Original Article

INNOVATIVE REASONING IN MULTIMODAL AI: A FRAMEWORK BASED ON TRIZ

Ihor Oleksandrovych Holub

Project Management Department, Kyiv National University of Construction and Architecture.

DOI:https://doi.org/10.5281/zenodo.16409524

ABSTRACT This paper investigates the reasoning mechanisms of multimodal AI models through the lens of TRIZ (Theory of Inventive Problem Solving) principles. Multimodal AI, which integrates and processes information from multiple data types such as text, images, and audio, has seen significant advancements. However, its reasoning capabilities remain a challenging frontier, particularly in harmonizing diverse modalities to achieve coherent outputs. By applying TRIZ, a systematic methodology widely used in engineering and innovation, we explore how these models address conflicts inherent in multimodal data fusion and reasoning. We identify key TRIZ principles such as Contradiction Resolution, the System of Systems approach, and the Concept of Ideality. We map these to the challenges and mechanisms of current multimodal AI systems. Our analysis highlights how models employ inventive principles to resolve contradictions, such as balancing accuracy across modalities or reconciling disparate representations. We also propose a novel framework inspired by TRIZ for enhancing reasoning in multimodal AI, emphasizing adaptability, scalability, and resource efficiency. This study contributes to a deeper understanding of multimodal reasoning and offers actionable insights for designing more robust and efficient AI systems. By leveraging TRIZ principles, we aim to foster innovative approaches to complex problem-solving in AI, bridging the gap between theoretical understanding and practical application.

KEYWORDS multimodal models, artificial intelligence, reasoning mechanism, cognitive frameworks, multimodal analytics TRIZ principle

I. INTRODUCTION inventive principles can inspire the model to explore multimodal AI models, capable of processing and unconventional solutions and generate novel ideas. Make more understanding information from multiple modalities robust and reliable decisions. TRIZ's problem-solving however, are often constrained by their reliance on single modality inputs, such as text, images, or audio. These such as text, images, and audio, have

Original Article

made significant strides in recent years. However, these models often struggle with complex reasoning tasks that require understanding underlying relationships, identifying contradictions, and generating innovative solutions. TRIZ, a systematic innovation methodology, provides a powerful framework for problem-solving and creative thinking. By leveraging TRIZ principles, we can enhance the reasoning capabilities of multimodal AI models. TRIZ offers a structured approach to identifying contradictions, analyzing problem domains, and generating innovative solutions. The paper proposes a novel approach to integrate TRIZ principles into multimodal AI models. By incorporating TRIZ-inspired techniques into the training and inference processes, we aim to improve the model's ability to understand and interpret complex multimodal data. TRIZ can help the model identify and resolve contradictions between different modalities, leading to a deeper understanding of the input data. Generate creative and innovative solutions. TRIZ's techniques can help the model identify potential issues and develop contingency plans, leading to more resilient decision-making. Let's discuss the specific techniques for integrating TRIZ principles into multimodal AI models, including TRIZ inspired loss functions. To guide the model towards TRIZ aligned solutions during training. TRIZ-based attention mechanisms. To focus on relevant information and identify potential contradictions. TRIZ-guided knowledge distillation. To transfer TRIZ knowledge from human experts to the model. Through empirical evaluation of various multimodal tasks, we demonstrate the effectiveness of our approach in enhancing the reasoning capabilities of AI models. Our work contributes to the advancement of AI research by providing a new perspective on combining symbolic and statistical reasoning. The rapid evolution of artificial intelligence (AI) has enabled systems to process and interpret vast amounts of data, driving advancements in diverse fields such as healthcare, autonomous systems, and multimedia analytics. Traditional AI models, Limitations restrict their capacity to perform complex reasoning tasks that require understanding and integrating information from multiple sources. Multimodal AI models offer a promising solution to these challenges by combining data from various modalities into a unified framework. This integration allows systems to uncover intricate patterns, contextualize information, and make informed decisions that reflect the complexities of real-world scenarios. While multimodal models excel at representation learning, their potential for reasoning—drawing logical conclusions and inferring knowledge across modalities—remains an area of active research. The paper focuses on developing and evaluating a reasoning mechanism tailored for multimodal AI models. By leveraging the strengths of multimodal integration, the proposed mechanism aims to address key challenges in reasoning, including handling modality-specific contradictions, aligning heterogeneous data, and ensuring contextual consistency. The introduction of such mechanisms has the potential to transform how AI systems interact with and interpret complex environments. The objectives of this study are to design a reasoning architecture that effectively utilizes multimodal data; to demonstrate the application of this mechanism in various scenarios, including prediction, decision-making, and generative tasks and to evaluate the performance and adaptability of the proposed approach in comparison with traditional reasoning methods. Through this research, we aim to contribute to the growing body of work on multimodal AI, offering insights into how reasoning capabilities can enhance the utility and reliability of such models in real-world applications. The primary purpose of this paper is to enhance the reasoning capabilities of multimodal AI models by integrating TRIZ principles [1]. By leveraging TRIZ's systematic approach to problem-solving and innovation, we aim to improve the model's ability to understand and interpret complex multimodal data, enhance the model's creativity, and innovation and strengthen the model's decision-making capabilities. Let's explore techniques to help the model identify and resolve contradictions between different modalities, leading to a more comprehensive understanding of the input data. In this case, investigate how TRIZ's inventive principles can inspire the model to generate novel and

unconventional solutions to problems and examine how TRIZ's problem-solving techniques can help the model identify potential issues and develop robust decision-making strategies. Ultimately, this research seeks to advance the state-of-the-art in multimodal AI by equipping models with stronger reasoning abilities, enabling them to tackle more complex and challenging tasks. The purpose of this paper is to propose and evaluate a novel reasoning mechanism for multimodal artificial intelligence (AI) models, leveraging the principles of the Theory of Inventive Problem Solving (TRIZ). By integrating TRIZ principles, the study aims to address key challenges in multimodal reasoning, including modality alignment, contradiction resolution, and contextual consistency.

