UTILIZATION OF GENERATIVE AI FOR THE CHARACTERIZATION AND IDENTIFICATION OF VISUAL UNKNOWNS

TECHNICAL SEMINAR REPORT

Submitted in partial fulfillment for the degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by

AFSAR

(P25UV24T065001)

1st Semester M.Tech

Computer Science and Engineering

Department of CSE, UVCE

Under the Guidance of

DR. SUNIL KUMAR G

Assistant Professor

Department of CSE

Department of Computer Science and Engineering

University of Visvesvaraya College of Engineering

K.R Circle, Bengaluru 560001

2024-2025

UNIVERSITY OF VISVESVARAYA COLLEGE OF ENGINEERING

K.R Circle, Bengaluru 560001

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that **AFSAR**, bearing register number **P25UV 24T065001**, has satisfactorily completed the technical seminar report entitled "**Utilization of generative AI for the characterization and identification of visual unknowns**" in partial fulfilment of the requirements for the 1^{st} Semester of the Degree of Master of Technology in Information Technology in the Department of Computer Science and Engineering of University of Visvesvaraya College of Engineering during the academic year 2024-2025.

GUIDE	**CHAIRPERSON**
Dr. Sunil Kumar G	Dr. Thriveni J
Assistant Professor	Chairperson
Department of CSE	Department of CSE,
UVCE	UVCE

Examiner 1:

Examiner 2:

DECLARATION

I, The Undersigned, hereby declare that the technical seminar report "**Utilization of generative AI for the characterization and identification of visual unknowns** " submitted for partial fulfilment of the requirements for the award of the degree of Master of Technology of the University of Visvesvaraya College of Engineering, Bengaluru, is a bonafide work undertaken by me under the supervision of **Dr.Sunil Kumar G**. This submission represents ideas expressed in my own words, and where ideas or words of others have been included, I have adequately and accurately cited and referenced the sources. I also declare that I have adhered to the ethical standards of academic honesty and integrity and have not misrepresented or fabricated any data, ideas, facts, or sources in my submission. I understand that any violation of the above may result in disciplinary action by the institute and/or the university and may also incur penal action from sources that have not been properly cited or from whom proper permission has not been obtained. This report has not previously served as the basis for the awarding of any degree, diploma, or similar title from any other university.

Place: Bengaluru **AFSAR**

Date: 02-04-2025 **P25UV24T065001**

ACKNOWLEDGMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without acknowledging the individuals who made it possible; their constant guidance and encouragement crowned my effort with success.

First, I express my sincere gratitude and respect to the **University of Visvesvaraya College of Engineering (UVCE)**, Bengaluru, for providing me with the opportunity to carry out my technical report work.

I am grateful to **Dr. Subhasish Tripathy**, Director of University Visvesvaraya College of Engineering, Bengaluru, for his unwavering support in completing this work.

I extend my sincere gratitude to **Dr. Thriveni J**, Head of the Department of Computer Science and Engineering at UVCE, for her invaluable guidance and steadfast support throughout this project.

I extend my deepest thanks to my guide, **Dr. Sunil Kumar G**, Assistant Professor in the Department of Computer Science and Engineering at UVCE, for providing all the resources that facilitated the smooth execution of my work. I also appreciate her assistance in reviewing the report to enhance its quality.

Last but not least, I wish to thank all the faculty members and staff in the Department of Computer Science and Engineering at UVCE for their constant support. I am immensely grateful to my parents and well-wishers for their continuous moral support and encouragement.

ABSTRACT

Artificial intelligence (AI) struggles with accurately interpreting out-of-library objects, limiting its effectiveness in real-world computer vision tasks. Analogical reasoning (AR), which applies abductive reasoning to infer knowledge from familiar scenarios, offers a potential solution. However, most visual AR applications focus on structured analogy-based problems rather than real-world image datasets. This study introduces the Image Recognition Through Analogical Reasoning Algorithm (IRTARA) and its generative AI counterpart, GIRTARA, designed to describe and predict out-of-library objects. IRTARA characterizes an unknown object by generating a term frequency list, which serves as input for GIRTARA to predict the object's identity.

