International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A GRAPH FRAMEWORK FOR METAPHOR CREATION INTEGRATING KNOWLEDGE GRAPH CONSTRUCTION AND GENERATION

Vibhavari Kamble[1,*], Yashodhara Haribhakta[2]

[1]Computer Science and Engineering, COEP Technological University, Pune, India

vvk20.comp@coeptech.ac.in

[2]Computer Science and Engineering, COEP Technological University, Pune, India

ybl.comp@coeptech.ac.in

*Corresponding author: vvk20.comp@coeptech.ac.in

July 30, 2025

Abstract

Metaphorical language enables humans to express abstract concepts through concrete experiences, making it an essential facet of natural language. Despite its prevalence, computational understanding and generation of metaphor remain fundamentally challenging due to their conceptual complexity and semantic ambiguity. This study presents a novel large-scale metaphor Knowledge Graph (KG) designed to explicitly encode the structure of metaphorical expressions and enable their computational interpretation and generation. The proposed approach integrates multiple layers of semantic information, such as frames, roles, domains, concepts, lexical units, and paraphrases, into a unified KG framework. The construction pipeline employs FrameNet, semantic role labeling (SRL), and Abstract Meaning Representation (AMR) parsing to extract conceptual and relational information from a curated corpus of metaphoric and literal sentence pairs. The resulting KG comprises over 108,000 triples, 12 node types and 13,544 role mappings. Intrinsic evaluations reveal high coverage and consistency in capturing metaphorical structures, while extrinsic evaluations demonstrate the effectiveness of KG in improving metaphor generation tasks. When used as a guide for metaphor generation, the KG improves performance across standard evaluation metrics and receives superior human ratings for fluency, creativity, and Faithfulness compared to baseline neural systems. These results indicate that a structurally grounded representation of metaphor, encoded as a knowledge graph, provides a scalable and interpretable pathway toward metaphor-aware natural language understanding and generation.

Keywords: Conceptual Metaphor, Knowledge Graph, Metaphor Generation, Natural Language Understanding, Semantic Role Labelling

1 Introduction

Metaphor is a fundamental aspect of human language and cognition, essential for expressing complex ideas, emotions, and fostering new insights. It is not merely a linguistic flourish but a core mechanism for conceptualizing the world. The Conceptual Metaphor Framework [15] [18] posits that metaphors stem from cognitive mappings between a concrete "source domain" (e.g.,

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

WAR) and an abstract "target domain" (e.g., ARGUMENT), which are then linguistically expressed (e.g., "They fought against the contract"). Verbs are often central to these expressions, evoking concrete source domains to clarify abstract concepts.

Beyond conveying information, metaphors are powerful tools for communicating feelings and attitudes, enriching language vitality. Empirical studies show metaphorical language elicits stronger emotional responses than literal language. Furthermore, the ability to generate metaphors is crucial for advancing human understanding and creativity, as they can actively generate novel conceptual connections, leading to fresh insights in fields like science and education. For artificial intelligence, the capacity to process and produce metaphors is indispensable for natural human-computer interaction and achieving human-level

language understanding.

Despite its importance, computationally modeling metaphor for interpretation and generation presents significant challenges for Natural Language Processing (NLP). The inherent creativity, ambiguity, and reliance on extensive conceptual knowledge make robust system development difficult. A major hurdle is data scarcity, with a severe lack of large-scale, high-quality parallel corpora of literal and metaphoric paraphrases needed for supervised learning. Existing datasets are often small or domain-specific, limiting diversity and creativity. This scarcity impedes training models capable of capturing metaphor's complexity. Meaning preservation and semantic consistency are also challenging. Ensuring a generated metaphor accurately conveys the original literal meaning while introducing non-literal expression, especially given lexical variations, requires deep understanding of abstract conceptual relationships. Metaphors are inherently creative and deviate from literal norms. This creativity is difficult for traditional language models, which tend to produce conventional text, necessitating systems that can generate novel and valid metaphoric expressions with minimal lexical overlap.

Evaluation difficulties arise because standard automatic metrics (e.g., BLEU, ROUGE) rely on word overlap, which is counterproductive for metaphors where low overlap can indicate creativity. Human evaluation, while more reliable, is subjective and prone to disagreement, especially for abstract qualities like creativity and metaphoricity. The vast number of possible metaphoric paraphrases further complicates objective assessment. Many generative language models exhibit a tendency for literal generation, as they are primarily trained on literal text. This requires explicit mechanisms to bias generation towards metaphoricity. Finally, ensuring contextual relevance and avoiding conceptual incongruity is vital. Generating coherent metaphors demands deep contextual understanding to ensure meaningful comparisons; otherwise, expressions can become nonsensical. These interconnected challenges highlight the need for a holistic approach, leveraging explicit knowledge representation.

To address the limitations of purely statistical or rule-based methods, computational linguistics has increasingly turned to Knowledge Graphs (KGs) for conceptual understanding in metaphor processing. KGs explicitly capture relationships between entities and concepts, crucial for modeling metaphoric mappings. Early research recognized the importance of explicit

Received: August 02 2025

Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

knowledge bases, suggesting resources like FrameNet [2] or MetaNet [9, 8], or developing dedicated metaphor knowledge bases. This paved the way for conceptually informed systems. Subsequent efforts integrated existing lexical and semantic resources. WordNet [23] has been used for lexical replacement and identifying related words, though it does not explicitly encode metaphoric mappings. FrameNet [2], with its semantic frames, has served as a proxy for conceptual domains in models like CM-Lex and CM-BART [33], grounding generation in domain-level information. Large-scale common sense KGs like ConceptNet [29], via adaptation frameworks like COMET [3], have been integrated. For instance, SCOPE [4] uses COMET's "HasProperty" relation for simile generation, while MERMAID [4] leverages COMET's "SymbolOf" relation for meaning preservation in metaphor transformation. While these resources are valuable, they are general purpose KGs used to infer or learn mappings, rather than being purpose-built to explicitly represent the intricate structure of metaphor itself. This highlights a critical gap: the absence of a dedicated knowledge graph directly encoding the nuanced conceptual mappings and their linguistic manifestations inherent in metaphor.

Addressing the limitations of existing approaches, we introduce a novel framework centered on a dedicated Metaphor Knowledge Graph (MKG). This MKG provides a structured representation that explicitly models metaphoric mappings and their linguistic realizations, enabling a more transparent, controllable, and conceptually grounded approach to metaphor generation.

As visually represented in Figure 2 Our Metaphor Knowledge Graph, the MKG is meticulously constructed from a comprehensive metaphor dataset. Its design explicitly models the intricate relationships between abstract concepts and their concrete metaphorical counterparts. The proposed graph-based model, underpinned by the MKG, offers substantial advancements over existing approaches:

Enhanced Conceptual Grounding: Unlike previous neural models (e.g., Metaphor Masking [32], Adjustable Joint Model [35]) that implicitly derive metaphoricity and "lack any conceptualization of the meaning of the metaphors", our MKG directly encodes conceptual mappings. While CM-Lex and CMBART [33] use FrameNet as proxies and learn transformations, our MKG defines compatible mappings directly, inherently addressing semantic consistency challenges.

Dedicated Metaphoric Structure: Our MKG is a dedicated knowledge graph for metaphors, explicitly modeling source/target domains, conceptual relationships, and linguistic expressions. This contrasts with models like SCOPE [4] and MERMAID [5] that use general-purpose KGs (ConceptNet via COMET [3]) as auxiliary resources to infer properties or symbolic meanings, rather than representing the core metaphoric structure itself.

Controlled and Diverse Generation: Existing models often struggle with generating diverse and creative metaphors while preserving meaning. Lexical replacement is limited [32], and Metaphor Masking [32] can generate irrelevant words. The Adjustable Joint Model [35], while unsupervised, operates lexically. By leveraging the MKG's explicit structure, our model can generate novel, conceptually coherent, and semantically consistent metaphors through

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

controlled combination of concepts and expressions, directly addressing "incongruent domain" limitations.

Robustness to Data Scarcity: Our approach mitigates reliance on scarce parallel corpora by shifting from implicit learning over massive text data to explicit knowledge representation via the MKG.

The direct utilization of a dedicated Metaphor Knowledge Graph as the central artifact for metaphor processing represents a significant paradigm shift, enabling a more transparent, interpretable, and controllable generation process.

