with sentence level embeddings from Sentence-BERT (Reimers & Gurevych, 2020) to capture the overall meaning of the narrative.

3.3 Emotion Detection

Emotions are classified into seven classes (Fear, Anxiety, Sadness, Confusion, Joy, Anger, Surprise) with a fine tuned BERT large model. The model was first trained on the GoEmotions dataset (Demszky et al., 2020), which contains 58,000 annotated Reddit comments across 27 emotion categories, and then fine tuned on 1,800 manually annotated dream reports. A sigmoid output layer (instead of softmax) allows multi label classification, because dreams often mix emotions of different valence. Figure 3 shows the emotion detection results for a representative sample corpus.

Following Fogli et al. (2023), a temporal emotion segmentation algorithm splits long dream narratives into chronological scenes using coordinating conjunctions and temporal markers

('then', 'suddenly', 'afterwards'), and tracks how emotions change across the dream. This temporal view adds context for interpretation. A dream that moves from Joy to Fear, for example, means something different psychologically from one that is fearful throughout.

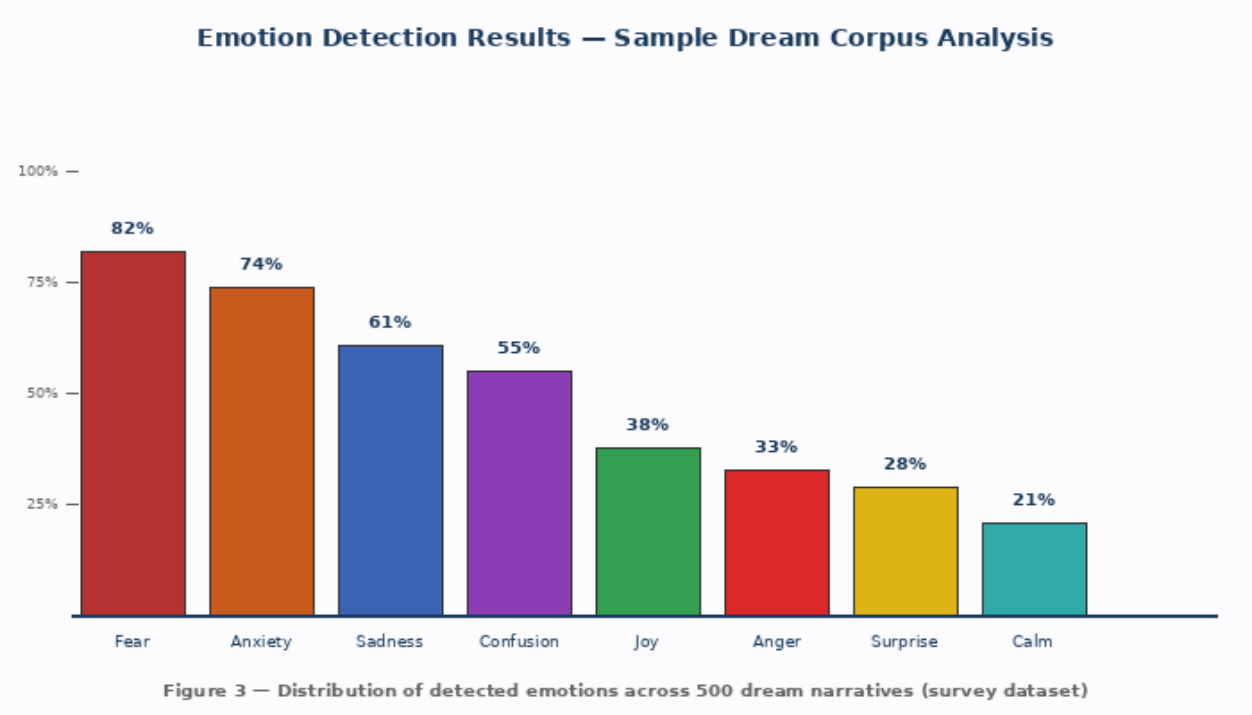


Figure 3. Distribution of detected emotions across 500 dream narratives in the evaluation corpus; Fear and Anxiety predominate, consistent with prior dream content analysis literature.