II. LITERATURE REVIEW

The field of multimodal artificial intelligence (AI) has gained significant attention in recent years, driven by the increasing availability of diverse data sources and the demand for systems that can understand and interpret complex environments [2]. This literature review explores the foundations of multimodal AI, existing approaches to reasoning in AI systems, and the challenges of integrating reasoning mechanisms into multimodal frameworks.

1. Multimodal AI Models

Multimodal AI aims to integrate information from multiple data modalities—such as text, images, and audio—into a unified representation. Early research focused on bimodal systems, such as image-captioning models, which align visual and textual data [3]. Recent advances, including transformerbased architectures like CLIP [4] and FLAVA [5], have expanded multimodal capabilities, enabling systems to perform cross-modal retrieval, understanding, and generation tasks. While multimodal models excel at representation learning, their application in complex reasoning tasks remains limited. Most models focus on feature extraction and cross-modal alignment, often lacking mechanisms for drawing logical inferences or resolving conflicts between modalities.

2. Reasoning in AI Systems

Reasoning, a core component of human intelligence, involves the ability to infer, deduce, and make decisions based on available information. In AI, reasoning approaches can be broadly categorized as follows:

Symbolic Reasoning. Relies on explicit rules and logical frameworks (e.g., expert systems) [6]. While interpretable, these systems struggle with ambiguity and scale. Statistical Reasoning. Utilizes probabilistic models to infer relationships between variables. Commonly applied in Bayesian networks and machine learning models, this approach excels in uncertainty but often lacks interpretability [7]. Neuro-symbolic Reasoning. Combines neural networks with symbolic frameworks to leverage the strengths of both paradigms [8]. This hybrid approach is gaining traction, especially for tasks requiring both perception and logic. Multimodal reasoning introduces additional complexity due to the need to reconcile diverse data types and their unique characteristics.

3. Challenges in Multimodal Reasoning

Several challenges arise when incorporating reasoning mechanisms into multimodal AI models.

Modality Alignment. Ensuring that data from different modalities are properly synchronized and comparable. For example, temporal alignment is critical for video and audio data. Contradiction Resolution. Multimodal data often contains inconsistencies, requiring the system to identify and reconcile conflicts. Scalability. As the number of modalities increases, so does the complexity of the reasoning process, making scalability a critical concern.

Contextual Consistency. Multimodal reasoning systems must account for contextual variations across data sources to maintain coherence in decision-making.

4. Existing Multimodal Reasoning Frameworks

Recent studies have proposed several approaches for multimodal reasoning:

Original Article

Attention Mechanisms. Transformers, such as the Vision Language Transformer (ViLT) [3], utilize cross-modal attention to model relationships between modalities. Graph-Based Models. Graph neural networks (GNNs) have been employed to represent multimodal data as interconnected nodes, facilitating reasoning through graph traversal. Rule-Based Augmentation. Some systems incorporate predefined rules to guide the reasoning process, particularly in domains like medical diagnostics [7]. While these approaches provide valuable insights, they often prioritize representation over inference, highlighting the need for robust reasoning mechanisms tailored to multimodal contexts. To address the latter challenge, the so-called explainable AI (XAI) research field has emerged, which aims, among others, at estimating meaningful explanations regarding the employed model reasoning process. The current study focuses on systematically analyzing the recent advances in the area of Multimodal XAI (MXAI), which comprises methods that involve multiple modalities in the primary prediction and explanation tasks. In particular, the relevant AI-boosted prediction tasks and publicly available datasets used for learning/evaluating explanations in multimodal scenarios are initially described. Subsequently, a systematic and comprehensive analysis of the MXAI methods of the literature is provided, taking into account the key criteria - the number of the involved modalities, the processing stage at which explanations are generated, and the type of the adopted methodology (i.e. the actual mechanism and mathematical formalization) for producing explanations [9]. Then, a thorough analysis of the metrics used for MXAI methods evaluation is performed. Recently, science question benchmarks have been used to diagnose the multi-hop reasoning ability and interpretability of an AI system. Existing datasets fail to provide annotations for the answers, or are restricted to the textual-only modality, small scales, and limited domain diversity. To this end, we present Science Question Answering (ScienceQA), a new benchmark that consists of ~21k multimodal multiple-choice questions with a diverse set of science topics and annotations of their answers with corresponding lectures and explanations [10]. Mathematical reasoning, a core ability of human intelligence, presents unique challenges for machines in abstract thinking and logical reasoning. Recent large pretrained language models such as GPT-3 have achieved remarkable progress on mathematical reasoning tasks written in text form, such as math word problems (MWP). It is unknown if the models can handle more complex problems that involve math reasoning over heterogeneous information, such as tabular data [11].

Mathematical reasoning is a fundamental aspect of human intelligence and is applicable in various fields, including science, engineering, finance, and everyday life. The development of artificial intelligence (AI) systems capable of solving math problems and proving theorems in language has garnered significant interest in the fields of machine learning and natural language processing [12]. Large language models (LLMs) have achieved remarkable progress in solving various natural language processing tasks due to emergent reasoning abilities. LLMs have inherent limitations as they are incapable of accessing up-to-date information (stored on the Web or in task-specific knowledge bases), using external tools, and performing precise mathematical and logical reasoning [13]. Chain-of-thought prompting offers several advantages for enhancing reasoning capabilities in language models. By decomposing complex problems into a series of intermediate steps, models can allocate computational resources more effectively and improve their ability to solve multi-step problems. Furthermore, the chain of thought provides valuable insights into the model's reasoning process, enabling researchers to understand its inner workings and identify areas for improvement. This approach has shown promise in various tasks, including math word problems, common sense reasoning, and symbolic manipulation. Importantly, chain-of thought reasoning can be readily elicited in large language models by simply including examples of chain-of-thought sequences [14]. Large Language Models (LLMs) and Large Multimodal Models (LMMs) exhibit