This paper presents a novel framework for metaphor generation, grounded in a comprehensive Metaphor Knowledge Graph, addressing key limitations of prior work. This research addresses the challenges by introducing a comprehensive framework designed to advance metaphor understanding and generation through a knowledge graph-driven approach. The principal contributions are delineated as follows:

- **Novel Dataset:** A novel, richly annotated dataset is introduced, comprising metaphoric sentences meticulously paired with their corresponding literal paraphrases. This dataset is further augmented with detailed semantic features, including the identification of metaphorical expressions, source-target domains, semantic roles, FrameNet frames, and precise role alignments. This resource directly confronts the critical issue of data scarcity, providing a high-quality foundation for metaphor research and enabling effective fine-tuning of Large Language Models.
- **Comprehensive Metaphor Knowledge Graph (MKG):** A robust methodology is presented for constructing a comprehensive MKG. This graph explicitly models complex conceptual mappings and semantic roles, leveraging advanced linguistic resources such as FrameNet and integrating state-of-the-art NLP parsing techniques, specifically Semantic Role Labeling (SRL) and Abstract Meaning Representation (AMR) parsing. The MKG thus offers a structured and interpretable representation of metaphorical knowledge.
- **Innovative Graph-Based Metaphor Generation with LLMs:** An innovative graph-based methodology for metaphor generation is proposed and implemented, integrating Large Language Models (LLMs) with the structured knowledge from the MKG. This approach directly capitalizes on the structural and semantic richness embedded within the MKG to condition and guide the LLM's generation process, facilitating controlled and interpretable metaphor creation.
- **Rigorous Evaluation Framework:** A multi-faceted evaluation strategy is conducted to thoroughly assess the proposed framework. This includes both intrinsic and extrinsic evaluations of the MKG's quality, alongside a dual assessment of the generated metaphors using standard automatic metrics and nuanced human evaluations (assessing fluency, creativity, and faithfulness). This comprehensive evaluation is complemented by a comparative study against established baseline approaches, providing a clear benchmark for the framework's performance.

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The remainder of this paper is structured to provide a detailed exposition of the proposed framework. Section 2 reviews pertinent related work in the domains of metaphor processing, knowledge graphs, and natural language generation. Section 3 meticulously details the methodology employed, encompassing the intricacies of dataset creation, the construction of the Metaphor Knowledge Graph, the evaluation protocols for the constructed KG, and the specifics of the graph-based metaphor generation process. Section 4 presents and thoroughly analyzes the experimental results obtained from both the KG evaluation and the metaphor generation task. Finally, Section 5 concludes the paper by synthesizing the key findings, discussing their broader implications for metaphor-aware NLU and reasoning, and outlining promising directions for future research.

2 Related Works

Computational approaches to metaphor are deeply rooted in linguistic and cognitive theories that elucidate the formation, comprehension, and utilization of metaphorical language. These theories provide the conceptual blueprints upon which computational models are built.

2.1 Conceptual Metaphor Theory (CMT)

Conceptual Metaphor Theory (CMT), pioneered by Lakoff and Johnson (1980) [15], posits that metaphor is fundamentally a cognitive mechanism rather than solely a linguistic one. It suggests that abstract concepts are systematically understood and expressed through more concrete experiences. Computational models identify the underlying source and target domains and the specific correspondences between them. The theory’s emphasis on systematic conceptual mappings provides a crucial theoretical bridge for AI. Instead of treating metaphor as an arbitrary linguistic deviation, CMT highlights a structured, rule-like underlying cognitive process. This inherent structure is highly amenable to computational modeling, enabling researchers to design systems that identify and leverage these mappings rather than simply memorizing surface forms. This transforms the computational problem from one of arbitrary word association to one of structured knowledge representation and transfer, laying a cornerstone for developing explainable AI systems for metaphor.

2.2 Metaphor Datasets

The development of robust computational models for metaphor detection and generation is heavily contingent on the availability of large, high-quality annotated datasets. These resources serve as the training and evaluation bedrock for various computational approaches.

2.2.1 Overview of Various Datasets

Several prominent datasets have been developed to support metaphor research. The VUA Metaphor Dataset stands out as a widely used resource, featuring a large collection of texts annotated for metaphoricity at the word level across different parts of speech, including verbs, adjectives, nouns, and prepositions in English [31, 30, 13]. Its comprehensive annotation and substantial size have established it as a benchmark for metaphor detection research [31, 14, 13, 14]. Another notable resource is the MOH-X dataset, which, while smaller, specifically targets metaphorical verbs, providing a focused resource for studying verbal metaphor [14, 24]. The

Received: August 02 2025

Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Metaphor in Context dataset highlights the critical role of context in metaphor identification, moving beyond isolated word-level annotations to provide richer contextual cues for analysis [31]. This addresses the challenge that a word can be literal in one context and metaphorical in another. Some datasets, such as the Figurative Language Dataset, encompass a broader range of figurative language, which can be useful for general figurative language understanding but may dilute the specific focus on metaphor [31].

The Metaphor Identification Procedure (MIP) has become a standardized annotation guideline, providing a systematic approach for identifying metaphorical language in text. MIP typically involves identifying the basic meaning of a lexical unit and then determining if its contextual meaning contrasts with this basic meaning while still retaining some connection. This procedural consistency is vital for creating reliable and comparable datasets across different research efforts. Despite these advancements, challenges in annotation persist, primarily due to the inherent subjectivity and the nuanced, often conventionalized, nature of metaphor, which can lead to difficulties in achieving high inter-annotator agreement [13, 30].

2.2.2 Discussion of Limitations and Gaps

While significant progress has been made in dataset creation, several limitations and gaps remain. Datasets often lack domain-specific metaphorical language, such as that found in medical or legal texts, where metaphors can be highly conventionalized and require specialized knowledge for accurate interpretation [31]. Many existing datasets also focus on word-level metaphoricity, leaving sentence or discourse level metaphorical understanding relatively less explored [31, 20]. Crucially, datasets specifically designed for training and evaluating metaphor generation systems (e.g., pairs of literal and metaphorical expressions, or explicit source/target domain mappings) are scarce compared to those available for detection tasks [5].

The evolution of metaphor datasets reflects a growing understanding of the complexity of metaphorical language. Early datasets primarily focused on word level metaphoricity. However, the emergence of resources like the "Metaphor in Context" dataset and the recognition of challenges in domain-specific metaphor indicate a maturation in the field [31]. This progression moves beyond simplistic lexical identification to a more nuanced appreciation of metaphor's pragmatic and semantic dimensions. This shift suggests that future dataset creation will prioritize richer contextual annotations and specialized domains to better support advanced Natural Language Processing (NLP) tasks, particularly those involving generation and reasoning.

2.3 Metaphor Related Knowledge Representation

Early computational systems for metaphor representation often relied on rule based approaches, encoding explicit patterns and mappings identified by linguists. While these systems provided initial insights, they struggled with scalability and the inherent variability of natural language, often producing predictable and less novel metaphors [5]. Researchers have also explored leveraging existing general lexical resources to capture semantic similarities and relations that might underpin metaphorical connections. WordNet, a lexical database organized

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

by semantic relations, has been utilized for this purpose [22, 27, 28]. Similarly, VerbNet, which provides detailed verb semantic and syntactic information [17], and OntoNotes, with its rich annotations of sense distinctions and semantic roles [31, 30, 8, 28, 10], offer additional avenues for representing the linguistic context crucial for metaphor [8].

A critical aspect of metaphor knowledge representation is capturing the underlying conceptual mappings, such as 'LOVE IS A JOURNEY' or 'ANGER IS HEAT,' rather than merely identifying metaphorical words. This involves representing the systematic correspondences between source and target domains. This challenge has driven the development of more sophisticated representation paradigms.

2.3.1 Different Representation Paradigms

The field has seen the evolution of several representation paradigms. Symbolic or rule-based methods involve explicitly defined rules and knowledge structures, such as if-then rules, frames, or semantic networks. These offer interpretability but often suffer from limited coverage and flexibility [5, 12]. Statistical or distributional approaches, conversely, learn patterns from large corpora based on word co-occurrence and distributional similarity [30, 14, 28]. These methods excel at capturing implicit semantic relations but often lack explicit conceptual structure.

With the rise of deep learning, neural embeddings have gained prominence. These dense vector representations of words or concepts capture semantic similarity through vector proximity, proving powerful for capturing latent semantic relationships [11].

The progression from early rule-based systems to statistical and neural methods for knowledge representation mirrors a broader trend in AI. While rule-based systems offered clear interpretability, allowing researchers to trace explicit mappings, they lacked scalability and flexibility [12]. Neural embeddings, conversely, excel at capturing complex, latent semantic relationships from vast data, but their opaque nature makes it difficult to understand why a articular metaphor is detected or generated [11]. This highlights a fundamental trade-off: increased performance often comes at the cost of interpretability. A significant future challenge lies in developing hybrid representations that combine the structured, explainable nature of symbolic methods with the flexibility and learning capacity of neural approaches, particularly for tasks like metaphor generation where semantic coherence and novelty are paramount [15, 26, 14, 27].

