Beyond Semantics: Information Retrieval with Emotion and Time

A Preprint

Joseph A. Evans*
Independent Researcher
josephev4@gmail.com

Abstract

Large Language Models (LLMs) have demonstrated strong performance on many Natural Language Processing (NLP) tasks. Following widespread adoption, research focused on improving their capabilities and finding solutions to drawbacks. One such drawback is hallucination: the phenomenon where a LLM generates misinformation with high confidence when asked about something that was not present in its training data. Retrieval Augmented Generation (RAG) addresses this by grounding LLMs with information from an external knowledge source. In traditional RAG systems, embedding models use pure semantic understanding to generate a vector for a similarity search against a knowledge base. However, this often overlooks the emotional and temporal nuances required for affective computing.

In this paper, we propose the Affective Link Score (ALS), a computation that represents the strength of a connection between events as a function of semantic, emotional, and temporal components. We further introduce Context Path Traversal, an algorithm for navigating memory graphs to retrieve affectively-grounded event sequences. Our empirical results demonstrate that our ALS framework achieves a 26.6% increase over uniformed baselines, converging to a stable validation loss of 0.6339. These results suggest that incorporating affective and temporal signals allows for a more robust reconstruction of narrative continuity than semantic retrieval alone.

1 Introduction

The introduction of the Transformer [1] led to breakthroughs in the field of Natural Language Processing (NLP). The attention mechanism allowed language models to better understand and generate natural language over larger bodies of text, compared to older architectures such as Recurrent Neural Networks (RNNs). Scaling up, we have many of the powerful large language models (LLMs) [2] today such as the GPT [3] and Gemini [4] series.

One of the issues that large language models deal with is hallucination [5], confidently generating misinformation, when prompted to respond to something that was not present in its training data. Many advances have been made in this area, notably Retrieval Augmented Generation (RAG) [6]. RAG is an information retrieval framework designed to augment a model's parametric knowledge (the knowledge gained during training) with non-parametric knowledge. This is achieved by indexing one or more external sources for storage and performing a semantic similarity search to select the most relevant text chunks during retrieval. Traditional RAG is highly-effective for grounding the model in factual information for question answering, but might fail to capture the finer details of human language, such as emotion, when centered around semantics.

Affective computing [7] is a growing, multidisciplinary field focused on improving the affective aspects of artificial intelligence. Emotion recognition, determining the emotions observed in textual, audio, and video content, is the key understanding task within this field, whereas emotional text generation and speech synthesis are the key generation tasks. Text generation within affective computing still relies on LLMs, inheriting the risk of hallucination. Where the center of generation is no longer pure semantics, but also emotion, the solution must also incorporate emotion.

*The method and system described in this paper are subject to US Provisional Patent Application No. 63/949,166, filed December 27, 2025.

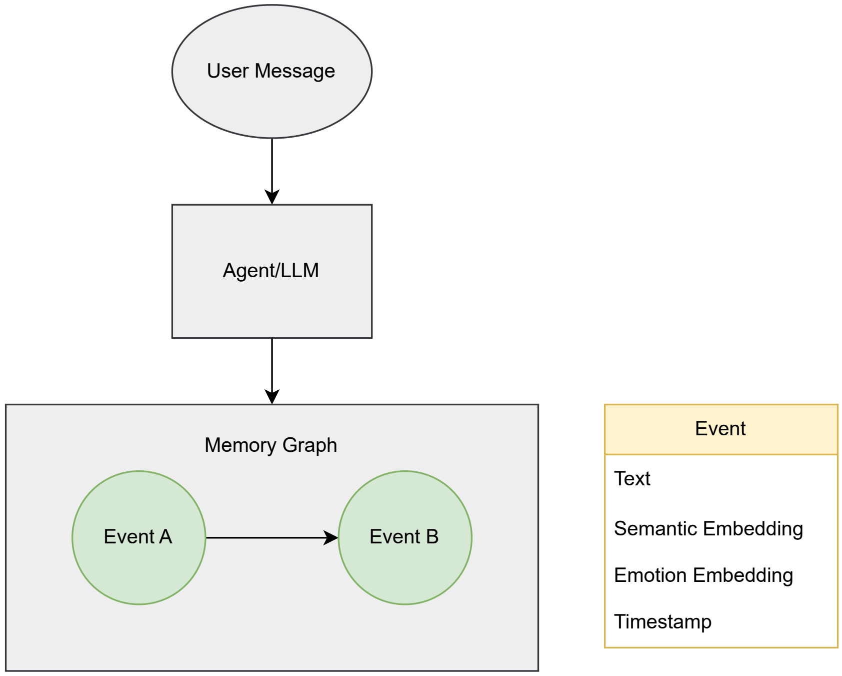


Figure 1: Memory graph creation - The agent processes input text to generate event nodes.