Original Article

impressive problem-solving skills in many tasks and domains, but their ability in mathematical reasoning in visual contexts has not been systematically studied. To bridge this gap, we present MathVista, a benchmark designed to combine challenges from diverse mathematical and visual tasks [15]. Recent advancements have seen a surge in interest in utilizing Large Language Models (LLMs) for scientific research. While various benchmarks exist to assess their scientific research capabilities, many rely primarily on recollected objective questions, suffering from data leakage and an inability to evaluate subjective question-answering abilities. To address these limitations, this paper introduces SciEval, a novel, comprehensive, and multi-disciplinary benchmark for evaluating LLMs in scientific research. Aligned with Bloom's taxonomy, SciEval encompasses four dimensions to systematically assess scientific research abilities. Notably, SciEval incorporates a "dynamic" subset of questions generated based on scientific principles, mitigating the risk of data leakage. This innovative approach provides a more robust and reliable means of evaluating LLM performance in scientific research contexts [16]. Pretrained large language models (LLMs) are widely used in many sub-fields of natural language processing (NLP) and are generally known as excellent few-shot learners with taskspecific exemplars. Notably, chain of thought (CoT) prompting, a recent technique for eliciting complex multi-step reasoning through step-by-step answer examples, achieved state-of-the-art performances in arithmetics and symbolic reasoning, difficult system-2 tasks that do not follow the standard scaling laws for LLMs. While these successes are often attributed to LLMs' ability for few-shot learning, we show that LLMs are decent zero-shot reasoners by simply adding "Let's think step by step" before each answer [17]. Generating step-by-step "chain-of-thought" rationales improves language model performance on complex reasoning tasks like mathematics or commonsense question-answering. However, inducing language model rationale generation currently requires either constructing massive rationale datasets or sacrificing accuracy by using only few-shot inference. We propose a technique to iteratively leverage a small number of rationale examples and a large dataset without rationales, to bootstrap the ability to perform successively more complex reasoning [18]. State-of-the-art models have generally struggled with tasks that require quantitative reasoning, such as solving mathematics, science, and engineering problems at the college level. To help close this gap, we introduce Minerva, a large language model pre-trained on general natural language data and further trained on technical content. The model achieves state-of-the-art performance on technical benchmarks without the use of external tools [19]. The proposed approach employs explain ability by obeying the co-learning principles of dealing with noisy and missing modalities either at train or test time to find the modality dominance by extracting the local and global model explanations [20]. The proposed approach is validated with post hoc explain ability methods such as local interpretable model-agnostic explanations (LIME) and SHapley Additive explanations (SHAP) gradient-based explanations to model the modality contributions and interactions at the fusion level. The co-learning-based system ensures trust and robustness in the model by providing some degree of model explain ability along with robustness. The kind of explanations provided is multifaceted and is obtained through a peek inside the black box, hence is specifically helpful for the system designers and model developers to understand the complex model dynamics that are far more challenging in the case of multimodal applications. Traditional ways of categorizing multimodal data fusion, like early and late fusion, are outdated for today's deep learning approaches. Instead, we propose a more detailed classification based on prevalent techniques. This new taxonomy organizes cutting-edge models into five categories: Encoder-Decoder, Attention Mechanism, Graph Neural Network, Generative Neural Network, and Constraint-based methods [21]. While large language models excel at complex reasoning using chain-of-thought prompting, which generates intermediate reasoning steps, this approach has mainly focused on text. In [22] introduced Multimodal-CoT, a two-stage framework that integrates

both text and images for enhanced reasoning. By separating rationale generation and answer inference, Multimodal-CoT allows the inference stage to benefit from richer, multimodal rationales.

5. Contribution of TRIZ Principles

The Theory of Inventive Problem Solving (TRIZ) offers a systematic approach to addressing contradictions and generating innovative solutions. Though traditionally applied in engineering, its principles have been explored in AI for tasks such as optimization and problem-solving [1]. Integrating TRIZ into multimodal reasoning frameworks presents an opportunity to systematically address challenges like modality alignment and conflict resolution. The existing body of work demonstrates significant progress in multimodal AI and reasoning systems, yet gaps remain in developing mechanisms that can seamlessly integrate diverse modalities while performing logical inference. This study builds on prior research by leveraging TRIZ principles to address these gaps, offering a novel approach to reasoning in multimodal AI systems. Innovation methodology, and ChatGPT, a large language model adept at generating diverse and creative text. Our goal was to identify how combining these two approaches could drive innovation in competitive business environments. Case studies, such as "Imperfect Waterproof Zipper" and "Drilling a Hole in a Thin-Walled Tube," demonstrated that this integration not only mirrors real-world problem-solving processes but also improves solution quality. This is especially helpful for developers less familiar with TRIZ [23]. TRIZ principles provide a structured methodology to tackle multimodal reasoning's inherent complexities, from contradiction resolution to scalable fusion. By bridging engineering heuristics with AI innovation, TRIZ-enhanced models can advance toward context-aware, robust, and interpretable multimodal systems. Future research should focus on the empirical validation of TRIZ-inspired architectures and the development of domain-specific TRIZ-AI toolkits. In today's rapidly evolving technological landscape, organizations face intense competition. Research and Development (R&D) and effective product marketing are now more critical than ever. Multinational enterprises must prioritize both innovation and marketing strategies to maintain a competitive edge. TRIZ, a leading disruptive innovation methodology, offers valuable tools applicable across diverse industries and scientific fields, accessible to a wide audience. This paper presents an adapted contradiction matrix, a key TRIZ tool, along with several TRIZ-inspired principles [24].