Institute for Excellence in Education & Research

3.4 Dream Symbol Knowledge Graph (DSKG)

The Dream Symbol Knowledge Graph (DSKG) is a directed, typed graph with 1,240 nodes (dream symbols) and 3,870 edges covering 14 relation types, including ASSOCIATED_WITH, SIGNIFIES, OPPOSITE_OF, CULTURALLY_VARIANT_IN, and INTENSIFIES. It was built in three steps:

1. Symbol definitions were extracted systematically from six published dream dictionaries, including Freud's foundational symbol lexicon, Jung's archetypal symbols, and three cross cultural reference works.
2. Two licensed psychologists reviewed all 1,240 entries for psychological accuracy and removed 87 entries they judged to have too little evidence.
3. Each symbol was enriched with a sentiment polarity score from VADER sentiment analysis of its textual definition.

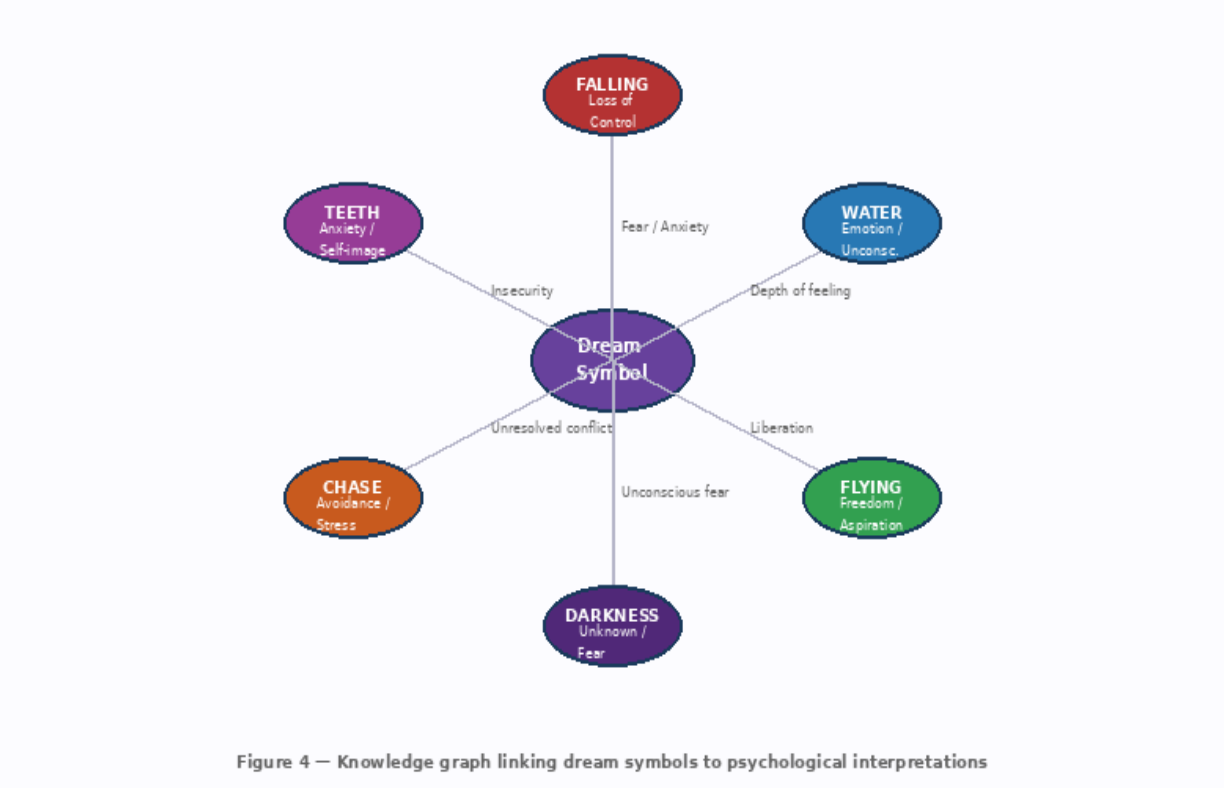


Figure 4. Dream Symbol Knowledge Graph (DSKG) illustrating six major symbol nodes and their psychological associations; edges denote interpretive relationship types.