In this work, we propose the Affective Link Score (ALS), a formula that captures event resonance across the semantic, emotion and time dimensions as a bounded number (0,1), and Context Path Traversal, an algorithm that reasons which events should be included as context from a memory graph, optimizing context for agents with finite context windows. The activation of events during traversal aims to simulate the activation of related memories in human episodic memory.

2 System Overview

An agent populates a directed memory graph by encoding events extracted from a user message into nodes that contain event text, semantic embeddings, emotional embeddings, and a timestamp. Then it forms weighted edges between nodes based on the agent's understanding of the causal relationship of events. This is the foundation for the episodic memory retrieval system. This system allows us to move from a 1-dimensional retrieval process to a network of interwoven experiences.

3 Affective Link Scoring

To determine the edge weights for the interwoven events mentioned in Section 2, we propose the **Affective Link Score** (ALS), which represents the strength of the connection between two events as a weighted sum of semantic, emotional, and temporal components. We compute ALS as:

$$ALS(u, v) = \sigma(\omega_{\mathrm{S}} S_{\mathrm{S}} + \omega_{\mathrm{E}} S_{\mathrm{E}} + \omega_{\mathrm{T}} S_{\mathrm{T}} + b) \tag{1}$$

where $\omega_{\mathrm{S}}, \omega_{\mathrm{E}}, \omega_{\mathrm{T}}$ are learned weights representing semantic, emotional and temporal importance, S_{S} is the semantic similarity score for events u and v, S_{E} is the emotional similarity score for events u and v, S_{T} is the temporal proximity score for events u and v, and b is bias.

Table 1: Nomenclature of ALS Components and Context Path Traversal Hyperparameters

Symbol	Definition	Aspect
S_{S}	Semantic Similarity	Textual Alignment
S_{E}	Emotional Similarity	Emotional Consistency
S_{T}	Temporal Proximity	Chronological Distance
ω_n	Learned Weights	Feature Importance
b	Bias	Model Tuning
K	Seed Nodes	Retrieval Width
D	Path Depth	Narrative Length

3.1 Semantic Similarity (S_{S})

Semantic similarity measures the semantic alignment of the text content of two events. In the context of agentic memory, this component is essential for grounding a language model with relevant facts, ensuring that retrieved memories are aligned with the current topic. We define S_{S} as the cosine similarity between the semantic embeddings for events u and v:

$$S_{\mathrm{S}} = \frac{E_{\mathrm{u}}^{\mathrm{sem}} \cdot E_{\mathrm{v}}^{\mathrm{sem}}}{\|E_{\mathrm{u}}^{\mathrm{sem}}\| \|E_{\mathrm{v}}^{\mathrm{sem}}\|} \tag{2}$$

where $E_{\mathrm{u}}^{\mathrm{sem}}$ is the semantic embedding generated for event u and $E_{\mathrm{v}}^{\mathrm{sem}}$ is the semantic embedding generated for event v.

3.2 Emotional Similarity (S_{E})

Emotional similarity measures the emotional alignment of the events the agent ponders. While semantic similarity measures factual alignment, this component ensures that the agent maintains emotional continuity between events during retrieval. By prioritizing affective resonance, we can replicate a human-like activation of related memories that share similar feelings even if the topic has changed. We define S_{E} as the cosine similarity between the emotion embeddings for events u and v:

$$S_{\mathrm{E}} = \frac{E_{\mathrm{u}}^{\mathrm{emo}} \cdot E_{\mathrm{v}}^{\mathrm{emo}}}{\|E_{\mathrm{u}}^{\mathrm{emo}}\| \|E_{\mathrm{v}}^{\mathrm{emo}}\|} \tag{3}$$

where $E_{\mathrm{u}}^{\mathrm{emo}}$ is the emotion embedding generated for event u and $E_{\mathrm{v}}^{\mathrm{emo}}$ is the emotion embedding generated for event v.