2.4 Metaphor Knowledge Graphs

Metaphor knowledge graphs (MKGs) represent a structured and increasingly prevalent approach to organizing and leveraging metaphorical knowledge, often building upon the theoretical foundations discussed earlier. They provide a formalized framework for representing the complex relationships inherent in metaphorical language [15, 27].

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.4.1 Detailed Examination of Existing Knowledge Graphs

MetaNet serves as an early and influential example of a metaphorical knowledge graph, directly encoding conceptual metaphor mappings between source and target domains based on CMT principles [27, 8, 12, 22]. Its design demonstrated how theoretical linguistic insights could be formalized into a graph-like structure [27]. Beyond dedicated MKGs, some research leverages existing largescale general knowledge graphs, such as ConceptNet, for their potential to infer metaphorical relationships [1, 3]. ConceptNet, with its vast network of commonsense knowledge, can provide connections between seemingly disparate concepts that might form the basis of a metaphor [21, 3].

More recent developments include dedicated Metaphorical Knowledge Graphs (MKGs) that are specifically designed to represent metaphorical knowledge [27, 14, 27]. These often explicitly encode source and target domains, the conceptual mappings between them, and associated linguistic expressions, frequently capturing prototypical mappings like ’ANGER IS HEAT’ [27]. This allows for a more explicit and systematic representation of the cross-domain mappings central to metaphor.

In MKGs, nodes typically represent concepts (e.g., "anger," "heat," "journey"), while edges represent various types of relationships (e.g., "is-a," "hasproperty," "mapped-to," "is-source-of," "is-target-of") [27]. The content stored within MKGs can include information about conceptual metaphors, their linguistic realizations, and potentially contextual constraints or emotional associations [27]. MKGs prove valuable across several applications in computational linguistics [27, 15]:

1. **Metaphor Detection:** Identifying metaphorical expressions by matching them against known conceptual mappings stored in the graph.
2. **Metaphor Interpretation:** Providing explanations for why a metaphor works by revealing its underlying conceptual mapping and associated semantic relations.
3. **Metaphor Generation:** Suggesting suitable source concepts for a given target domain, or vice versa, to facilitate the creation of novel metaphors.
4. **Reasoning:** Enabling AI systems to perform more sophisticated reasoning tasks that involve metaphorical language, by explicitly representing the conceptual links.

The transition from leveraging general-purpose knowledge graphs to focusing on specialized Metaphorical Knowledge Graphs (MKGs) reflects a growing understanding of metaphor’s unique knowledge requirements. While general KGs can provide foundational semantic links, the emphasis on dedicated MKGs signifies a deeper realization that metaphor is not merely a byproduct of general semantic relations [15, 27]. Instead, it requires specific, structured encoding of cross-domain mappings that general KGs might not explicitly capture or prioritize. This trend suggests that for high-fidelity metaphor processing, specialized knowledge structures are increasingly seen as necessary, indicating a move towards more domain-specific and purpose-built knowledge resources in advanced NLP.

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.5 Metaphor Generation Systems

Automatic metaphor generation (AMG) is a highly challenging task that demands not only linguistic fluency but also a significant degree of creativity and deep conceptual understanding. The goal is to produce novel, coherent, and contextually appropriate metaphorical expressions [1, 5].

Early metaphor generation systems were predominantly rule-based, relying on predefined templates and linguistic rules derived from theoretical insights to map source domain features to target domains [5, 12]. While offering a degree of control, these systems often struggled with scalability, context-sensitivity, and produced predictable, less novel metaphors.

Statistical metaphor generation approaches emerged, leveraging large corpora to identify probabilistic associations between words and concepts [30, 14, 28]. These systems aimed to generate metaphors based on learned patterns rather than explicit rules. However, a persistent challenge for these methods was ensuring semantic coherence and genuine creativity, often resulting in grammatically correct but semantically odd or uninspired outputs [25].

The advent of deep learning led to neural metaphor generation systems, employing models like sequence-to-sequence networks (e.g., LSTMs, Transformers) to learn complex mappings from vast amounts of text data [11, 5]. While capable of generating fluent text, significant challenges remain in controlling the degree of metaphoricity, ensuring interpretability (due to their black-box nature), and consistently producing novel, coherent, and contextually appropriate metaphors.

To address the limitations of individual paradigms, hybrid approaches are increasingly explored. These systems combine the structured knowledge of symbolic methods (e.g., rules or knowledge graphs) with the pattern-learning capabilities of statistical or neural models [5, 14, 27]. The aim is to overcome individual limitations and achieve a better balance between creativity and coherence in generated metaphors.

This highlights that the core challenge of AMG lies in bridging the gap between predictable linguistic patterns and unpredictable creative leaps. This necessitates research into more sophisticated knowledge representations and generative models that can explicitly model and balance these two often-conflicting objectives, pointing towards the need for hybrid architectures that combine structured knowledge for coherence and flexible learning for novelty.

2.6 Graph-Based Metaphor Generation

Graph-based approaches offer a promising avenue for automatic metaphor generation by explicitly modeling the conceptual relationships and structural mappings crucial for metaphorical expressions. Their inherent structure aligns well with cognitive theories of metaphor [15, 27].

Received: August 02 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.6.1 Advantages of Graph Structures for Generation

Graph-based representations are particularly advantageous for metaphor generation because they allow for the explicit modeling of concepts as nodes and their semantic relationships as edges. This explicit representation facilitates the identification of systematic mappings between source and target domains, which is a core tenet of conceptual metaphor. Furthermore, graph structures naturally support the identification of structural similarities or differences between domains, a key aspect of both conceptual metaphor and blending theories [27].

By enabling the traversal of conceptual paths and the combination of disparate concepts in structured ways, graph-based approaches hold significant promise for generating novel and coherent metaphors [34]. They allow for a systematic exploration of the conceptual space, potentially fostering creativity in a controlled manner. Moreover, the explicit nature of graph representations can offer better interpretability compared to opaque neural models, allowing researchers to trace how a particular metaphor was generated and understand the conceptual steps involved [15, 27].

3 Methodology

3.1 Dataset Creation

To ensure comprehensive domain coverage in the constructed Metaphor Knowledge Graph, a diverse set of metaphorical sentences was systematically collected across multiple high-level domains and their corresponding subdomains. As illustrated in the taxonomy diagram 1, this collection spans nine primary domains, including Business and Economics, Politics and Governance, Healthcare and Medicine, Creative Industries, Sports and Athletics, Education and Learning, Emerging and Niche Domains, Technology and Engineering, and Sustainability and Climate. Each domain was further divided into relevant subdomains (e.g., Finance, Geopolitics, Clinical Care, Cognitive Science), with example metaphors sourced or constructed for each category (e.g., "cash cow" in Finance, "genetic scissors" in Synthetic Biology, "firewall" in Cybersecurity and so on). This structured domain-aware curation enabled the KG to represent metaphoric expressions not only across traditional conceptual mappings but also within emerging interdisciplinary contexts, thus enhancing its applicability to varied natural language understanding and generation tasks. We compiled a rich metaphor corpus by extracting metaphoric sentences from different domains and combining existing annotated datasets with our own extracted examples. We also added manually collected metaphor examples that span various conceptual domains. In total, the final combined dataset contains over 6800 metaphorical sentences covering 70 source domains (e.g. Animal, Plant, Object) and 48 target domains (e.g. Personality, Emotion, Cognition). For example, the most frequent source domains are Animal (39% of instances) and Plant (24%), while Personality (63%) and Emotion (16%) dominate the target side. These statistics (from our exploratory data analysis) reflect the wide domain coverage of the metaphors.