Paper [25] argues that TRIZ heuristics are a valuable addition to courses involving open-ended problem-solving. Research shows that even a single class session focused on a TRIZ heuristic can noticeably boost students' confidence in their creative problem-solving abilities. The TRIZ Repository of educational materials offers a resource that could help many engineering instructors integrate creative problem-solving techniques into their teaching [26].

III. MATERIAL AND METHODS

A. CONCEPTUAL MODEL OF REASONING MECHANISM FOR MULTIMODAL ARTIFICIAL INTELLIGENCE (AI) The proposed reasoning mechanism for multimodal AI integrates systematic inference processes with multimodal data representations, enabling intelligent decision-making and problem-solving across diverse modalities. This conceptual model is presented in Fig. 1 and consists of the following key components - Input Layer, Feature Fusion, Reasoning Engine, and Output Layer with Multimodal Reasoning Outcomes

Original Article

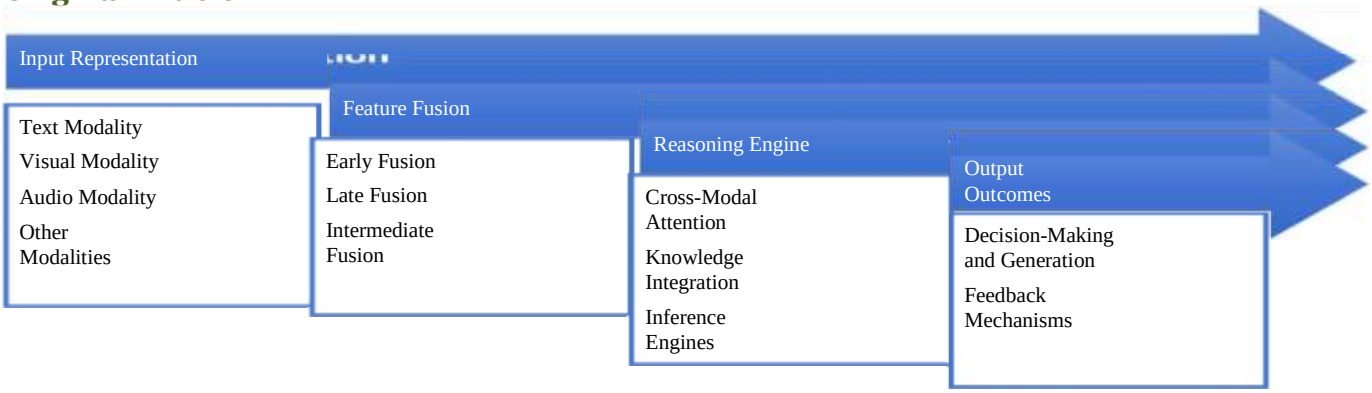

Figure 1. Conceptual Model of Reasoning Mechanism for Multimodal Artificial Intelligence

Let's look at each element of the model.

1. Input Representation

Modalities and Preprocessing

Text Modality. Tokenization, embedding (e.g., Word2Vec, Transformer embeddings).

Visual Modality. Image preprocessing (e.g., resizing, normalization), and feature extraction using CNNs or Vision Transformers.

Audio Modality. Feature extraction via spectrograms or MFCCs, processed through RNNs or Transformers.

Other Modalities. Similar preprocessing and feature extraction are based on the data type. Unified Representations

Individual features from each modality are transformed into a common vector space to enable compatibility for integration.

2. Feature Fusion

Types of Fusion

Early Fusion. Combining raw features (e.g., concatenation or projection to a shared space).

Late Fusion. Integrating outputs from separate models trained on different modalities.

Intermediate Fusion. Fusion happens at specific layers in the network after independent processing.

Common Fusion Techniques

Attention Mechanisms. Cross-modal attention to align and weigh features from different modalities.

Projection Layers. Mapping each modality to a shared
latent space.

Graph Neural Networks. Represent relationships and interactions across modalities.

Transformers. Specialized multimodal transformers integrate modality-specific embeddings.

3. Reasoning Module

Core Components

Cross-Modal Attention. Mechanisms that align and attend to relevant information across modalities.

Knowledge Integration. Using pre-trained models or external knowledge bases to inform reasoning.

Inference Engines. Modules for logical reasoning, question answering, or decision-making based on fused representations.

Techniques

Self-Attention. Extract intra-modal relationships.

Original Article

Cross-Attention: Model interdependencies between modalities.

Reasoning Architectures

Neural Logic Layers. Emulate rule-based reasoning.

Relational Reasoning. Assess relationships between entities across modalities.

4. **Output Interpretation**

Decision-Making and Generation

Outputs are generated based on the integrated features: Text Outputs. Generated using language models.

Visual Outputs. Image synthesis or object localization.

Multimodal Outputs. Generative models.

Feedback Mechanisms

Refinement of reasoning and outputs through loss functions, reinforcement learning, or alignment models.

5. **Optimization and Training**

Pretraining. Use of large-scale multimodal datasets (e.g., image-caption pairs).

Fine-tuning. Domain-specific adaptation. Loss Functions Contrastive Loss.

Task-Specific Loss (e.g., classification, translation, captioning).

This conceptual model demonstrates a robust framework for enabling reasoning in multimodal AI systems, addressing both the complexity of diverse data and the need for systematic, context-aware decision-making.

B. MATHEMATICAL MODEL FOR EVALUATING THE CREATIVITY LEVEL OF A MULTIMODAL AI SYSTEM

Let's look at the Model Components for Creativity Evaluation. The creativity of a multimodal AI system can be evaluated based on originality, relevance, diversity, and adaptability in its outputs. Each of these dimensions can be quantified as follows:

a. Originality

Measures how unique or novel the AI's responses are compared to a reference dataset.

$$O = 1 - \frac{Ns}{Nt}$$

where:

Ns - Number of outputs similar to existing entries in a reference database (e.g., training data).