3.3 Temporal Proximity (S_{T})

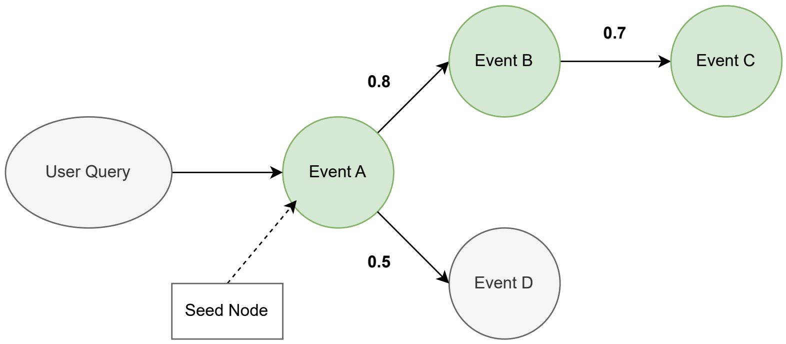
Temporal proximity measures how closely two events occurred in time. Directly causal events often occur in close succession to each other. By prioritizing temporal proximity during retrieval, the agent is able to weave together events that maintain narrative flow. We define S_{T} as:

$$S_{\mathrm{T}} = \frac{1}{1 + |t_{\mathrm{u}} - t_{\mathrm{v}}|} \tag{4}$$

where t_{u} is the time recorded for event$u$ and t_{v} is the time recorded for event v.

4 Context Path Traversal

The information retrieval process begins when a user message is received. The agent generates a semantic embedding for the query and runs a similarity search against the nodes in the memory graph to select the top-k (a configurable hyperparameter) seed nodes, ensuring the contextual entry points are topically relevant. Starting from each seed node, we traverse the memory graph selecting unvisited edges whose combined weights maximize the path's total ALS until an exit condition (e.g., maximum path depth) is reached. We call the path that results a context path. The nodes along each context path contribute a semantic and emotionally-rich narrative that can be selected for use as context for the final content generation task.



Edge	Weight	Cumulative Score
AB	0.8	0.8
BC	0.5	1.5

Path	Context Contributed
A	[Text from A]
A,B	[Text from A,B]
A,B,C	[Text from A,B,C]

Figure 2: Context Path Traversal - The agent selects weights that maximize the total ALS score.

Algorithm 1 Context Path Traversal via Affective Link Scoring

Require: Seed Nodes $\mathcal{K}$, Max Depth D, Learned Weights $\{\omega_\mathrm{S}, \omega_\mathrm{E}, \omega_\mathrm{T}\}$, Bias b, Sigmoid Function σ
Ensure: Context Path $\mathcal{P}$

```
n ← argmax_{k∈K}(S_S(Query, k)) {Select the best starting point}
P ← {n}
while |P| < D do
    N ← Neighbors(n)
    if N is empty then
        break
    end if
    for all m ∈ N do
        Compute S_S, S_E, S_T for edge (n, m)
        ALS_nm ← σ(ω_S S_S + ω_E S_E + ω_T S_T + b)
    end for
    n* ← argmax_{m∈N}(ALS_nm) {Select the most resonant neighbor}
    P ← P ∪ {n*}
    n ← n*
end while
return P
```

5 Experiment Design

5.1 Evaluation Framework

The goal of Context Path Traversal is to augment semantically-relevant context with information that allows the language model to produce affectively resonant and time-aware responses. As such, we want to model the evaluation to determine how emotionally intelligent responses are when using our algorithm. However, Context Path Traversal relies on ALS as its decision mechanism, which means that we must first test ALS before the full retrieval framework.