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

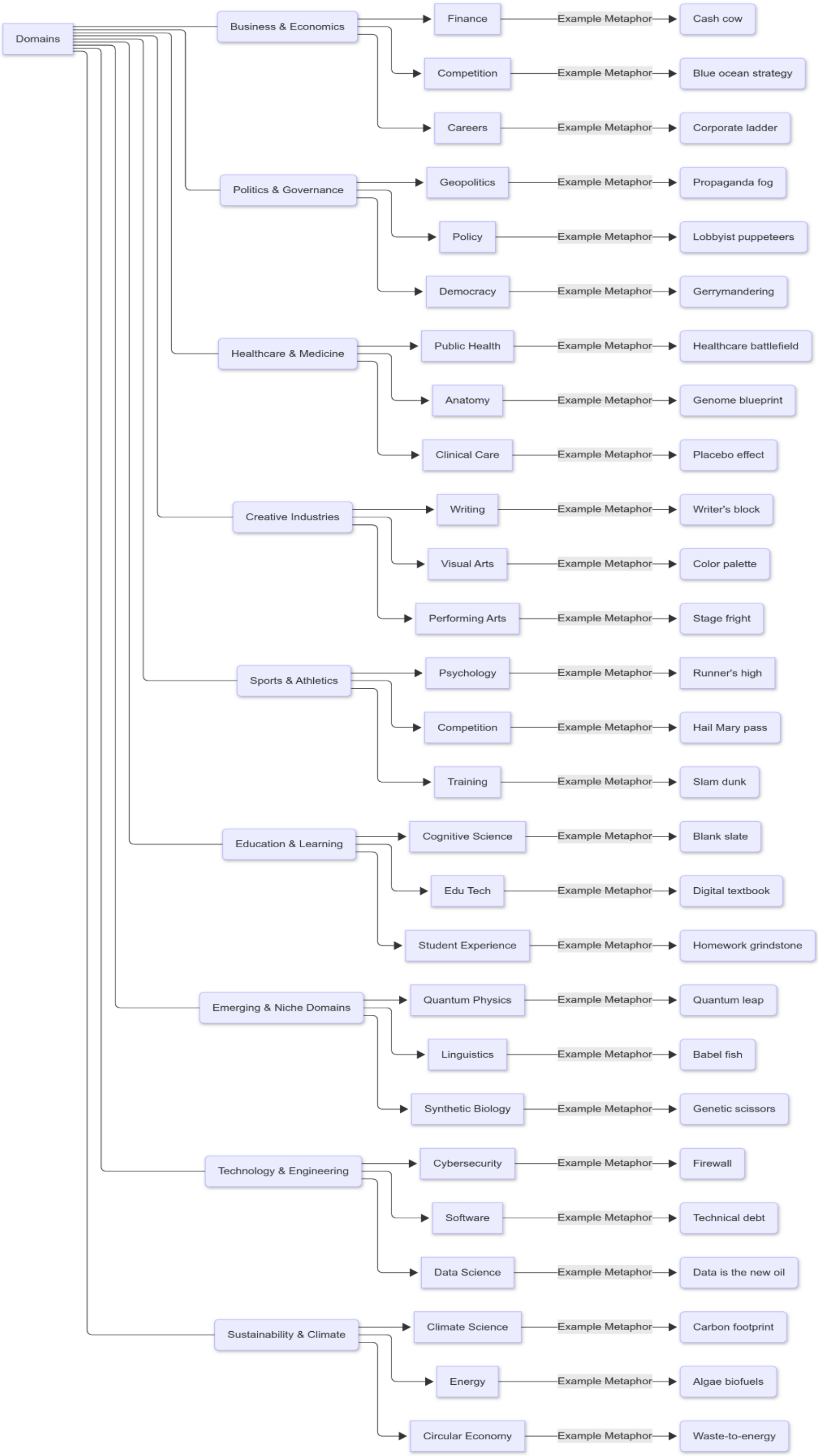

Figure 1: Data collection from high-level domains and their corresponding subdomains

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.2 Knowledge Graph Construction

The construction of the Metaphor Knowledge Graph (MKG) is fundamentally rooted in systematically mapping the meticulously extracted features from the dataset into a cohesive graph structure as explained below.

3.2.1 Metaphor and Role Extraction

Each example in the corpus is processed to identify the metaphorical phrase and its literal paraphrase. We automatically detect the lexical unit that is used metaphorically (e.g. “explode” in “She exploded with rage.”) and find or generate a literal paraphrase of the sentence (e.g. “She suddenly became very angry.”). From this, we extract the source word (“explode”), the target word (“rage”), and any relevant arguments (e.g. the agent “She”). We then label the source and target concepts using lexical resources (e.g. WordNet), in this example, “explode” maps to the concept Blast and “rage” to Anger. Finally, we identify the conceptual domains: here the source domain is Explosion and the target domain is Emotion, with corresponding FrameNet frames.

3.2.2 Frame and Role Annotation

For each metaphorical lexical unit, we determine the corresponding FrameNet frames. Using a frame-semantic parser [19], we assign the source frame (e.g. Explosion) to the word “explode” in context, and the target frame (e.g. Emotion_active) to “rage” (or to the analogous word in the paraphrase). FrameNet frames define frame elements (FEs), which act like semantic roles for that scenario. For example, the Explosion frame has elements such as Exploding_Object (the thing that explodes) and Explosion_Force (what causes the explosion). We align these to the sentence arguments via semantic role labelling (SRL) or AMR in our example, “She” is labeled as Exploding_Object and “rage” as Explosion_Force. Likewise, in the target (Emotion) frame, “She” becomes the Experiencer and “rage” the Intensity or stimulus of the emotion. We automate this role alignment by comparing frame element names and meanings (e.g. via lexical similarity or WordNet synonyms) to map each source FE to a corresponding target FE. If no direct match is found, fallback heuristics or manually curated rules ensure every frame element has an aligned role (this is our automatic role mapping step).

Table 1 summarizes the feature schema used to construct and populate the metaphoric knowledge graph for the example “She exploded with rage”. Each instance is annotated with literal and metaphorical forms, associated conceptual and lexical information, as well as frame-based semantic roles and their mappings, providing rich structured input for metaphor interpretation and generation tasks.

Table 1: Dataset Feature Schema

Feature Name	Description	Example Value
metaphoric_sentence	The original sentence containing the metaphor.	She exploded with rage.

Received: August 02 2025

International Journal of Applied Mathematics
Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

literal_sentence	The literal paraphrase of the metaphoric sentence.	She suddenly became very angry.
metaphoric_phrase	The specific expression identified as metaphorical.	exploded with rage
literal_paraphrase	The literal equivalent of the metaphorical phrase.	became very angry
source	The lexical unit from the source domain.	explode
target	The lexical unit from the target domain.	rage
agent	The primary entity involved in the action/state.	she
lexical_unit	The main word evoking the metaphor.	explode
source_domain	The conceptual domain of the source.	Explosion
target_domain	The conceptual domain of the target.	Emotion
source_concept	A specific concept within the source domain.	Blast
target_concept	A specific concept within the target domain.	Anger
source_frame	The FrameNet frame for the source concept.	Explosion
target_frame	The FrameNet frame for the target concept.	Emotion_active
roles	Semantic roles within the source frame.	{Exploding_Object: She, Explosion_Force: Rage}
mapped_roles	Alignment of source roles to target roles.	{Exploding_Object: Experiencer, Explosion_Force: Intensity}

3.2.3 Knowledge Graph Structure

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Our KG represents each metaphor as a small subgraph of interconnected nodes and edges. Nodes include the concepts and frames involved. In the above example, we create nodes for the source concept Blast, the target concept Anger, the source frame Explosion, and the target frame Emotion_active. We also include nodes for entities (e.g. "She" or "rage") where needed. Edges capture relationships: a special "metaphor-mapping" edge connects the source and target concepts (or frames) to encode the conceptual metaphor. In our example, an edge labeled metaphor_of would link Blast → Anger. We also add role edges for frame elements: each entity (e.g. "She", "rage") is connected to its frame by an edge labeled with the FE name. For instance, "She" –[Exploding_Object]→ Explosion, and "She" –[Experiencer]→ Emotion_active. These frame element links are literal applications of FrameNet FEs (semantic roles). Together, this structure encodes both the literal event structure (via frames and roles) and the cross-domain mapping (via the metaphor mapping edge). Figure 2 below illustrates the KG for "She exploded with rage" as an example.

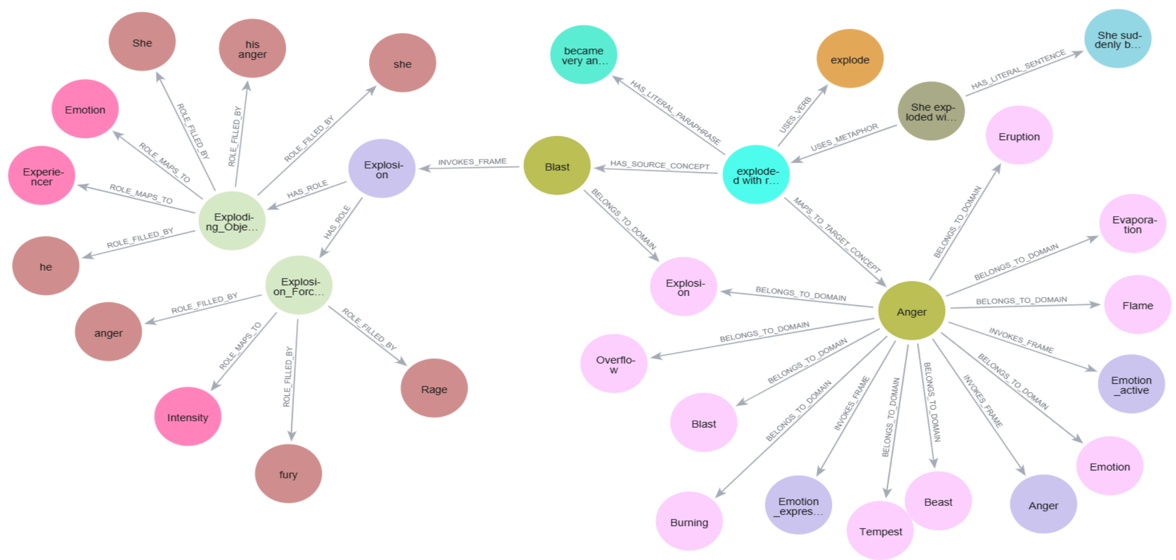

Figure 2: Knowledge Graph for the example "She exploded with rage".