To assess IRTARA's effectiveness, both quantitative and qualitative evaluations are conducted, comparing automated outputs with human-generated descriptions. The accuracy of GIRTARA's predictions is measured using cosine similarity analysis, achieving up to a 65% match with true labels. Results indicate that IRTARA consistently produces high-quality term frequency lists across multiple evaluation methods, demonstrating its reliability in characterizing out-of-library objects. Meanwhile, GIRTARA effectively leverages this information for object prediction, showing promising performance in interpreting unfamiliar visual data. This research highlights the potential of analogical reasoning in advancing AI-driven object recognition, paving the way for more adaptable and interpretable AI models. Future work will focus on refining the approach to improve predictive accuracy and expand applicability across diverse real-world scenarios.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Understanding the Limitations of Current AI Systems

Artificial Intelligence (AI) is often misunderstood as a technology capable of perfectly replicating human thought processes. However, this perception does not align with the actual capabilities of most AI systems (McCarthy, 2004). The majority of AI in use today is classified as "weak AI," meaning its functionality is limited to the specific tasks and datasets it was trained on (IBM, 2024). When weak AI encounters new situations or objects, its response depends on whether the input matches its pre-existing knowledge. These interactions result in four possible outcomes:

- **Known Knowns:** The AI correctly identifies a trained concept (in-library).
- **Unknown Knowns:** The AI mistakenly labels a trained concept as unknown
- **Known Unknowns :** The AI incorrectly classifies an unfamiliar concept as known.
- **Unknown Unknowns:** The AI correctly recognizes an unfamiliar concept as truly unknown.

Among these, unknown unknowns—referred to as "true unknowns" in this paper—represent the greatest challenge for AI systems. Addressing true unknowns effectively is a crucial step toward the development of "strong AI," which aims to achieve broader generalization in perception and cognition (IBM, 2024).

1.2 AI's Role in Image Recognition and the Challenge of True Unknowns

AI is widely used across industries, including healthcare, defense, business, and emerging technologies like handwriting recognition, depth perception, and augmented reality (Google, 2021). In these fields, the ability to correctly identify and describe visual information—especially when it includes objects absent from training data—is increasingly critical. A key example of this challenge is zero-shot learning, where AI must process and classify images of previously unseen objects (Socher et al., 2013; Pourpanah et al., 2022; Sun et al., 2021).

Among these, unknown unknowns—referred to as “true unknowns” in this paper—represent the greatest challenge for AI systems. Addressing true unknowns effectively is a crucial step toward the development of ”strong AI,” which aims to achieve broader generalization in perception and cognition (IBM, 2024).

1.3 Advancing Computer Vision with Analogical Reasoning and Generative AI

The potential of analogical reasoning in computer vision lies in its ability to infer meaning from unfamiliar objects without extensive retraining or additional labeled data. Traditional machine learning approaches struggle with this limitation, often requiring large-scale data collection and retraining when encountering new information. In contrast, analogical reasoning allows AI to classify and describe objects beyond its original training scope, mimicking a human-like capacity for adaptive learning. Generative AI models, which can create detailed textual or visual outputs from minimal input, are emerging as a promising complement to analogical reasoning (Combs et al., 2023a). This paper explores how these two approaches can be integrated to improve AI's ability to recognize and describe true unknowns. Specifically, we propose two novel frameworks

- Image Recognition Through Analogical Reasoning Algorithm (IRTARA)
- Generative Image Recognition Through Analogical Reasoning Algorithm (GIRTARA)

Chapter 2

LITERATURE SURVEY

2.1 Convolutional Neural Networks (CNN) in Image Processing

Convolutional Neural Networks (CNN) in Image Processing The paper "Aonvolutional Neural Networks (CNN) in Image Processing" proposes the CNN, Convolutional Neural Networks have significantly improved image recognition and processing tasks. These networks are designed to extract spatial hierarchies from images, enabling them to perform well in tasks such as object detection, classification, and segmentation. However, CNNs are inherently limited to patterns seen in training data, making them ineffective in recognizing truly unknown objects unless combined with other methodologies.

2.2 Analogical Reasoning in AI

Combs et al., 2022; Ichien et al., 2020; Gentner, 1983; Holyoak and Thagard, 1989; Hofstadter and Mitchell, 2022 Analogical reasoning has been predominantly explored in text-based domains. Traditional methods involve mapping relations between words (e.g., "king is to queen as man is to woman") and analyzing sentence structures for relational meaning. Early models were psychology-based but have evolved into neural network-based methods like Word2Vec, GloVe, and fastText. While well-established in natural language processing (NLP), analogical reasoning remains underutilized in image-based AI applications.