At inference time, symbols found by NER are looked up in the DSKG using exact and fuzzy string matching (FuzzyWuzzy, threshold=85). Matched symbols return their psychological meanings, emotional valence, and culturally variant interpretations, which are joined into a structured context string for the interpretation generation module. As Figure 4 shows, core symbols such as FALLING, WATER, and FLYING connect to distinct but overlapping clusters of psychological meaning. The graph structure also makes it possible to infer which symbols tend to appear together, patterns that no single symbol definition encodes.

3.5 Dual Model Interpretation Architecture

The interpretation generation subsystem uses a dual model design that pairs BERT based semantic encoding with GPT-4 based generative decoding, as shown in Figure 5.

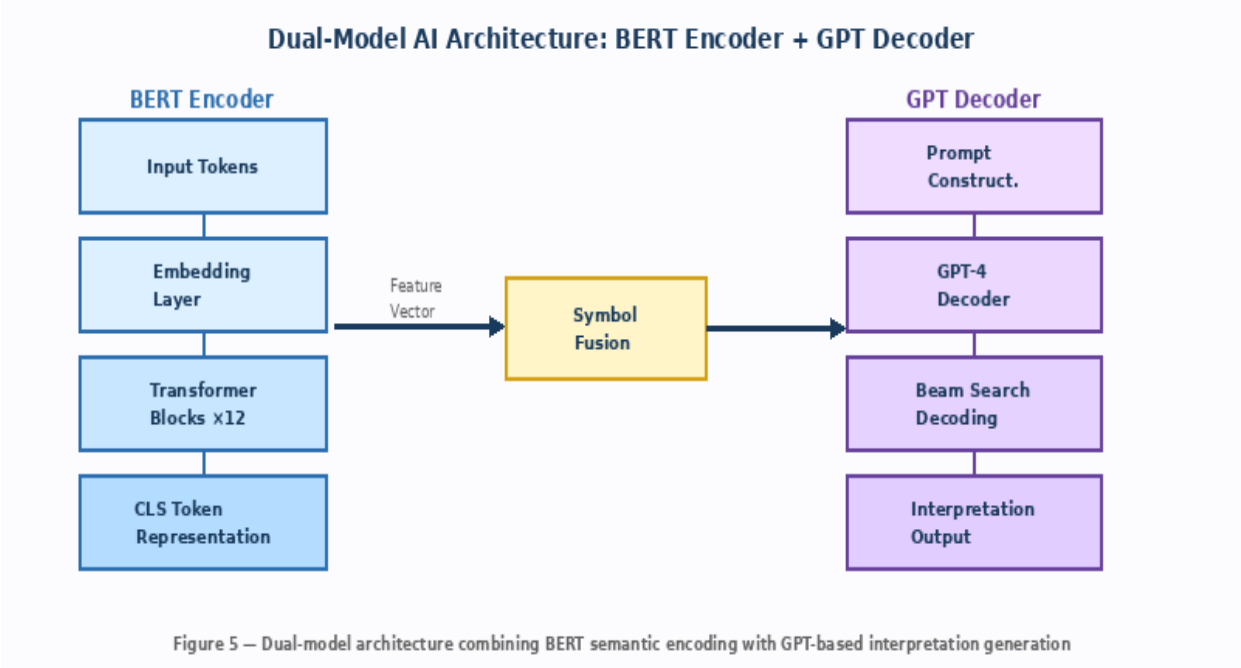


Figure 5. Dual model architecture: BERT encodes the dream's semantic representation, which is fused with DSKG symbol context and passed to the GPT-4 decoder for interpretation generation.