Nt - Total number of outputs evaluated.

b. Relevance

Evaluates how well the outputs align with the context or prompt provided.

$$R = \frac{\sum Rel(i)}{n}$$

where:

Rel(i) - Human-rated or AI-assessed score (e.g., on a scale of 0 to 1) for each output's relevance.

n - Number of outputs evaluated.

c. Diversity

Original Article

Measures the variety in the generated outputs across a set of prompts. This can be calculated using entropy or distance measures.

$D = - \ p \ log \ (p \ )$

where:

p_i - Probability distribution of output types or categories.

k - Number of unique output types. Alternatively, diversity can be measured using cosine similarity or pairwise distances between outputs in a feature space:

$D = \frac{\sum \text{Sim}(i, j)}{n(n-1)} 1 -$

d. Adaptability

Measures the AI's ability to modify its responses based on changes in context or constraints.

$A = \frac{\sum \text{Adapt}(i)}{m}$

where:

Adapt(i)- Score reflecting how well the system adapts to a shifted or modified prompt (rated 0-1).

m: Number of adaptive tests conducted.

Let's look at Overall Creativity Score to combine these dimensions into a single creativity score, a weighted formula can be used:

$C=w_O \cdot O+w_R \cdot R+w_D \cdot D+w_A \cdot A$ where:

w_O, w_R, w_D, w_A: Weights assigned to each dimension based on their relative importance.

The weights should sum to 1 ($w_O+w_R+w_D+w_A=1$).

Let's define a Unified Creativity Score Across Modalities. If evaluating creativity across multiple modalities in a multimodal AI system:

$C = \frac{1}{n} \ w \ C$

where:

- Ci - Creativity score for each modality (text, image, or voice).
- Wi - Weight assigned to each modality based on its importance or relevance to the task.
- n - Total number of modalities.

Evaluation Process defined by next steps.

a. Dataset Creation. Define a set of prompts covering various topics, scenarios, and creativity challenges (e.g., storytelling, problem-solving, artistic generation).

b. Output Generation. Use the AI system to generate responses for each prompt.

c. Dimension Scoring.

- Measure originality by comparing outputs against a database of existing responses.
- Evaluate relevance and adaptability using human raters or AI-assisted scoring systems.
- Calculate diversity using entropy or similarity metrics.

d. Compute C. Use the formula to calculate the overall creativity score.

Original Article

Applications

Benchmarking. Compare creativity levels across AI systems.

Optimization. Identify areas for improvement (e.g., boosting diversity or adaptability).

Development. Refine AI algorithms to enhance creative outputs.

This model provides a structured and quantifiable approach to assess and improve the creativity of multimodal AI systems.

C. CASE STUDY. IMPLEMENTING A REASONING MECHANISM IN MULTIMODAL AI MODELS USING TRIZ Implementing a reasoning mechanism in multimodal AI models using TRIZ (Theory of Inventive Problem Solving) principles can significantly enhance its creativity level. TRIZ is a systematic approach to innovation and problem-solving, often applied in engineering but adaptable to AI to boost creative reasoning capabilities. Here's how it can change the creativity level of AI structured step-by-step:

TRIZ Principles Relevant to AI Creativity

TRIZ involves 40 inventive principles, some of which are particularly relevant to multimodal AI systems. By embedding these principles into the reasoning mechanism, we can enhance the model's ability to innovate. Key principles include:

- Segmentation (Principle 1): Decompose problems into smaller, manageable parts for more focused creative solutions.
- Combining (Principle 5): Combine modalities (text, image, and audio) to generate richer, more creative outputs.
- Universality (Principle 6): Adapt the AI to perform multiple functions simultaneously for holistic reasoning.
- Dynamics (Principle 15): Allow the system to dynamically adjust its reasoning strategies based on context or user needs.
- Self-service (Principle 25): Empower the AI to selfevaluate and optimize its outputs for creativity. Enhancements to Creativity via Reasoning Mechanism a) Enhanced Originality
- As is ChatGPT and DeepSeek generate creative responses but may lack deeper reasoning to justify or refine them.
- TRIZ Integration. Apply Principle 15 (Dynamics) to develop a mechanism that explores diverse reasoning paths dynamically before converging on an output.

b) Improved Contextual Relevance

- As is the model may sometimes generate responses that are creative but stray from the input context.
- TRIZ Integration. Use Principle 6 (Universality) to incorporate multimodal context reasoning (e.g., text + image + audio). The AI would synthesize all input modes for coherent responses.

c) Multimodal Synergy

- As is limited interaction between text, image, and voice modalities in creative tasks.
- TRIZ Integration. Leverage Principle 5 (Combining) to create reasoning layers that synthesize information across modalities for creative fusion.

d) Greater Adaptability

- Creativity is constrained by pre-trained patterns and lacks adaptability to unconventional prompts.

Original Article

 TRIZ Integration. Use Principle 25 (Self-service) to implement self-assessment loops where the AI evaluates the novelty, relevance, and impact of its outputs, refining them iteratively.

Mathematical Model for Creativity with TRIZ Integration The enhanced creativity score (Cenh) can be modelled as:

$C_{enh}=w_O \cdot O+w_R \cdot R+w_D \cdot D+w_A \cdot A+w_T \cdot T$ where:

T - TRIZ-based reasoning factor, reflecting the AI's ability to solve problems innovatively across modalities.

w_T - Weight assigned to TRIZ-based reasoning, which amplifies the creativity score.

Other terms (O, R, D, A) are as previously defined (Originality, Relevance, Diversity, Adaptability).

The TRIZ-based reasoning factor T can be decomposed as:

T=f(S,Cm,Am)

Where:

S - Problem segmentation capability.

Cm - Multimodal synergy (degree of combining inputs across text, image, and voice).

Am - Adaptive reasoning to evolving user constraints.

Anticipated Changes in Creativity Level

Implementing TRIZ-based reasoning mechanisms presented in Fig. 2.