2.3 Visual Analogical Reasoning

Evans, 1964; Raven and Court, 1938; Sadeghi et al., 2015; Lu et al., 2019a The application of analogical reasoning in images has been mostly limited to geometric pattern-based tasks, such as Raven's Progressive Matrices. Algorithms like Visalogy attempt to apply analogy reasoning to object attributes, such as color and shape (e.g., "A red car is to a blue car as a red bike is to what?"). However, these methods struggle with analogies based on action, orientation, or semantic context.

2.4 Generative AI in Various Domains

Bihl and Talbert, 2020; Doshi et al., 2023; Singh, 2023; Raimondi et al., 2023; Eggmann et al., 2023 Generative AI has expanded rapidly, impacting fields like healthcare, education, and research. It has enabled advanced contextual understanding, making it useful in applications such as diagnosis assistance, automated content generation, and predictive modeling. Generative AI extends beyond simple data generation to higher levels of knowledge processing.

2.5 Generative AI in Programming and Coding

Chen et al., 2021; Dohmke, 2023; OpenAI, 2023a,b,c,d,e,f AI-driven code generation tools like Codex and GitHub Copilot leverage large language models (LLMs) to assist developers. Codex was initially built on GPT-3 but has been succeeded by more advanced models, such as GPT-4. These tools provide autocomplete functionality, refactoring suggestions, and even entire function implementations, significantly enhancing coding efficiency.

2.6 Large Language Models (LLMs) and Analogical Reasoning

Webb et al., 2023; Yu et al., 2023; Mitchell et al., 2023; Petersen and van der Plas, 2023 Studies have examined the analogical reasoning abilities of LLMs, revealing both strengths and weaknesses. While LLMs can recognize textual analogies, they struggle with deeper conceptual reasoning, particularly in complex or multimodal contexts (e.g., integrating text and images). Research continues to explore how LLMs can be improved for such tasks.

2.7 Multimodal Learning: Bridging Text and Images

Baltrušaitis et al., 2019; Tsimpoukelli et al., 2021; Li et al., 2022 Multimodal learning aims to integrate multiple data types, such as text, images, and audio, for more comprehensive AI understanding. Models like CLIP and Flamingo leverage cross-modal training, improving the ability to understand and generate analogies across different media. However, challenges remain in achieving deeper conceptual reasoning beyond pattern matching.

2.8 Challenges in Image-Based Analogical Reasoning

Li et al., 2023; Ichien et al., 2023; Webb et al., 2023; Combs and Bihl, 2024 Unlike text-based analogical reasoning, image-based analogy systems face challenges such as high computational requirements, the need for structured datasets, and limited generalizability. Some research has attempted to integrate generative AI to enhance analogy-based image reasoning, but this remains an emerging field.

2.9 Zero-Shot and Few-Shot Learning in AI

Brown et al., 2020; Radford et al., 2021; Lin et al., 2023 Zero-shot and few-shot learning approaches allow AI models to generalize from limited examples. GPT-3 and CLIP are notable examples, demonstrating impressive performance in unseen tasks. In visual analogy reasoning, zero-shot learning helps AI infer relationships without extensive labeled datasets, but it struggles with nuanced and abstract concepts.

2.10 AI in Search Engines and Conversational Agents

Mehdi, 2023; Pichai, 2023; OpenAI, 2023a,b,c,d,e,f AI-powered chatbots and search engines, such as Microsoft Bing Chat and Google Bard, have introduced advanced search and conversational AI functionalities. Bing Chat integrates a version of GPT-4 optimized for web searches, while Bard transitioned from LaMDA to the more advanced Pathways Language Model (PaLM). These models enhance user interaction by providing context-aware responses.

Chapter 3

METHODOLOGY

3.1 Proposed System

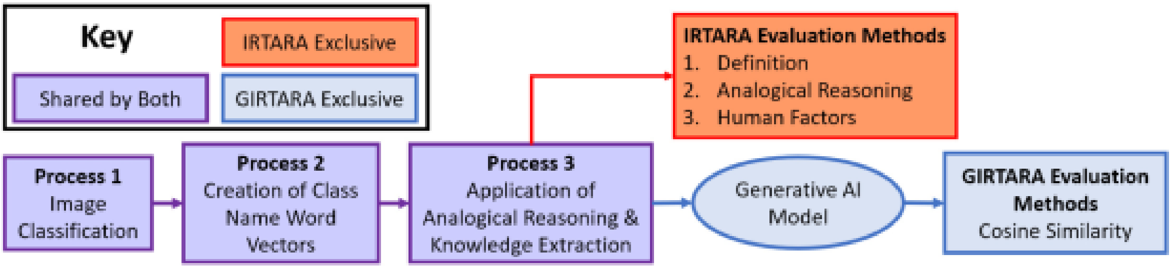

Figure 3.1: GIRTARA and IRTARA framework.