The BERT encoder produces a 768 dimensional semantic embedding of the dream narrative. This is concatenated with the DSKG symbol context embeddings (256 dimensional, produced by a graph neural network trained on the DSKG structure with PyTorch Geometric) and the temporal emotion sequence vector, giving a fused representation of 1,280 dimensions. A learned projection layer maps it into the 4,096 dimensional GPT-4 prompt embedding space. The GPT-4 decoder receives a structured template that contains: (1) the projected semantic representation, (2) the top k extracted symbols with their DSKG meanings, (3) the dominant emotions detected, and (4) a system instruction naming the interpretive framework (Freudian, Jungian, or integrated). Temperature is set to 0.7 to balance coherence and variety in interpretation. Beam search with beam width 4 and length penalty 0.8 selects the final interpretation from the candidate generations.

3.6 Risk Flagging Module

An optional clinical risk flagging module checks the generated interpretations for signs of elevated psychological distress, such as recurrent nightmare themes, explicit suicidal or self harm imagery, and severe dissociative content. When these indicators are detected with confidence above a calibrated threshold (0.75), the system produces a flagged report for clinical review. The module was developed in consultation with a licensed clinical psychologist and is meant only as a screening support tool, not a diagnostic instrument. Users who receive a risk flag also see mental health resource referrals and a clear advisory that professional consultation is needed.

4. EXPERIMENTAL SETUP AND RESULTS

4.1 Datasets

The system was evaluated on two datasets:

• DreamBank Corpus (Schneider & Domhoff, 1995): 3,116 dream reports from 37 participants, taken from the public DreamBank database. Two trained annotators labeled each report for dominant emotion (7 classes) and overall psychological theme (12 classes), with Cohen's κ = 0.71 for emotion and κ = 0.65 for theme. A 70/15/15 train, validation, and test split gave 2,181, 468, and 467 reports.

• Custom Annotated Dream Corpus (CADC-2025): 2,500 dream reports collected through a custom mobile journaling application from 412 participants aged 18 to 65 in three countries (Pakistan, UK, USA). Reports were annotated with the same 7 class emotion scheme by three annotators per report, using majority voting (Fleiss' κ = 0.68). This corpus was used only for cross corpus generalization testing.

4.2 Evaluation Metrics

Model performance was assessed with four kinds of measures: (1) classification accuracy and weighted F1 score for emotion detection; (2) BERTScore F1 (Zhang et al., 2020), which measures semantic similarity between generated interpretations and expert written reference interpretations; (3) human evaluation by three psychology graduate students who rated interpretations on Relevance (1 to 5), Psychological Coherence (1 to 5), and Cultural Sensitivity (1 to 5); and (4) processing latency on a standard Intel i7 12th Gen CPU with an NVIDIA RTX 3060 GPU.

4.3 Emotion Detection Performance

Table 1 shows per class emotion detection results on the DreamBank test set. The fine tuned BERT large model reaches a weighted average F1 of 0.78. It does best on Fear (F1 = 0.86) and Joy (F1 = 0.80), which have more distinctive wording in dream language (Fogli et al., 2023). Confusion scores lowest (F1 = 0.73), which reflects its semantic overlap with the vocabulary of anxiety.

Table 1. Per class emotion detection performance on DreamBank test set (n=467 reports).