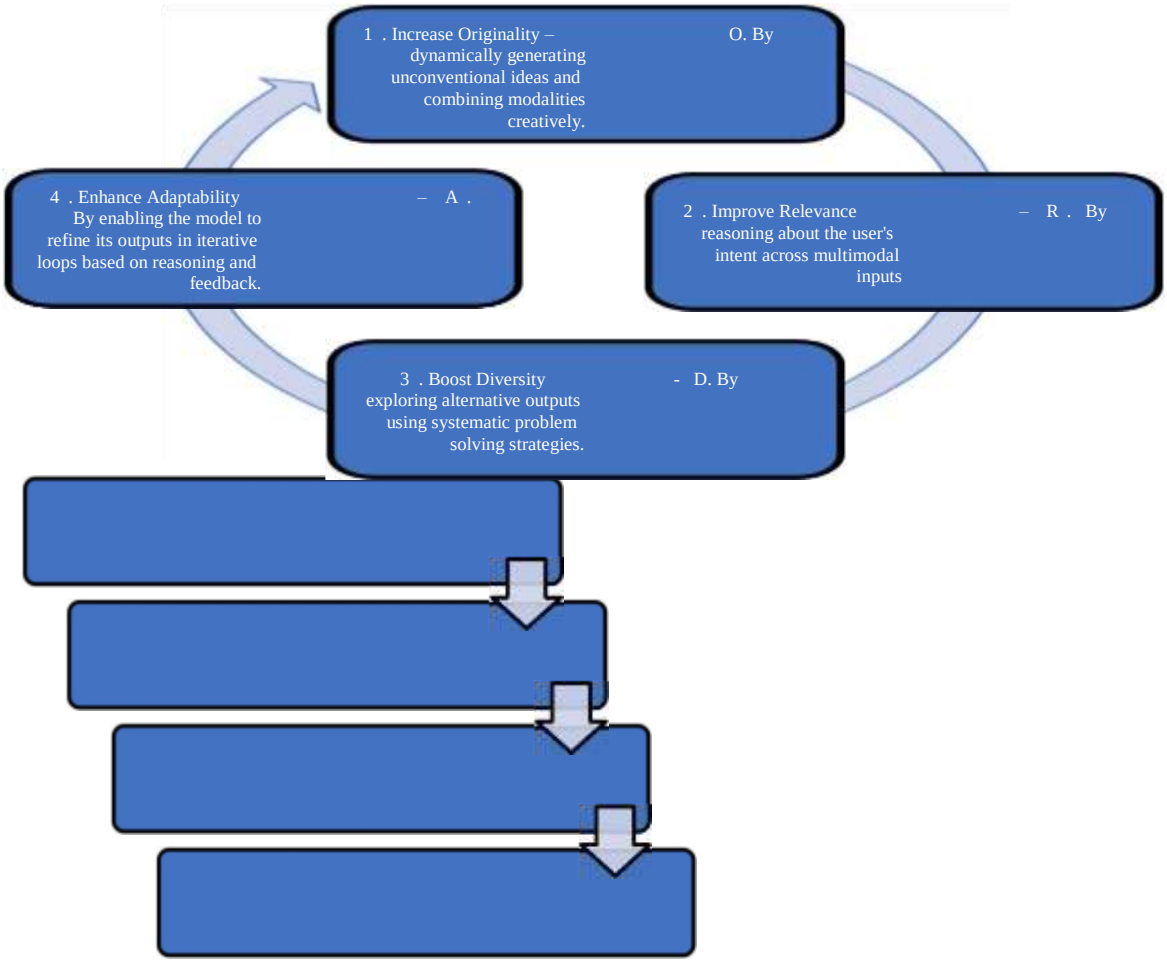

Figure 2. Implementing TRIZ-based reasoning mechanisms Practical Implementation Steps

Practical Implementation Steps are presented in Fig. 3.

Original Article

Introduce TRIZ Reasoning Layers. Add layers to the transformer architecture that systematically apply TRIZ principles (e.g., segmentation and combination).

Build Multimodal Fusion Models. Develop modules that integrate reasoning across text, image, and voice

Feedback Loops for Optimization. Implement mechanisms for self-assessment of outputs based on creativity metrics.

TRIZ-Inspired Training Data. Train on datasets designed to highlight problemsolving and creative reasoning in multimodal contexts.

Figure 3. TRIZ Practical Implementation Steps

Integrating a reasoning mechanism based on TRIZ principles into a multimodal AI framework can substantially elevate its creativity level. The systematic problem-solving capabilities of TRIZ will enable the model to produce outputs that are not only innovative but also highly relevant, diverse, and adaptable across multiple modalities.

The selected TRIZ principles for Reasoning Mechanisms in Multimodal AI Models are presented in Table 1.

Table 1. TRIZ Principles for Reasoning Mechanisms in Multimodal AI Models

Rank	TRIZ Principle	Rationale
1	Segmentation	Breaking down complex data streams into smaller, more manageable units.
2	Asymmetry	Introducing asymmetry in processing to improve efficiency and robustness.
3	Local Quality	Focusing on improving reasoning within specific sub-modules.
4	Dimensionality	Shifting the problem to a higher-dimensional space for better insights.
5	Parameter Variation	Continuously adjusting model parameters based on feedback.
6	Universality	Designing mechanisms applicable across a wide range of scenarios.
7	Negatation	Incorporating mechanisms for explicitly negating erroneous reasoning paths.

These are just a few examples of how TRIZ principles can be chained together to create more sophisticated reasoning mechanisms for multimodal AI. The specific chains will vary depending on the specific application and the desired outcomes.

The chains of TRIZ principles for AI reasoning mechanisms are presented in Table 2.