We will structure the evaluation into the following two tasks:

1. Binary classification for edge formation
2. LLM-generated ratings for emotional intelligence within responses

5.1.1 Binary Edge Classification

ALS produces a value between 0 and 1 which represents the probability that a connection forms between two memories. As that number moves closer to 1, the model's confidence that an edge will form between two memories increases. The

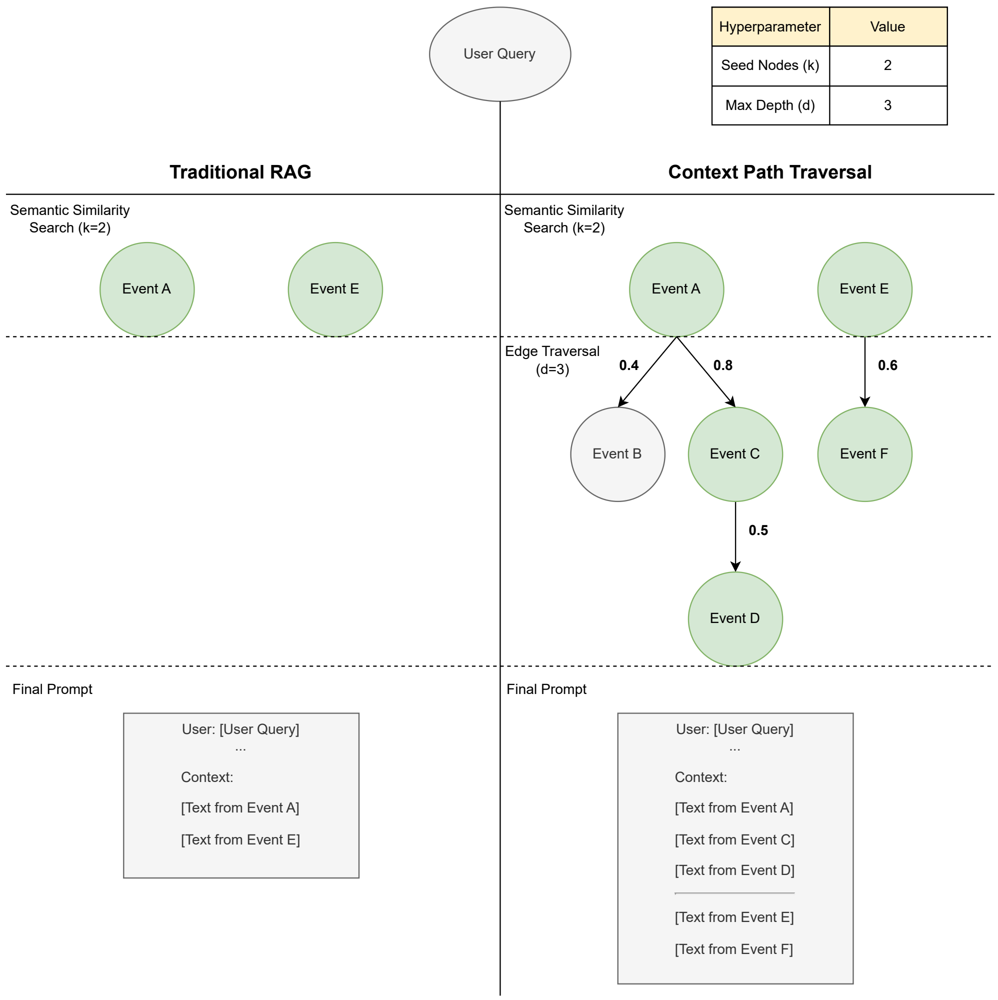


Figure 3: RAG vs. Context Path Traversal - The agent selects weights that maximize the total ALS score.

opposite is true when the number tends to 0. Since the number is bound (0,1), we will train an ALS model to predict whether an edge exists between two nodes in a memory graph.

5.1.2 LLM-Generated Response Rating

In order to evaluate the full retrieval process, we will setup a pipeline that takes a user query, augments that query with information retrieved from a knowledge base and pass that to a LLM to generate a response. The response will be evaluated by a LLM against a fixed set of criteria (eg. relevance, empathy and coherence) to determine a rating between 0 and 1 for how emotionally intelligent it is.

The above pipeline will be ran against a large set of queries and memories (knowledge entries), using both traditional RAG and the Context Path Traversal in parallel to highlight potential differences. For both approaches, the same queries, knowledge entries, and hyperparameters will be used. For the reason that RAG uses top-k to determine the number of chunks retrieved and Context Path Traversal uses top-k to determine the number of seed nodes, in addition to a maximum depth parameter for how many events can be accumulate along a context path, we will choose values that target a common maximum context window instead.