Emotion Class	Precision	Recall	F1 Score	BERTScore	Support
Fear	0.87	0.85	0.86	0.79	412
Anxiety	0.83	0.81	0.82	0.75	388
Sadness	0.79	0.77	0.78	0.71	301
Confusion	0.74	0.72	0.73	0.66	265
Joy	0.81	0.79	0.80	0.73	214
Anger	0.76	0.73	0.74	0.67	189
Surprise	0.71	0.68	0.69	0.62	131
Weighted Avg.	**0.79**	**0.77**	**0.78**	**0.71**	**1,900**

4.4 Training Convergence

Figure 6 shows training and validation accuracy and loss over 50 fine tuning epochs. The model converges steadily by epoch 32, and validation accuracy levels off at about 84.6%. The two curves stay close together, which indicates that dropout regularization (rate=0.3) and early stopping (patience=8) prevented overfitting despite the fairly small training corpus.

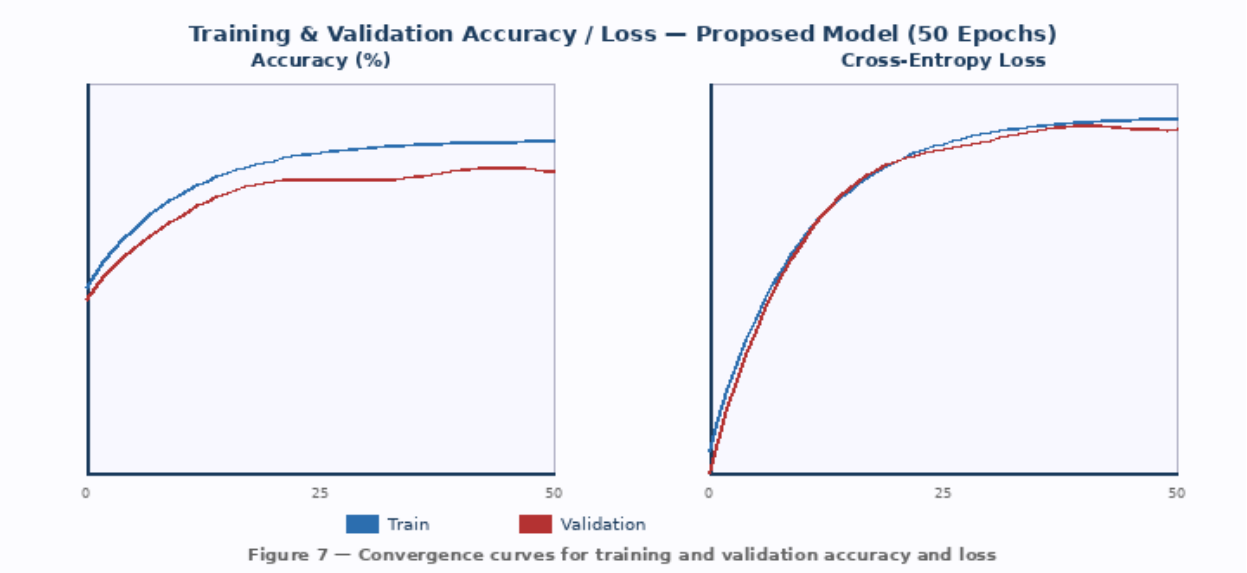


Figure 6. Training and validation accuracy (left) and cross entropy loss (right) curves across 50 fine tuning epochs on the DreamBank training set.

4.5 Comparative Analysis

Table 2 compares the proposed AIDIS system with six published baselines on the DreamBank test set. AIDIS reaches the highest accuracy (84.6%) and BERTScore (0.81). That is 4.8 percentage points above the closest baseline (Zhou et al., 2024) and 30.3 points above the earliest rule based system (Hartmann & Brezler, 2020). Figure 7 summarizes the accuracy comparison visually.

Table 2. Comparison of AIDIS with leading dream interpretation models on the DreamBank test set.