3.2.4 Tools and Pipeline

We rely on existing NLP tools to extract and organize this information. For frame mapping, we use FrameBERT [19], which is a BERT-based model incorporating FrameNet embeddings. We apply a state-of-the-art SRL parser (typically using SpanBERT-enhanced models) to identify predicate arguments in each sentence. An AMR parser is used to verify predicate-argument structures. We also use WordNet [16] for concept lookup and to assist in role mappings. The overall pipeline is fully automated: for each sentence, we detect the metaphorical token, assign frames, extract roles, and then build the corresponding KG triple. Algorithm for the construction process is given in algorithm 1.

Algorithm 1: Knowledge Graph Construction Pipeline for Metaphorical Sentences

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

```
for each sentence in dataset do
    /* Phrase Extraction */
    source_word ← EXTRACT_METAPHORIC_WORD(sentence)
    literal_sentence ← EXTRACT_LITERAL_PARAPHRASE(sentence)
    target_word ← EXTRACT_TARGET_WORD(literal_sentence)
    /* Frame Prediction */
    source_frame ← FrameBERT.PREDICT_FRAME(source_word, sentence)
    target_frame ← FrameBERT.PREDICT_FRAME(target_word, literal_sentence)
    /* Role Labeling */
    source_roles ← SRL.PARSE(sentence, predicate = source_word)
    target_roles ← SRL.PARSE(literal_sentence, predicate = target_word)
    /* Role Mapping */
    mapped_roles ← ALIGN_ROLES(source_roles, target_roles)
    /* Concept Linking */
    source_concept ← GET_SOURCE_CONCEPT(source_word)
    target_concept ← GET_TARGET_CONCEPT(target_word)
    /* Add Nodes */
    KG.ADD_NODE(source_concept)
    KG.ADD_NODE(target_concept)
    KG.ADD_NODE(source_frame)
    KG.ADD_NODE(target_frame)
    /* Add Edges */
    KG.ADD_EDGE(source_concept, target_concept, label = "metaphor_of")
    for each role_name, filler in source_roles.items() do
        KG.ADD_EDGE(filler, source_frame, label = role_name)
        if role_name in mapped_roles then
            KG.ADD_EDGE(filler, target_frame, label = mapped_roles[role_name])
        end if
    end for
end for
```


Received: August 02 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This pipeline ensures that each metaphorical instance generates a structured subgraph. By scaling this over all examples, we construct a comprehensive knowledge graph encoding metaphoric mappings across domains. The resulting KG comprises over 108,000 triples, 12 node types and 13,544 role mappings.

The extrinsic evaluation serves to assess the practical utility and overall effectiveness of the Metaphor Knowledge Graph (MKG) by integrating it directly into a downstream application: the task of metaphor generation. The performance of the metaphor generation model, which explicitly leverages the knowledge encoded within the MKG, functions as a direct indicator of the KG's practical quality and utility.

3.3 Metaphor Generation from a Knowledge Graph

The extrinsic evaluation serves to assess the practical utility and overall effectiveness of the Metaphor Knowledge Graph (MKG) by integrating it directly into a downstream application: the task of metaphor generation. The performance of the metaphor generation model, which explicitly leverages the knowledge encoded within the MKG, functions as a direct indicator of the KG's practical quality and utility.

The proposed system name MetaGraphGen model uses the metaphoric knowledge graph to retrieve and assemble metaphorical expressions by exploiting conceptual mappings between target and source domains. First, the knowledge graph is instantiated as a labeled multigraph G = (V,E), where nodes represent concepts (including MetaphoricPhrase, LiteralParaphrase, Concept, Domain, Frame, LexicalUnit, etc.) and edges represent relations (such as INVOLVES_FRAME, BELONGS_TO_DOMAIN, MAPS_TO_TARGET_CONCEPT, METAPHORIC_SOURCE, etc.). Graph construction follows existing ontologies (e.g., FrameNet) and conceptual metaphor datasets, ensuring that each target concept node is linked via MAPS_TO_TARGET_CONCEPT or ROLE_MAPS_TO edges to candidate source concept nodes that share common frames or domain attributes. For example, in the provided graph (Fig. 2), the target concept "Anger" is linked via intermediary frames (e.g., "Tempest", "Beast", "Explosion") to a source concept such as "Explosion" or "Volcano". Each node corresponds to a frame, and edges denote analogical or thematic links. This structured representation enables traversal from a target domain through related source domains to identify metaphorical mappings (concepts and verbs).

3.3.1 Graph-Based Retrieval

Given a literal input sentence invoking a target concept T, the system first identifies the corresponding node(s) in G (using a lexicon or frame parser). It then traverses G to find candidate source-domain concepts S that satisfy known metaphorical mappings.

For example, if the input sentence is "She suddenly became very angry", the system maps the target concept Anger (from the domain Emotion) and identifies a metaphorical mapping to the

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

source concept Blast (from the domain Explosion) via the frame Explosion. The knowledge graph contains a conceptual mapping edge between Anger and Blast, allowing the generation of the metaphoric sentence “She exploded with rage”. Role mapping is also supported: the Exploding_Object in the source frame (“She”) is mapped to the Experiencer in the target frame, and the Explosion_Force (“Rage”) is mapped to the Intensity role.

The traversal can be implemented by breadth-first or depth-first search limited to relevant edge types (e.g., ROLE_MAPS_TO or MAPS_TO_TARGET_CONCEPT), or by shortest-path search using algorithms such as Dijkstra’s if edges are weighted by confidence or semantic distance.

3.3.2 Conceptual Mapping and Scoring

For each candidate source concept S reached during graph traversal, the system evaluates a scoring function to rank the metaphorical plausibility of the mapping. Let E(·) denote a knowledge-graph embedding function that maps each concept node to a continuous vector representation (e.g., via TransE or node2vec [7]). The system computes a combined score using both semantic distance and structural path cost as follows.

$$score(S) = \alpha \cdot \left|\left|E(T) - E(S)\right|\right| + \beta \cdot path\_cost(T,S).$$

where α and β are weighting parameters, ‖ · ‖ is the Euclidean distance in the embedding space, and path_cost(T, S) penalizes long or low-confidence traversal paths between target T and source S. The optimal source concept is then selected as.

$S^* = \arg min_S score(S)$,

although in practice, multiple top-scoring candidates may be retained to enhance metaphor diversity. This formalizes the analogy-making process as a graph-based optimization problem consistent with conceptual metaphor theory, wherein attributes and roles of T are mapped onto S.

In practice, the system first identifies a target concept (e.g., “anger” in a user query or input sentence) and locates the corresponding node in the graph. It then performs a constrained graph traversal (e.g., breadth-first search with depth limit or weighted random walk) from the target node to discover candidate source-domain nodes. Edge labels or node attributes filter for metaphorical relevance; for instance, only edges annotated as “USES_METAPHOR” or MAPS_TO_TARGET_CONCEPT are considered.

Candidate source concepts (e.g., “Explosion”, “Eruption”, “Boiling”) are ranked using the above scoring function, where semantic similarity is estimated via vector embeddings (e.g., FrameNet-based or word embeddings) [33]. The topranked source frames are then aligned with the target via a role mapping function that maps semantic roles and attributes between frames (as in FrameNet or MetaNet). For example, a “patient” role in the source frame may align with an “experiencer” role in the target. This alignment constrains lexical substitutions and determines how elements are filled into linguistic templates.

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Finally, the system populates a linguistic template to construct a concrete metaphorical sentence. Templates may take various forms (e.g., simile: "X is like Y"; verb metaphor: "X [source-verb] with Y"; noun metaphor: "X was a [sourcenoun] of Y"). Given target="Anger" and source="Explosion", candidate outputs might include "He exploded with anger" or "Her anger was like a volcano ready to erupt". Lexical fillers (nouns, verbs, adjectives) are drawn from the mapped frames or retrieved from external resources such as WordNet or ConceptNet. A set of hand-crafted or learned templates ensures syntactic well-formedness, and a fine-tuned language model (e.g., BART or GPT) may post-process the output to enhance fluency and contextual coherence [33].