5.2 Data Generation

We use a LLM to generate a synthetic dataset of nodes and edges for the memory graph and a set of queries that are topically or emotionally relevant to the memories. To best simulate human-like memory retrieval, the data used must comprise human-like memories.

We generate a set of five stories (A-E) consisting of common life events for a variety of personalities and circumstances. Some stories emphasize more emotionally impactful events and others have minimal fluctuations. The agent extracts events from the stories and stores them in a serialized, directed graph structure that can be loaded for training and inference. Examples are below:

```
{
    "event_id": 0,
    "global_id": 0,
    "story_id": "A",
    "timestamp": "2025-01-10T15:00:00Z",
    "event_text": "I just received a promotion offer, but it requires relocating away from my best friend. I feel a mix of elation about the career step and deep sadness about the personal loss.",
    "semantic_tags": [
        "career achievement",
        "personal sacrifice",
        "major decision",
        "relocation"
    ],
    "emotional_state": "Conflict (Joy & Sadness)",
    "semantic_vec": [...],
    "emotional_vec": [...],
    "id": 0
}
```

Listing 1: Event from story A

```
"edges": [
    {
        "weight": 0.19047851318626563,
        "source": 0,
        "target": 2
    },
    {
        "weight": 0.8406625865942473,
        "source": 0,
        "target": 3
    }
]
```

Listing 2: Causal edges from story A with placeholder weights

5.3 Model Architecture and Training

ALS is the core computation that drives the memory graph retrieval approach. To determine the influence each component has in the memory selection process, we train a single-layer perceptron to learn the weights from a large dataset of known causal edges. The model is composed of 3 inputs, a weight for S_{S}, S_{E}, and S_{T}, respectively. Finally, the output is processed through a sigmoid activation function to compress the value to a range of 0 to 1. After training, the individual weights can be extracted for direct use within the ALS computation. Here is the transformation applied:

$$ALS = \sigma(\omega_{\mathrm{S}} S_{\mathrm{S}} + \omega_{\mathrm{E}} S_{\mathrm{E}} + \omega_{\mathrm{T}} S_{\mathrm{T}} + b) \tag{5}$$

where σ is the logistic sigmoid function defined by:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

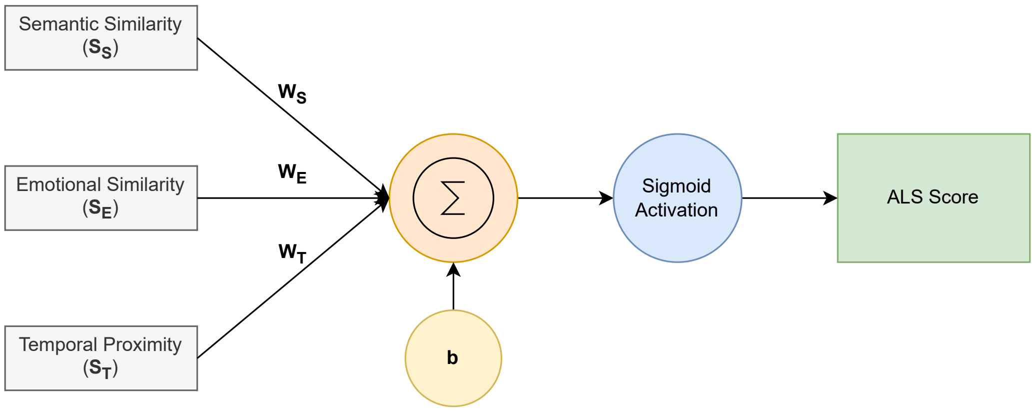


Figure 4: Single-layer perceptron - A weighted sum is performed before adding the bias term and passing the output through a sigmoid activation to get the ALS value.

6 Results

6.1 Binary Edge Classification (Complete)

Table 2: ALS Model BCE Loss Metrics

Metric	Baseline (Pre-train)	Convergence (Epoch 2)	Final (Epoch 30)
Training Loss	0.8637	0.6551	0.6403
Validation Loss	0.8637	0.6425	0.6339

Table 3: ALS Model Weights

Weight	Value
Semantic Weight (ω_S):	0.0791
Emotional Weight (ω_E):	-0.5179
Temporal Weight (ω_T):	3.1470

Table 2 summarizes the loss metrics for our ALS model when trained on the binary classification task of edge causality with Binary Cross Entropy (BCE) loss. Before training, the model's loss was 0.8637, performing worse than a random guess (0.693). During training, the model converged rapidly to the 0.64 range by epoch 2, reaching a final loss of 0.6339 at epoch 30. The model achieved a 26.6% improvement compared to the baseline, successfully capturing the non-linear relationship between the affective features and link causality. This is visualized by the loss curve in Figure 5.