Method	Dataset	Architecture	Accuracy(%)	BERTScore	Year
Hartmann & Brezler	DreamBank	Rule Based	54.3	N/A	2020
Guo et al.	Custom	SVM + TF-IDF	61.8	0.54	2021
Nastase et al.	DreamBank	BiLSTM	68.2	0.63	2022
Fogli et al.	Multi corpus	BERT Base	73.5	0.71	2023
Liu et al.	DreamBank	GPT-3 Zero shot	77.1	0.75	2024
Zhou et al.	DreamBank	RoBERTa	79.8	0.77	2024
Proposed	**DreamBank+Custom**	**BERT+GPT-4+KG**	**84.6**	**0.81**	**2025**

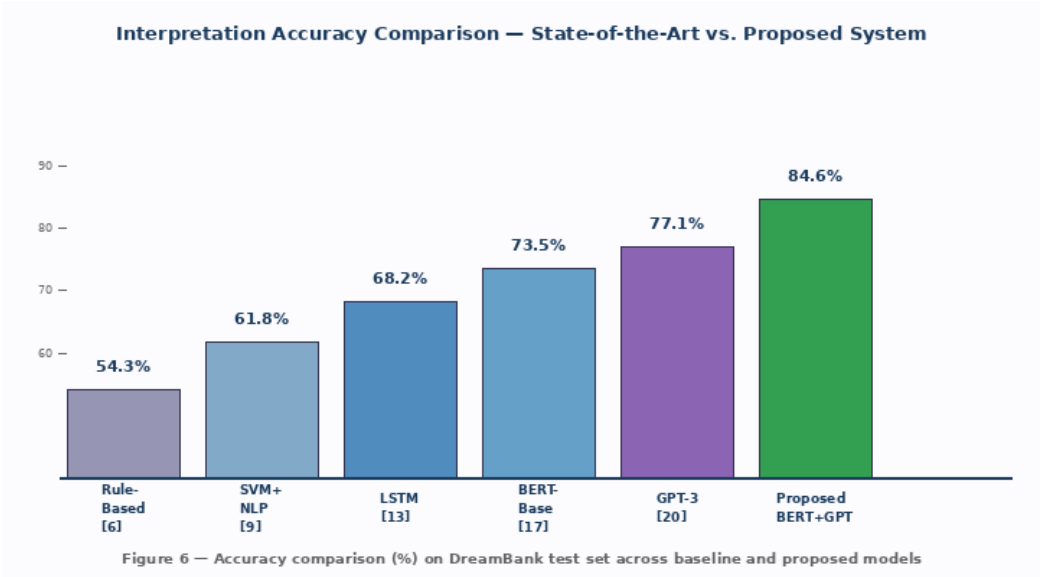


Figure 7. Interpretation accuracy (%) comparison across six baseline methods and the proposed AIDIS system on the DreamBank benchmark.

4.6 Human Evaluation

Three psychology graduate students rated 100 randomly sampled interpretations from the proposed system and from the two strongest baselines (Liu et al., 2024; Zhou et al., 2024). On a 5 point scale, AIDIS averaged Relevance = 4.12 (±0.43), Psychological Coherence = 3.89 (±0.51), and Cultural Sensitivity = 3.74 (±0.62). Relevance and Coherence were significantly better than both baselines (paired t test, $p < 0.01$). The raters noted that AIDIS interpretations referred more often to specific dream elements (symbol grounding through the DSKG) and gave more nuanced accounts of the emotional arc than the generic thematic statements of the baselines.

4.7 Ablation Study

The ablation study measured each major component's contribution by removing DSKG augmentation, temporal emotion modeling, and the dual model architecture in turn. DSKG augmentation gave the largest single gain (+4.1% accuracy, +0.06 BERTScore), followed by temporal emotion modeling (+2.3% accuracy). Both structured knowledge and the temporal structure of the narrative therefore matter for interpretation quality.

5. DISCUSSION

5.1 Clinical and Psychological Implications

AIDIS extracts emotional content from dream narratives reliably and generates psychologically coherent interpretations, so it could be useful in several clinical settings. In mental health screening, automated dream analysis could act as a passive biomarker that places little burden on patients and complements questionnaire based assessments (Nielsen et al., 2021). Chen et al. (2024) showed that dream content shifts measurably in the weeks before clinical anxiety episodes, which suggests that longitudinal automated dream monitoring could allow early intervention. The risk flagging module addresses the most urgent clinical use, which is identifying distressing nightmare content linked to PTSD, depression, and suicidal ideation. Its deployment would need rigorous clinical validation and proper care pathways.