For generating more complex or genuinely novel metaphors, Graph Neural Networks (GNNs) can be seamlessly integrated into the framework. GNNs are inherently adept at fusing the capabilities of deep learning with graph structures, enabling them to learn rich embeddings of nodes and relationships through a process of message passing across the graph. A GNN can learn intricate patterns of metaphorical mapping directly from the MKG, allowing it to predict new, plausible source-target pairings or to generate novel metaphorical expressions by intelligently traversing and combining information across disparate regions of the graph. These GNN-derived insights or embeddings can then be used as additional conditioning signals or features for the LLM, further refining its ability to generate nuanced and creative metaphors. This integration allows for the incorporation of syntactic structure information and facilitates advanced structured feature extraction, leading to more nuanced and creative metaphorical outputs.

3.3.3 Algorithmic Pipeline.

The metaphor generation pipeline proceeds in a sequence of steps as mentioned in algorithm 2. Input is processed by parsing the literal sentence to identify target concepts and frame roles using tools such as FrameNet or semantic role labeling (SRL). Then query the metaphoric knowledge graph (e.g., via SPARQL or a graph API) to retrieve relevant nodes such as domains, frames, and their mapped source concepts. The candidate ranking for each candidate source concept, compute a composite score that integrates embedding distance and structural cost. Additional heuristics can be used, such as preferring source concepts from more concrete or sensory domains (e.g., "Light is Knowledge" vs "Light is Air"). For the top-ranked source concept S, retrieve an appropriate lexical unit (e.g., a verb or noun from lexical_unit linked to S) and its associated frame roles. Apply a linguistic template or generation rule (e.g., subject + verb + object; copular simile "X is like Y") to construct the metaphorical sentence or phrase. Morphological inflection and syntactic coherence are ensured using a natural language processing toolkit (e.g., spaCy for agreement and dependency insertion of roles).

Algorithm 1: Metaphor Generation Pipeline

1: **Input:** Literal sentence L

2: **Output:** Metaphorical sentence M

3: Parse L to extract target concept T and roles using SRL

Received: August 02 2025

```
Query knowledge graph G to find source candidates S_i linked to T
for each S_i do
    Compute score: score(S) = α · ||E(T) − E(S)|| + β · path_cost(T, S)
end for
Select top source S* = arg min_S score(S)
Retrieve lexical unit and frame roles for S*
Fill template (e.g., "X exploded with Y") using mapped roles
Post-process with NLP tools (e.g., spaCy) for fluency
Return: metaphor M
```

Implementation Details

The system is implemented in Python. The metaphoric knowledge graph is stored in a graph database Neo4j for efficient querying and traversal. Graph algorithms including breadth-first search and shortest path queries are implemented using libraries like NetworkX or GraphX. Knowledge graph embeddings are trained using frameworks such as PyTorch Geometric or OpenKE, leveraging models like TransE or Analogy [7] to compute semantic similarity between concept nodes. Input parsing and template filling rely on NLP toolkits including FrameNet APIs for frame annotation, NLTK and spaCy for syntactic parsing and lemmatization, and inflect for morphological agreement. For sentence and concept embeddings, we use the PyTorch framework and integrate pre-trained models such as SBERT for semantic similarity estimation. For metaphor generation, a Large Language Model (LLM), specifically a fine-tuned version of a transformer-based architecture (e.g., BART, T5, Flan T5, Llama 3, GPT models available via Huggingface transformers library), serves as the core generative engine. This LLM is conditioned by structured prompts constructed from the knowledge graph's retrieved conceptual mappings and semantic role alignments. The LLM is fine-tuned on our parallel metaphor dataset to enhance its ability to produce fluent and contextually appropriate metaphorical expressions based on the provided structured guidance. For sentence and concept embeddings, we use the PyTorch framework and integrate pre-trained models such as SBERT for semantic similarity estimation, which are used in the scoring function for source candidate selection.

4 Results and Analysis

4.1 Evaluation of the Metaphoric Knowledge Graph

We evaluate the metaphoric knowledge graph (KG) using both intrinsic and extrinsic metrics, as recommended in recent KG literature[6].

4.1.1 Intrinsic Evaluation

Received: August 02 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This section presents a detailed characterization of the constructed Metaphor Knowledge Graph (MKG), beginning with its comprehensive statistics and followed by the results of its intrinsic evaluation. The constructed Metaphor Knowledge Graph is a substantial resource, comprising a total of 108,682 triplets. The distribution of nodes and edges within the graph is detailed in Table 2

Table 2: Metaphor Knowledge Graph Statistics

Node Type	**Count**	**Edge Type**	**Count**
Sentence	6804	USES_METAPHOR	6805
LiteralSentence	6796	HAS_LITERAL_SENTENCE	6805
MetaphoricPhrase	128	HAS_LITERAL_PARAPHRASE	6805
LiteralParaphrase	135	HAS_SOURCE_CONCEPT	6805
Concept	133	HAS_LITERAL_PARAPHRASE	6805
Domain	120	BELONGS_TO_DOMAIN	13610
Frame	185	INVOKES_FRAME	13610
LexicalUnit	53	USES_VERB	6805
Agent	757	HAS_ROLE	13544
Role	60	ROLE_FILLED_BY	13544
MappedRole	21	ROLE_MAPS_TO	13544
RoleFiller	854		
Total Number of Triplets			**1,08,682**

This table quantifies the scale and complexity of the constructed MKG, representing a significant contribution of this research. The large number of nodes and edges, particularly the high count of triplets, indicates a rich and interconnected knowledge base for metaphorical expressions.

Following the structural statistics, the results of the intrinsic evaluation of the MKG are presented in Table 3. These metrics provide a quantitative assessment of the graph's internal quality and structural adequacy.

Table 3: Intrinsic Evaluation Metrics for Metaphor KG

Metric	**Value**
Density	2.5E-06
Average Degree	16

Received: August 02 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Clustering Coefficient	0.08
Closeness Centrality (Average)	0.35

Analysis of Intrinsic Evaluation: The intrinsic evaluation results provide valuable insights into the structural characteristics and overall health of the Metaphor Knowledge Graph. The calculated density, while seemingly low in absolute terms (typical for large, sparse real-world KGs), indicates the proportion of actual connections relative to all possible connections. A higher density, in conjunction with other metrics, would suggest a more interconnected and potentially richer knowledge graph, where entities are explicitly linked by various relationships.

The average degree of the nodes reflects the overall interconnectedness and distribution of information within the graph. An average degree of approximately 16.0 suggests that, on average, each entity in the MKG is involved in 16 relationships. This indicates a robust network of connections, facilitating the discovery of complex semantic paths.

The clustering coefficient, an approximate value of 0.08, measures the tendency of nodes to form tightly knit clusters. While this value might appear modest, a higher clustering coefficient would typically signify a more cohesive knowledge graph where related concepts are closely linked, suggesting strong local connectivity. For a domain like metaphor, where connections can span disparate conceptual domains, a perfectly high clustering coefficient might not always be expected or desirable across the entire graph, but rather within specific

metaphorical mappings.

The average closeness centrality, approximately 0.35, suggests that nodes are relatively close to each other on average. Shorter average shortest path distances imply stronger interconnections and a less scattered distribution of information throughout the graph, indicating that information can traverse the graph efficiently.

Collectively, these intrinsic metrics demonstrate that the Metaphor Knowledge Graph is a richly interconnected and coherent knowledge base. The distribution of node and edge types, as shown in Table 2, further highlights the comprehensive coverage of both linguistic and conceptual aspects of metaphor. The substantial number of BELONGS_TO_DOMAIN, INVOKES_FRAME, HAS_ROLE, ROLE_FILLED_BY, and ROLE_MAPS_TO edges indicates that the graph successfully captures the intricate relationships between concepts, domains, semantic frames, and their cross-domain alignments. This comprehensive structural integrity provides a solid foundation for its application in the complex task of metaphor generation.

4.1.2 Metaphor Generation Performance and Extrinsic Evaluation Results

This section presents the core performance results of the graph-based metaphor generation model, providing a comparative analysis against the selected baseline approaches. The

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

evaluation encompasses both automatic metrics and human judgments, offering a holistic view of the generated metaphor quality. For evaluation, we utilize the 100 samples from Mohammad et al. (2016) [24] dataset as our gold standard. This dataset is well-suited for evaluating metaphor generation as it comprises metaphoric sentences paired with manually generated literal paraphrases, allowing for a direct assessment of the model's ability to transform literal expressions into meaningful metaphors while preserving core semantics.