The proposed system is explained in Fig. 3.5. The Image Recognition Through Analogical Reasoning Algorithm (IRTARA) and the Generative Image Recognition Through Analogical Reasoning Algorithm (GIRTARA) are frameworks designed to enhance artificial intelligence's capability to interpret and characterize out-of-library (OOL) objects—those not present in a system's training data.

GIRTARA operates on a framework similar to IRTARA, with the primary distinction being its focus on generating descriptive characterizations for OOL objects. It follows the same initial steps: processing raw image data through a CNN, creating class name word vectors, and applying analogical reasoning. The outcome is a term frequency list that provides a descriptive characterization of the unfamiliar object, facilitating a more nuanced understanding of OOL images.

Both frameworks aim to bridge the gap in AI's ability to interpret and provide contextual information about objects that were not part of their training datasets, thereby enhancing the adaptability and comprehension of AI systems in real-world scenarios.

3.2 Workflow of Proposed System

The IRTARA and GIRTARA frameworks work through a three-step process involving image classification, creation of class name word vectors, and analogical reasoning with knowledge extraction. First, image classification is performed using a deep learning-based Convolutional Neural Network (CNN), where an input image is analyzed to extract visual features at multiple levels. The CNN layers detect edges, shapes, and high-level patterns,

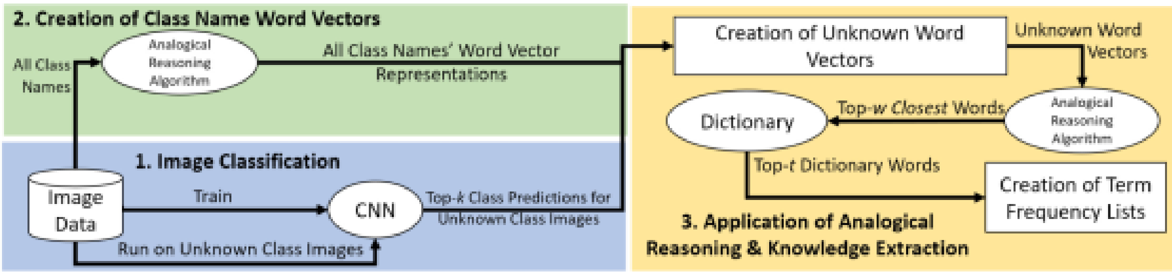

Figure 3.2: Workflow of proposed system (Process 1 - 3)

assigning a probability distribution to known object classes. Once classification is complete, the next step involves converting the predicted class names into numerical word vectors using word embedding techniques like GloVe or Word2Vec. These embeddings map the class names into a semantic space, preserving contextual relationships between words. The final step, analogical reasoning and knowledge extraction, involves comparing the vector representation of an unknown object with known class vectors using cosine similarity. If a close match is found, the system generates analogical insights, drawing relationships between the unknown and known objects. Additionally, knowledge is extracted from external sources such as Wikipedia or knowledge graphs to provide contextual information. This process enables the AI model to recognize, relate, and describe unknown entities intelligently, making it useful for applications in autonomous systems, object recognition, and AI-driven knowledge discovery.

3.3 IRTARA's Convolution Network Architecture.

GIRTARA improves upon IRTARA by introducing Generative AI into the process. The key difference lies in how the term frequency list is handled: GIRTARA refines the term frequency list generated by the analogical reasoning step by passing the top-5 words to a Generative AI model, which identifies the semantic meaning of these words and finds the most likely interpretation for the unknown image. This semantic enhancement allows GIRTARA to produce results with greater accuracy, as it understands the relationships between words and interprets them contextually. Gen AI and evaluate their similarity to true unknown objects. It operates similarly to IRTARA but may have unique characteristics in term selection, ranking, or evaluation. Word Selection: GIRATARA extracts the top five most frequent words associated with unknown objects from its dataset. Querying AI Models: The first five words from GIRATARA's term frequency lists were used to query various generative AI algorithms (e.g., ChatGPT, Bard, Bing AI). Prediction Analysis: The AI models were asked to infer the most likely object based on the five words. Similarity Calculation: Cosine similarity scores were used to compare AI predictions to the true unknown object, utilizing word vector representations from the GloVe-wiki-gigaword-300