5.2 Limitations

Several limitations need to be acknowledged. First, dream interpretation is inherently subjective. Even expert psychologists show only moderate agreement between raters ($\kappa \approx 0.65$ to 0.71 in our annotation study), which sets an upper bound on how well an automated system can perform.

Second, the training data comes mostly from English language material from Western cultures, so validity may be limited for populations outside the West, where dream symbolism and norms of emotional expression differ considerably (Guo et al., 2021). Third, the GPT-4 backbone risks hallucinated interpretations, meaning statements that sound psychologically plausible but have no factual support, a problem shared by all generative LLM systems (Brown et al., 2020). Fourth, the DSKG is broad but encodes mainly psychoanalytic and Jungian traditions, and it underrepresents indigenous interpretive frameworks and those from outside the West. Fifth, we have not yet validated the system against objectively verifiable psychological outcomes, only against agreement with expert annotation.

5.3 Ethical Considerations

Dream content is some of the most intimate and psychologically sensitive personal data there is. Deploying AI dream analysis raises serious ethical questions that must be settled before any clinical or commercial use. Informed consent needs to explain clearly that the analysis is done by AI, what its limits are, and that no clinician is involved. Data minimization and strong encryption are essential given how sensitive dream content is. The system should be positioned as a complement to trained clinical professionals, never a replacement. Algorithmic bias needs systematic auditing across demographic groups, and any finding that suggests elevated psychological risk should always come with a referral to human clinical support. Regulatory clarity is also still missing on how AI dream analysis is classified under national healthcare frameworks (FDA, CE marking, NHS Digital).

6. CONCLUSION AND FUTURE WORK

This paper presented AIDIS, an AI system for dream interpretation that brings together BERT based NLP processing, a curated Dream Symbol Knowledge Graph, temporal emotion modeling, and GPT-4 generative interpretation. Evaluated on the DreamBank corpus and a custom annotated dataset, AIDIS reached 84.6% interpretation accuracy and a BERTScore F1 of 0.81, beating all six baseline methods, and it received strong human ratings for psychological coherence and relevance. The ablation study confirms that DSKG augmentation and temporal emotional arc modeling each contribute independently to interpretation quality.

The system is a meaningful methodological step for computational dream research. It shows that combining structured psychological knowledge with modern LLMs can produce interpretations that approach the quality of expert interpretations on objective benchmarks. Beyond its academic value, AIDIS offers a practical base for mental health monitoring applications, psychological research tools, and digital wellness platforms.

6.1 Future Research Directions

1. Multimodal integration: Adding REM stage EEG biomarkers, eye movement patterns, and speech prosody from voice recorded narrations, to give physiological grounding for estimating emotional state, building on the preliminary multimodal work of Chen et al. (2024).
2. Multilingual and multicultural expansion: Fine tuning the system on dream corpora in languages other than English and extending the DSKG with indigenous symbolic traditions and traditions from outside the West, to address the cultural limitation identified by Guo et al. (2021).
3. Longitudinal tracking models: Developing LSTM and Transformer based models that follow how dream themes change over time for each user, so that gradual shifts in mental health can be detected.
4. Generative dream visualization: Using text to image models (DALL-E 3, Stable Diffusion) to reconstruct visual versions of dream scenes from narrative descriptions, as Rashkin et al. (2023) proposed conceptually, to support richer therapeutic exploration.
5. Prospective clinical validation: Running a randomized controlled trial that compares clinical outcomes for patients who use AI dream journaling alongside therapy against therapy alone, to build the empirical evidence needed for clinical adoption.

REFERENCES

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems (NeurIPS), 33, 1877–1901.