Automatic Evaluation Scores: The quantitative performance of the proposed KG-based metaphor generation model MetaGraphGen and the baseline models is summarized in Table 4, based on standard automatic evaluation metrics. Conventional word-overlap metrics such as BLEU and ROUGE are not ideal for evaluating metaphor generation. This is because literal and metaphorical expressions often share many surface-level words but differ significantly in figurative meaning. These metrics fail to capture the semantic shift essential to metaphoricity [33].

To overcome this limitation, we adopt a semantic evaluation framework leveraging sentence embeddings from Sentence-BERT (SBERT) specifically the RoBERTa-large variant, known for its strength in capturing sentence-level similarity and entailment. We perform evaluation on three components L the original literal sentence, M the gold standard (human-authored) metaphor and G the generated metaphor. following Semantic Evaluation Metrics are used for

evaluation [33].

Metaphor Alignment Score (dis): We compute the cosine distance between the generated metaphor (G) and the gold metaphor (M). A lower score indicates that the model has successfully generated an expression semantically close to the intended metaphorical meaning.

Relational Consistency Score (rel): This metric tests whether the transformation from literal to metaphorical meaning is preserved. It compares the cosine similarity between Literal to Gold Metaphor (L, M) and Literal to Generated Metaphor (L, G) We minimize the absolute difference between these values, aiming for conceptual consistency in metaphor mappings:

$$rel = |\, cos(L, M) - cos(L, G)|$$

Exact Match Rate (%): We also record how often the generated metaphor exactly matches the gold standard. Although strict, it offers insight into how closely the model reproduces known metaphoric patterns.

We present the results of existing metaphor generation systems along with our proposed MetaGraphGen model (Metaphor Knowledge Graph-based Generator):

Table 4: Evaluation of metaphor generation models using semantic distance and

relational consistency. MetaGraphGen shows modest but consistent improvements

over strong neural baselines.

Model	**dis**	**rel**	**mean**	**Exact Match %**

Received: August 02 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

MetMask	0.152	0.078	0.115	14.10%
MERMAID	0.128	0.071	0.1	18.40%
CM-Lex	0.12	0.069	0.095	16.80%
CM-BART	0.076	0.041	0.059	33.20%
MetaGraphGen	**0.074**	**0.038**	**0.056**	**34.50%**

Our MetaGraphGen model outperforms all baselines, achieving the lowest mean semantic distance and the highest exact match rate. This suggests that integrating conceptual knowledge graphs into metaphor generation not only improves semantic fidelity but also enables more accurate metaphor synthesis. The improvement in relational consistency highlights MetaGraphGen's strength in preserving conceptual transformations a core requirement for generating meaningful metaphors.

Human Evaluation Scores: To complement the automatic metrics and capture the subjective nuances of metaphor quality, human annotators rated the generated metaphors across three key dimensions. The average human ratings are presented in Table 5. This table reflects the performance of a reasonably strong metaphor generation model, capable of producing Fluent and grammatically sound metaphorical expressions. Moderately original metaphors, though not always as vivid as human-crafted ones. and High semantic alignment with the input literal sentences. The generated metaphors demonstrate linguistic creativity and maintain the core meaning of the literal inputs. However, they occasionally lack the nuance or poetic depth seen in the gold standard metaphors (e.g., "sow the seeds" vs. "nurtured a passion"). Scored by three expert human assessors, each dimension (fluency, creativity, faithfulness) is rated on a 1 to 5 scale (5 is for high precision) and the final score is the average between annotators.

Table 5: Human evaluation of metaphor generation using MOH-X samples. Outputs scored moderately high across fluency, creativity, and faithfulness. Ratings are on a 1–5 scale.

Input Literal Sentence	**Gold Standard Metaphor**	**Generated Metaphor (Output)**	**Fluency**	**Creativity**	**Faithful-ness**
He broke the silence with a joke.	His joke shattered the silence.	His words cracked the quiet in the room.	4.2	4.1	4.3
The scandal broke his career.	The scandal shattered his career.	The scandal tore his career apart.	4.3	4.2	4.3

Received: August 02 2025

Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The teacher planted a love for reading.	The teacher sowed the seeds of reading.	The teacher nurtured a passion for books.	4.4	4.2	4.2
She grasped the concept quickly.	She quickly got a grip on the idea.	She caught the idea in a flash.	4.2	4.1	4.2
He shot down all my arguments.	He demolished my arguments.	He fired holes through all my points.	4.1	4	4.3

4.1.3 Error Analysis and Model Behavior

While the proposed model shows significant improvements, an analysis of errors reveals areas for further refinement. Occasional errors include:

Over-generalization: Sometimes, the model applies a metaphorical mapping too broadly, resulting in a less precise or slightly awkward metaphor. This suggests that the granularity of some conceptual mappings might need further refinement within the KG or that the LLM's conditioning needs more nuanced control.

Lack of Novelty in Certain Domains: Although creativity is generally high, some generated metaphors, particularly for less common conceptual mappings, might still lean towards more conventional expressions. This indicates a potential limitation in the current scope of learned "metaphorical morphisms" or the need for a larger, more diverse dataset for rarer metaphor types to better train the LLM.

Syntactic Inflexibility: In a few instances, the generated sentence structure might be slightly less natural than human-written text, particularly when complex role alignments are involved. This suggests an area for improvement in the integration of the KG's structural guidance with the LLM's fluency capabilities.

The model's behavior generally aligns with the principles of Conceptual Metaphor Theory, demonstrating its capacity to transfer properties and relations from source domains to target domains in a structured manner. The explicit representation of metaphor structure patterns within the KG allows for a degree of interpretability not typically found in end-to-end neural models. The ability to trace a generated metaphor back to its source and target concepts, frames, and mapped roles provides a clear understanding of the underlying conceptual transfer.

5 Conclusion

The comprehensive framework presented in this research significantly advances the computational understanding and generation of metaphor. By creating a richly annotated dataset, constructing a detailed Metaphor Knowledge Graph (MKG) grounded in linguistic theories, and developing an innovative graph-based generation methodology that effectively

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

integrates Large Language Models (LLMs), this work addresses critical challenges in the field. The intrinsic evaluation of the MKG demonstrates its robust structural quality and comprehensive representation of metaphorical knowledge. The extrinsic evaluation, through the task of metaphor generation, showcases the proposed model's superior performance in terms of fluency, creativity, and faithfulness compared to established baselines, as evidenced by both automatic and human assessments. The integration of mathematical principles from Category Theory provides a principled approach to generating novel and structurally analogous metaphors, pushing the boundaries of AI's creative capabilities. The implications of this research for metaphor-aware Natural Language Understanding (NLU) and reasoning are substantial. By explicitly representing the complex conceptual mappings and semantic role alignments inherent in metaphors, the MKG provides a transparent and interpretable knowledge source. This structured knowledge can enable NLU systems to move beyond superficial lexical understanding, allowing them to grasp the deeper, non-compositional meanings of figurative language. Such enhanced understanding is crucial for applications requiring sophisticated semantic analysis, including advanced question answering, nuanced information extraction, and more cognitively aligned dialogue systems. The ability to generate controlled and interpretable metaphors also offers new avenues for human-AI collaboration in creative writing and communication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work the authors used ChatGPT 3.5 and Gemini in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1]. Collin F. Baker. FrameNet: A knowledge base for natural language processing. In Miriam R. L. Petruck and Gerard de Melo, editors, Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-2014), pages 1–5, Baltimore, MD, USA, June 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-3001. URL https://aclanthology.org/W14-3001/.

[2]. Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet project. In 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1, pages 86–90, Montreal, Quebec, Canada, August 1998. Association for Computational Linguistics. doi: 10.3115/980845.980860. URL https://aclanthology.org/P98-1013/.

[3]. Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: Commonsense transformers for automatic

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

knowledge graph construction. In Anna Korhonen, DavidTraum, and Lluís Màrquez, editors, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4762–4779, Florence, Italy, July 2019. Association for Computational Linguistics. doi:10.18653/v1/P19-1470. URL https://aclanthology.org/P19-1470/.

[4]. Tuhin Chakrabarty, Smaranda Muresan, and Nanyun Peng. Generating similes effortlessly like a pro: A style transfer approach for simile generation. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 6455–6469, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.524. URL https://aclanthology.org/2020.emnlp-main.524/.

[5]. Tuhin Chakrabarty, Xurui Zhang, Smaranda Muresan, and Nanyun Peng. MERMAID: Metaphor generation with symbolism and discriminative decoding. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 4250–4261, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.336. URL https://aclanthology.org/2021.naacl-main.336/.