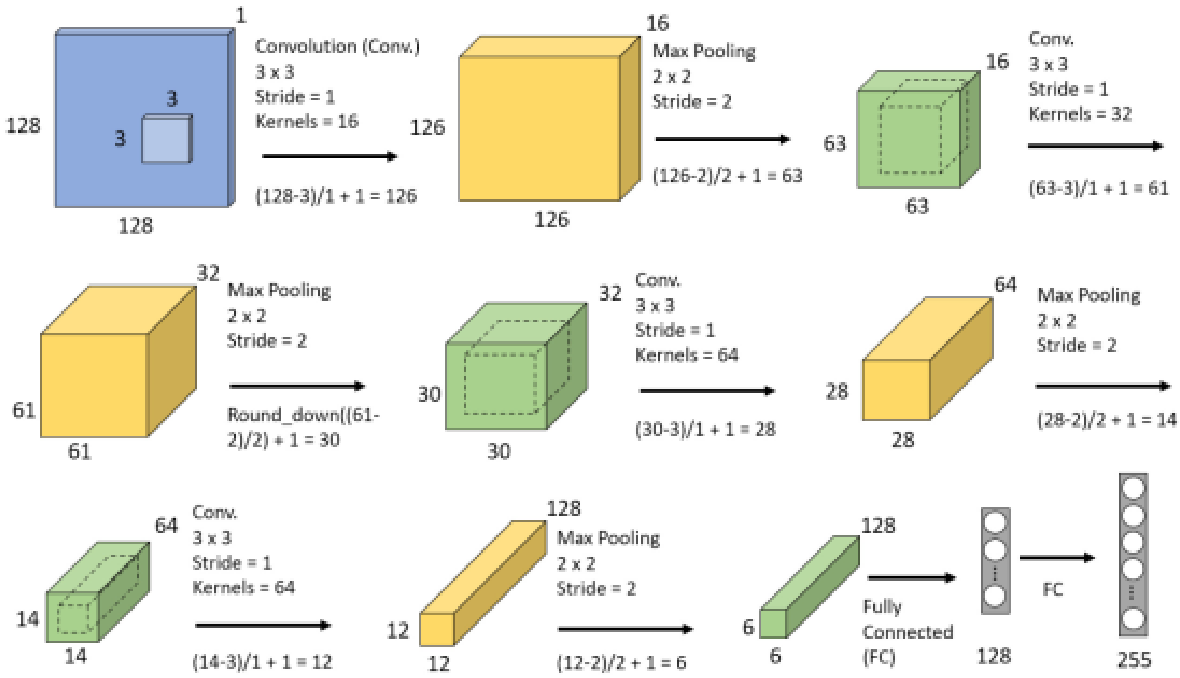

Figure 3.3: IRTARA's Convolution Network Architecture.

model. Future improvements could include combining multiple word vectors for better phrase representation.

3.4 Image Classification Using CNN

- The process begins by classifying the unknown image using a Convolutional Neural Network (CNN). This is a typical step in computer vision tasks, where a CNN is trained on a labeled dataset to learn to recognize various classes of images.
- The CNN outputs class probabilities for the unknown image.
- GIRTARA improves upon IRTARA by introducing Generative AI into the process. The key difference lies in how the term frequency list is handled.
- GIRTARA refines the term frequency list generated by the analogical reasoning step by passing the top-5 words to a Generative AI model, which identifies the semantic meaning of these words and finds the most likely interpretation for the unknown image.
- This semantic enhancement allows GIRTARA to produce results with greater accuracy, as it understands the relationships between words and interprets them contextually.

- Gen AI and evaluate their similarity to true unknown objects. It operates similarly to IRTARA but may have unique characteristics in term selection, ranking, or evaluation.
- Word Selection: GIRATARA extracts the top five most frequent words associated with unknown objects from its dataset.
- Querying AI Models: The first five words from GIRATARA's term frequency lists were used to query various generative AI algorithms (e.g., ChatGPT, Bard, Bing AI).
- Prediction Analysis: The AI models were asked to infer the most likely object based on the five words.
- Similarity Calculation: Cosine similarity scores were used to compare AI predictions to the true unknown object, utilizing word vector representations from the GloVe-wiki-gigaword-300 model.
- Future improvements could include combining multiple word vectors for better phrase representation.