After training, the core weights (ω_S, ω_E, ω_T) were extracted from the model to determine which component (semantic similarity, emotional similarity, or temporal proximity) is most closely related to edge formation. As shown in Table 3, the weight associated with temporal proximity, 3.1470, had the greatest value, with a significant gap between it and the other weights. It should be noted that, while temporal proximity had the greatest statistical impact on the model's performance, the range of values that it can produce vary greatly compared to those of cosine similarity used in the embedding comparisons. This presents a challenge for time normalization and unit selection (seconds, weeks, etc.) according to the task.

Semantic similarity was the second most influential factor, followed by emotional similarity, with 0.0791 and -0.5179, respectively. This indicates that the signal for the feature with the lowest weight, emotional similarity, functioned as noise when temporal proximity was large and the model penalized it. A similar trend was observed for identical tests ran on the stories, both individually and together, where temporal proximity was the greatest, and either semantic similarity or emotional similarity followed.

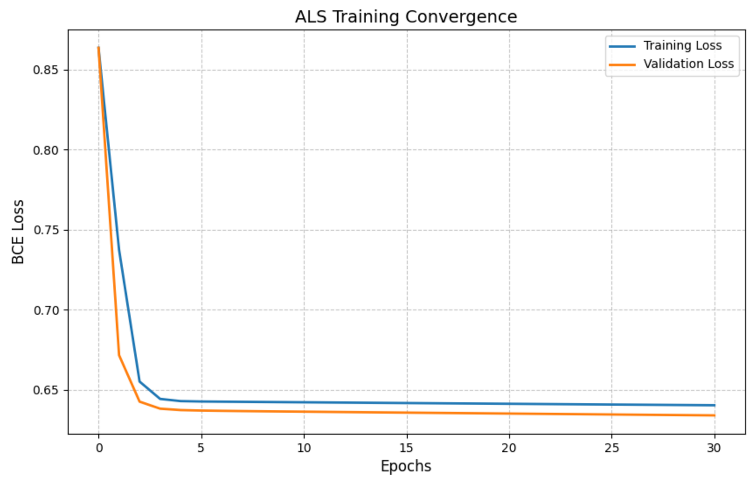


Figure 5: ALS Model Edge Classification Loss Curve

6.2 LLM-Generated Response Rating (In-Progress)

The qualitative evaluation (Task 2) is currently in-progress. This stage will use a LLM-as-a-judge technique to rate the affective content generated by augmentation through Context Path Traversal versus the standard top-k baselines.

7 Conclusion

In this work, we introduced the Affective Link Score (ALS) and Context Path Traversal as a novel framework for multi-dimensional information retrieval. By integrating semantic similarity, emotional similarity and temporal proximity into a memory graph retrieval system, we enable language models to generate affectively-relevant content while maintaining narrative flow. Task 1 demonstrates a 26.6% improvement in memory connection prediction over uninformed baselines, with the dominant weight (ω_T = 3.1470) indicating that temporal proximity is the primary factor in episodic coherence. This work opens new research avenues in affective computing. Future work will explore spreading activations through non-causal memory paths and evaluate the qualitative impact of emotion state transitions on user satisfaction through Task 2.

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[2] Humza Naveed, Asad Ullah Khan, Shi Qiu, et al. A comprehensive overview of large language models, 2024.

[3] OpenAI et al. Gpt-4 technical report, 2024.

[4] Gemini Team et al. Gemini: A family of highly capable multimodal models, 2025.

[5] Lei Huang et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55, January 2025.

[6] Patrick Lewis, Ethan Perez, Aleksandra Piktus, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *CoRR*, abs/2005.11401, 2020.

[7] Guanxiong Pei, Haiying Li, Yandi Lu, et al. Affective computing: Recent advances, challenges, and future trends. *Intelligent Computing*, 3:0076, 2024.