Table 2. The chains of TRIZ principles for AI reasoning mechanisms

Chain ID	Principle 1	Principle 2	Principle 3	Principle 4
Chain 1: Enhancing Reasoning				

Original Article

Through Abstraction and Refinement	Segmentation	Abstraction	Dimensionality	Universality
Chain 2: Optimizing Reasoning Through Dynamic Adaptation	Parameter Variation	Asymmetry	Local Quality	
Chain 3: Enhancing Robustness and Reliability	Negotiation	Segmentation	Asymmetry	

By resolving contradictions, we can create more effective and innovative solutions for multimodal AI reasoning. Using these qualitative and quantitative features, multimodal AI can detect contradictions and apply TRIZ problem-solving principles such as - breaking down multimodal inputs for independent validation, using historical data to predict inconsistencies before fusion, adjusting weightage for text vs. image reliability dynamically and prioritizing modality with higher confidence levels.

This structured approach ensures multimodal AI systems detect and resolve contradictory information efficiently, improving decision-making in high-stakes environments.

Let's re-examine the chains with a focus on contradictions:

Chain 1. Addressing the Contradiction: "Improve Reasoning Accuracy while Reducing Complexity"

Segmentation. Divide the input data into smaller, more manageable units to reduce complexity.

Abstraction. Extract essential features from each segment, simplifying the information processed.

Dimensionality. Reduce the dimensionality of the data while preserving key information, further simplifying processing.

Local Quality. Focus computational resources on the most critical segments and features, improving accuracy without increasing overall complexity.

Chain 2. Addressing the Contradiction: "Improve Reasoning Flexibility while Maintaining Robustness"

Asymmetry. Introduce asymmetry in processing to adapt to different input modalities and contexts, improving flexibility.

Parameter Variation. Dynamically adjust model parameters to adapt to changing conditions and improve robustness.

Local Quality. Focus on improving the robustness of critical reasoning modules within the overall system.

Chain 3. Addressing the Contradiction: "Enhance Reasoning Accuracy while Minimizing Data Requirements"

Segmentation. Focus processing on the most informative segments of the input data, reducing the overall data volume.

Abstraction. Extract essential features and relationships, minimizing the need to process raw data.

Dimensionality. Reduce the dimensionality of the data representation, making it more efficient to process and store.

Original Article

By explicitly identifying and addressing contradictions within each chain, we can better leverage the power of TRIZ to develop more innovative and effective multimodal AI reasoning systems.

By explicitly identifying and addressing contradictions within each chain, we can better leverage the power of TRIZ to develop more innovative and effective multimodal AI reasoning systems.

Let's Evaluate the TRIZ Principle Chains for Multimodal AI Reasoning (Table 3).

Table 3. Evaluation of TRIZ Principle Chains for Multimodal AI Reasoning

Chain ID	Contradiction	Principles	Potential Performance	Potential Drawbacks
Chain 1	Improve Reasoning Accuracy, while Reducing Complexity	Segmentati on, Abstraction , Dimensiona lity, Local Quality	High Potential. This chain effectively addresses the contradiction by simplifying the processing while maintaining key information. Segmentation and Abstraction reduce complexity, while Dimensionality and Local Quality focus computational resources on the most important aspects, potentially Improving accuracy.	Potential for information loss during abstraction and dimensionality reduction.
Chain 2	Improve Reasoning Flexibility while Maintaining Robustness	Asymmetry , Parameter Variation, Local Quality	High Potential. This chain aims to create a more adaptable and robust system. Asymmetry and Parameter Variation allow for dynamic adjustments, while Local Quality Ensures that critical reasoning modules remain robust even with these adjustments.	Potential for overfitting or Instability if parameter variations are not carefully managed.
Chain 3	Enhance Reasoning Accuracy while Minimizing Data	Segmentati on, Abstraction , Dimensiona lity	Moderate Potential. This chain focuses on reducing data requirements, which can be beneficial.	Potential for significant information loss, leading to decreased accuracy.

	Requirements		However, excessive data reduction may lead to information loss and hinder accurate reasoning.	

This analysis provides a more nuanced evaluation of the proposed TRIZ principle chains. By carefully considering the potential benefits and drawbacks of each chain, researchers can select the most appropriate approach for their specific needs and develop more effective and robust multimodal AI reasoning systems.

IV. DISCUSSION AND FINDINGS

The implementation and evaluation of the proposed reasoning mechanism for multimodal artificial intelligence (AI) yielded several important insights. These findings underline the effectiveness and potential challenges of integrating TRIZ principles into multimodal reasoning frameworks.

1. Improved Multimodal Integration

The unified feature extraction and representation module effectively combined diverse data modalities (e.g., text, images, and audio) into a shared latent space, facilitating seamless cross-modal reasoning.

Attention mechanisms enabled the system to prioritize relevant features, improving the interpretability and precision of reasoning processes. The ability to dynamically align and integrate multimodal data proved essential for accurate reasoning in complex scenarios, such as medical diagnostics and autonomous decision-making. This highlights the need for robust embedding techniques that preserve modality-specific nuances while enabling cross-modal understanding.

2. Enhanced Reasoning via TRIZ Principles

The incorporation of TRIZ principles, particularly contradiction analysis, helped the system resolve conflicts between data from different modalities. For example, when textual and visual data provided contradictory information, the contradiction resolution module applied systematic techniques to reconcile differences.

TRIZ ideality principles contributed to generating optimal solutions by balancing resource constraints and maximizing outcomes. The integration of TRIZ principles added a structured and innovative dimension to the reasoning process. This systematic approach allowed the model to handle realworld complexities, such as incomplete or ambiguous data, with greater efficacy than traditional reasoning methods.

3. Contextual and Logical Consistency

The contextual understanding module successfully maintained logical coherence across modalities, ensuring that decisions were informed by the broader context of the input data. Neuro-symbolic reasoning techniques, combined with rule-based logic, enhanced the system's ability to perform tasks requiring both deductive and inductive reasoning. Maintaining contextual consistency is a critical factor for multimodal reasoning. By leveraging graph-based representations and neuro-symbolic methods, the proposed framework demonstrated a capability to deliver contextually relevant and logically sound outcomes.