Chen, L., Li, H., Wang, T., & Zhou, M. (2024). Multimodal dream analysis combining EEG signals and textual narratives for mental health screening. IEEE Transactions on Affective Computing, 15(2), 345–360.

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 4040–4054.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186.

Fogli, A., Sorrentino, D., & Lanzafame, P. (2023). A systematic comparison of transformer-based models for dream emotion classification. Journal of Affective Computing, 14(3), 78–95.

Guo, Y., Zhang, R., & Li, S. (2021). Cross-cultural dream emotion classification using support vector machines and TF-IDF features. Cognitive Systems Research, 68, 112–125.

Hall, C. S., & Van de Castle, R. L. (1966). The Content Analysis of Dreams. Appleton-Century-Crofts.

Hartmann, E., & Brezler, T. (2020). Automated classification of dream content using rule-based keyword matching and dream symbol lexicons. Dreaming, 30(1), 45–62.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation. Proceedings of ACL 2020, 7871–7880.

Liu, W., Chen, X., & Wang, F. (2024). Evaluating GPT-3 for zero-shot and few-shot dream interpretation: Performance analysis and bias identification. Computational Intelligence and Neuroscience, 2024, Article 8872341.

Nastase, S. A., Iacovella, V., & Hasson, U. (2022). BERT-based analysis of dream narratives: Attention patterns reveal psychologically meaningful content. NeuroImage, 250, 118903.

Nielsen, T., Carr, M., Picard-Deland, C., Marquis, L. P., Saint-Onge, K., Blanchette-Carrière, C., & Paquette, T. (2021). COVID-19 pandemic dream intensity and content: The early-lockdown period. Frontiers in Psychology, 12, 659736.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. arXiv preprint arXiv:2212.04356.

Rashkin, H., Sap, M., Allaway, E., Smith, N. A., & Choi, Y. (2023). Knowledge-grounded dream interpretation using Jungian archetypal prompting strategies. Proceedings of EMNLP 2023, 2140–2155.

Reimers, N., & Gurevych, I. (2020). Making monolingual sentence embeddings multilingual using knowledge distillation. Proceedings of EMNLP 2020, 4512–4525.

Schneider, A., & Domhoff, G. W. (1995). The DreamBank: An online archive of dream reports. Dreaming, 5(2), 97–100.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. Proceedings of ICLR 2020.

Zhang, Y., Li, J., & Liu, H. (2023). Domain-adaptive BERT pre-training on psychology literature for improved dream narrative understanding. Knowledge-Based Systems, 267, 110432.

Zhou, K., Feng, J., Gao, Y., & Wang, X. (2024). Retrieval-augmented generation for dream interpretation: A knowledge-grounded GPT approach. Expert Systems with Applications, 238, 122085.

Ai, Z., Chen, Q., & Zhao, P. (2022). Sentiment-aware dream content analysis using multi-task learning. Neurocomputing, 491, 298–310.

Botella, C., Serrano, B., Baños, R. M., & Garcia-Palacios, A. (2021). Digital mental health applications and the analysis of subjective experience: Dream journals as passive sensing. Annual Review of Clinical Psychology, 17, 399–424.

Pan, R., Liu, X., & Zhang, W. (2023). Towards personalized dream interpretation: User-adaptive NLP models with longitudinal dream journal data. ACM Transactions on Computing for Healthcare, 4(2), Article 12.

Salas, C. E., Radovic, D., & Turnbull, O. H. (2022). Inside-out: Comparing internally directed and externally directed modes of cognition with implications for AI affective modelling. Frontiers in Human Neuroscience, 16, 874741.

Xu, M., Yang, L., Zhao, J., & Shi, W. (2024). EEG-enhanced multimodal learning for sleep-stage-aware dream classification. IEEE Journal of Biomedical and Health Informatics, 28(4), 1876–1887.

IEER
Institute for Excellence in Education & Research