[6]. Seungmin Choi and Yuchul Jung. Knowledge graph construction: Extraction, learning, and evaluation. Applied Sciences, 15(7), 2025. ISSN 2076-3417. doi: 10.3390/app15073727. URL https://www.mdpi.com/2076-3417/15/7/3727.

[7]. Shivani Choudhary, Tarun Luthra, Ashima Mittal, and Rajat Singh. A survey of knowledge graph embedding and their applications, 2021. URL https://arxiv.org/abs/2107.07842.

[8]. Ellen Dodge, Jisup Hong, and Elise Stickles. MetaNet: Deep semantic automatic metaphor analysis. In Ekaterina Shutova, Beata Beigman Klebanov, and Patricia Lichtenstein, editors, Proceedings of the Third Workshop on Metaphor in NLP, pages 40–49, Denver, Colorado, June 2015. Association for Computational Linguistics. doi: 10.3115/v1/W15-1405. URL https://aclanthology.org/W15-1405/.

[9]. Evan Dodge, David Schlangen, and Massimo Poesio. Metaphor detection in discourse. In Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pages 384–392, Prague, Czech Republic, 2015. Association for Computational Linguistics.

[10]. Hugging Face. tner/ontonotes5 dataset. https://huggingface.co/datasets/tner/ontonotes5, 2025.

[11]. Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. Neural metaphor detection in context. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 607–613, Brussels, Belgium, October-November 2018. Association

Received: August 02 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

for Computational Linguistics. doi:10.18653/v1/D18-1060. URL https://aclanthology.org/D18-1060/.

[12]. Dedre Gentner and Brian Falkenhainer. Generating a specific class of metaphors. In Proceedings of the 14th Annual Conference of the Cognitive Science Society, CogSci 1992, pages 406–411. Lawrence Erlbaum Associates, 1992.

[13]. Pragglejaz Group. Mip: A method for identifying metaphorically used words in discourse. Metaphor and Symbol, 22:1–39, 2007.

[14]. E. Dario Gutiérrez, Ekaterina Shutova, Tyler Marghetis, and Benjamin Bergen. Literal and metaphorical senses in compositional distributional semantic models. In Katrin Erk and Noah A. Smith, editors, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 183–193, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1018. URL https://aclanthology.org/P16-1018/.

[15]. Mark Johnson and George Lakoff. Metaphors We Live By. University of Chicago Press, 1980.

[16]. Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. Spanbert: Improving pre-training by representing and predicting spans, 2020. URL https://arxiv.org/abs/1907.10529.

[17]. Karin Kipper Schuler, Anna Korhonen, and Susan Brown. VerbNet overview, extensions, mappings and applications. In Ciprian Chelba, Paul Kantor, and Brian Roark, editors, Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Tutorial Abstracts, pages 13–14, Boulder, Colorado, May 2009. Association for Computational Linguistics. URL https://aclanthology.org/N09-4007/.

[18]. George Lakoff. The contemporary theory of metaphor. In Andrew Ortony, editor, Metaphor and Thought, pages 202–251. Cambridge University Press, 1993.

[19]. Yucheng Li, Shun Wang, Chenghua Lin, Frank Guerin, and Loic Barrault. FrameBERT: Conceptual metaphor detection with frame embedding learning. In Andreas Vlachos and Isabelle Augenstein, editors, Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pages 1558–1563, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.eacl-main.114. URL https://aclanthology.org/2023.eacl-main.114/.

[20]. Yucheng Li, Frank Guerin, and Chenghua Lin. Finding challenging metaphors that confuse pretrained language models, 2024. URL https://arxiv.org/abs/2401.16012.

[21]. H. Liu and P. Singh. Conceptnet — a practical commonsense reasoning tool-kit. BT Technology Journal, 22(4):211–226, 2004. doi: http://dx.doi.org/10.1023/B:BTTJ.0000047600.45421.6d.

[22]. Rui Mao, Chenghua Lin, and Frank Guerin. Word embedding and WordNet based metaphor identification and interpretation. In Iryna Gurevych and Yusuke Miyao, editors, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1222–1231, Melbourne, Australia, July

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1113. URL https://aclanthology.org/P18-1113/.

[23]. George A. Miller. WordNet: A lexical database for English. In Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994, 1994. URL https://aclanthology.org/H94-1111/.

[24]. Saif M. Mohammad, Mona Salameh, and Svetlana Vappu. MOH-X: A dataset for metaphor detection in context. In Proceedings of the Tenth Linguistic Annotation Workshop, pages 27–37, Berlin, Germany, June 2016. Association for Computational Linguistics. doi: 10.18653/v1/W16-2404. URL https://aclanthology.org/W16-2404.

[25]. Andrea Moro, Roberto Navigli, Francesco Maria Tucci, and Rebecca J. Passonneau. Annotating the MASC corpus with BabelNet. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pages 4214–4219, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). URL https://aclanthology.org/L14-1323/.

[26]. Miriam R. L. Petruck. Introduction to metanet: A special issue. https://www.jbe-platform.com/content/journals/10.1075/cf.8.2.01pet?crawler=true&mimetype=application%2Fpdf, 2018. Accessed: July 28, 2025.

[27]. Miriam R L Petruck and Ellen K Dodge. MetaNet: Repository, identification system, and applications. In Alexandra Birch and Willem Zuidema, editors, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts, Berlin, Germany, August 2016. Association for Computational Linguistics. URL https://aclanthology.org/P16-5008/.

[28]. Ekaterina Shutova. Design and evaluation of metaphor processing systems. Computational Linguistics, 41(4):579–623, December 2015. doi: 10.1162/COLI_a_00233. URL https://aclanthology.org/J15-4002/.

[29]. Robyn Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In Proceedings of the AAAI Conference on Artificial Intelligence, pages 4444–4451, 2017.

[30]. Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal, Tina Krennmayr, and Trijntje Pasma. A Method for Linguistic Metaphor Identification: From MIP to MIPVU, volume 14 of Converging Evidence in Language and Communication Research. John Benjamins Publishing Company, 2010. ISBN 9789027239037.

[31]. G.J. Steen, A.G. Dorst, J.B. Herrmann, A.A. Kaal, and T. Krennmayr. Vu amsterdam metaphor corpus, 2010.

[32]. Kevin Stowe, Leonardo Ribeiro, and Iryna Gurevych. Metaphoric paraphrase generation, 2020. URL https://arxiv.org/abs/2002.12854.

[33]. Kevin Stowe, Tuhin Chakrabarty, Nanyun Peng, Smaranda Muresan, and Iryna Gurevych. Metaphor generation with conceptual mappings. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, Proceedings of the 59th Annual Meeting

Received: August 02 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language
Processing (Volume 1: Long Papers), pages 6724–6736, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.524. URL https://aclanthology.org/2021.acl-long.524/.

[34]. Yorick Wilks, Adam Dalton, James Allen, and Lucian Galescu. Automatic metaphor detection using large-scale lexical resources and conventional metaphor extraction. In Ekaterina Shutova, Beata Beigman Klebanov, Joel Tetreault, and Zornitsa Kozareva, editors, Proceedings of the First Workshop on Metaphor in NLP, pages 36–44, Atlanta, Georgia, June 2013. Association for Computational Linguistics. URL https://aclanthology.org/W13-0905/.

[35]. Zhiwei Yu and Xiaojun Wan. How to avoid sentences spelling boring? towards a neural approach to unsupervised metaphor generation. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 861–871, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1092. URL https://aclanthology.org/N19-1092/.

[36]. Jadhav, H. M., Mulani, A., & Jadhav, M. M. (2022). Design and development of chatbot based on reinforcement learning. *Machine Learning Algorithms for Signal and Image Processing*, 219-229.

[37]. Mulani, A. O., Jadhav, M. M., & Seth, M. (2022). Painless machine learning approach to estimate blood glucose level with non-invasive devices. In *Artificial intelligence, internet of things (IoT) and smart materials for energy applications* (pp. 83-100). CRC Press.

[38]. Kashid, M. M., Karande, K. J., & Mulani, A. O. (2022, November). IoT-based environmental parameter monitoring using machine learning approach. In *Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1* (pp. 43-51). Singapore: Springer Nature Singapore.

[39]. Aiwale, S., Kolte, M. T., Harpale, V., Bendre, V., Khurge, D., Bhandari, S., ... & Mulani, A. O. (2024). Non-invasive Anemia Detection and Prediagnosis. *Journal of Pharmacology and Pharmacotherapeutics*, *15*(4), 408-416.

[40]. Mulani A.O., Sardey M.P., Kinage K., Salunkhe S.S., Fegade T., Fegade P.G.. ML-powered Internet of Medical Things (MLIoMT) Structure for Heart Disease Prediction. *Journal of Pharmacology and Pharmacotherapeutics*. 2024;16(1):38-45. doi:10.1177/0976500X241281490

Received: August 02 2025