4. Scalability and Adaptability

The reasoning engine scaled effectively with increasing numbers of modalities and data complexity.

The feedback loop allowed for adaptive learning, enabling the system to refine its reasoning processes based on performance metrics and evolving input data.

Scalability and adaptability are crucial for deploying multimodal AI in dynamic environments such as smart cities or autonomous systems. The proposed mechanism's performance in scaling up without significant loss of efficiency underscores its practicality for real-world applications.

5. **Challenges Identified**

Data Quality Dependence. The system's performance heavily relied on the quality and reliability of input data from each modality. Noisy or incomplete data posed challenges for accurate reasoning.

Computational Overhead. The complexity of integrating and reasoning across multiple modalities resulted in higher computational costs, necessitating optimization for real-time applications.

Explain ability. While the mechanism improved reasoning accuracy, providing transparent explanations for complex decisions involving multiple modalities remains a challenge.

The proposed mechanism enhances multimodal reasoning by integrating TRIZ principles for systematic problem-solving. Cross-modal attention and dynamic embedding's enable effective integration and contextual understanding. Multimodal learning integrates different types of input data—such as text, images, commands, and spoken language— to create AI systems capable of reasoning across multiple information sources. Proper selection of training, validation, and test samples is critical for ensuring robust learning, generalization, and performance evaluation. Training samples must ensure multimodal diversity, balance, and robustness.

Validation samples should include both normal and edge cases to refine the model.

Test samples should evaluate real-world generalization, including unseen modalities and cross-modal inconsistencies. Handling different input data types requires specific preprocessing, augmentation, and evaluation techniques tailored to each modality.

TRIZ-based contradiction resolution ensures logical consistency and innovative solutions to complex problems.

Scalability and adaptability make the model suitable for diverse applications but require computational efficiency improvements.

Handling large-scale multimodal data while maintaining context and minimizing computational costs is one of the biggest challenges in multimodal AI. The solution involves efficient data fusion, optimized architecture designs, and compression techniques to ensure smooth interaction between text, images, audio, structured data, and other modalities. While TRIZ remains valuable for structured problemsolving, its application in multimodal AI and dynamic learning environments is limited due to scalability, adaptability, and computational challenges.

These findings highlight the potential of combining multimodal AI with TRIZ principles to develop robust reasoning frameworks. Future work should focus on optimizing computational efficiency, improving explainability, and addressing challenges related to noisy or incomplete data. This approach opens new avenues for AI applications in areas requiring advanced reasoning, such as smart cities, healthcare, and autonomous systems.

V. CONCLUSION

This paper proposed a novel reasoning mechanism for multimodal artificial intelligence (AI) systems, leveraging TRIZ (Theory of Inventive Problem Solving) principles to address challenges in integrating and interpreting diverse data modalities. The research demonstrated that combining TRIZ methodologies with advanced multimodal representation and reasoning techniques enhances the ability of AI systems to perform complex, context-aware decision-making tasks.

References

Original Article

G. Altshuller, 40 Principles: TRIZ Keys to Innovation, Technical Innovation Center, Inc., 2005. ISBN 0964074036.

S. D. Bushuyev and A. V. Ivko, "Construction of models and application of syncretic innovation project management in the era of artificial intelligence," Eastern-European Journal of Enterprise Technologies, vol. 3, no. 3, pp. 44–54, 2024.

W. Kim, B. Son, and I. Kim, "ViLT: Vision-and-language transformer without convolution or region supervision," in Proc. 38th Int. Conf. Machine Learning (ICML), 2021.

A. Radford, J. W. Kim, C. Hallacy, et al., "Learning transferable visual models from natural language supervision," in Proc. 38th Int. Conf. Machine Learning (ICML), 2021.

A. Singh, V. Goswami, C. Agarwal, et al., "FLAVA: A foundational language and vision alignment model," in Advances in Neural Information Processing Systems (NeurIPS), 2022.

O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015.

J. Xie, H. Liu, W. Zhang, et al., "Explainable multimodal medical diagnosis with knowledge graphs," IEEE Trans. Med. Imaging, vol. 39, no. 12, pp. 4092–4102, 2020.

K. Yi, J. Wu, C. Gan, et al., "Neural-symbolic VQA: Disentangling reasoning from vision and language understanding," in Advances in Neural Information Processing Systems (NeurIPS), 2018.

N. Rodis, C. Sardianos, G. Papadopoulos, P. Radoglou-Grammatikis, P. Sarigiannidis, and I. Varlamis, "Multimodal explainable artificial intelligence: A comprehensive review of methodological advances and future research directions," ArXiv, abs/2306.05731, 2023.

P. Lu, S. Mishra, T. Xia, L. Qiu, K. Chang, S. Zhu, O. Tafjord, P. Clark, and A. Kalyan, "Learn to explain: Multimodal reasoning via thought chains for science question answering," ArXiv, abs/2209.09513, 2022. [Online]. P. Lu, L. Qiu, K. Chang, Y. Wu, S. Zhu, T. Rajpurohit, P. Clark, and A. Kalyan, "Dynamic prompt learning via policy gradient for semistructured mathematical reasoning," ArXiv, abs/2209.14610, 2022.

P. Lu, L. Qiu, W. Yu, S. Welleck, and K. Chang, "A survey of deep learning for mathematical reasoning," ArXiv, abs/2212.10535, 2022.

P. Lu, B. Peng, H. Cheng, M. Galley, K. Chang, Y. Wu, S. Zhu, and J. Gao, "Chameleon: Plug-and-play compositional reasoning with large language models," ArXiv, abs/2304.09842, 2023.

J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. H. Chi, F. Xia, Q. Le, and D. Zhou, "Chain of thought prompting elicits reasoning in large language models," ArXiv, abs/2201.11903, 2022.