3.5 Imbalanced Data Handling

The dataset was imbalanced with far fewer stroke cases compared to non-stroke cases. To address this issue:

- Random Over Sampling (ROS) was used to balance the dataset by duplicating minority class samples.
- After oversampling, the dataset contained an equal number of stroke and non-stroke cases.

3.6 Word Embedding Generation through Analogical Reasoning

Now, for the unknown image, a word vector needs to be created to represent the image's class. This vector is formed by considering the top-k class predictions (whose probabilities are above the threshold). Each of these classes) contributes to the formation of the unknown word vector. Example: Using the probabilities of Class A (0.60) and Class B (0.25), the unknown word vector is calculated as a weighted sum of the word vectors corresponding to these classes. Mathematically, the unknown word vector.

3.7 Analogical Reasoning for Word Matching

The training process begins with data preparation and feature engineering. Large-scale datasets were collected and preprocessed using various natural language processing (NLP) techniques, such as tokenization, stopword removal, and vectorization. One key approach was the use of Term Frequency (TF) lists, where models were trained to associate specific words with unknown objects. Additionally, models leveraged both supervised and unsupervised learning techniques to enhance their predictive capabilities. While supervised learning relied on labeled datasets for fine-tuning, unsupervised learning helped uncover latent patterns within the data. A total of 11 classifiers were implemented for stroke prediction, along with their corresponding formulas and definitions of terms:

3.7.1 Formula

A supervised learning algorithm that separates data into classes using an optimal hyperplane for effective classification. The equation is:

$$V_{\text{unknown}} = \sum_{i \in I} V_i \cdot P_i$$

Terms:

- V_i: is the word vector of class
- P_i: is the probability of class
- I: is the set of top k classes that meet the threshold criterion

3.8 Dictionary Lookup and Word Filtering

The proposed system is explained in Fig. 3.5. Once the closest words are identified, the next step is to look up their definitions in a dictionary. This helps understand the semantic context of the words. Words that are stop words or that lack significant meaning are removed. The remaining words are termed "definition words". Example: From the list of words ["feline," "pet," "animal," "catlike"], words like "animal" may be considered too broad, while "feline" and "pet" provide more useful context for identifying a cat. It is a predictive analysis-oriented algorithm based on probability theory. We have used logistic regression to train the model to find its accuracy and compared the model using a sequential DL model.

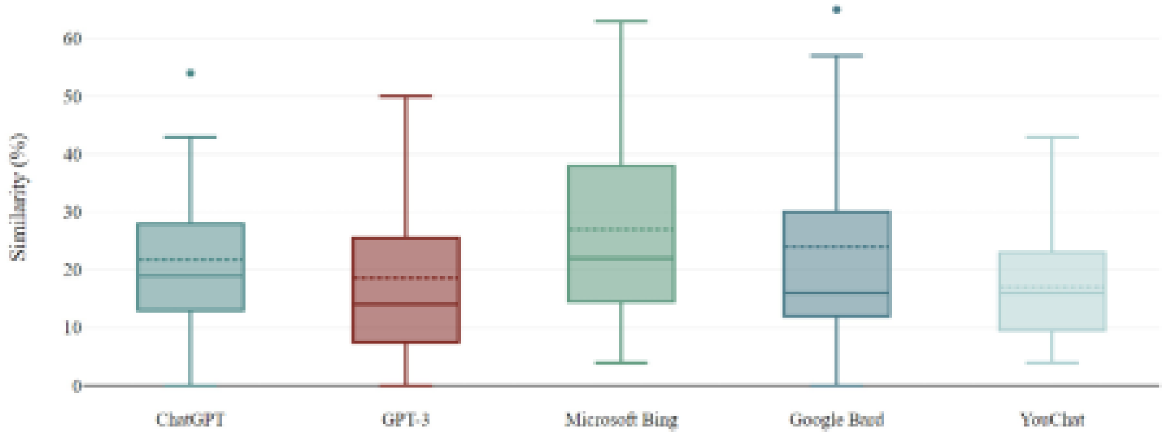

Figure 3.4: Similarity box and Whisker plots.

3.9 Term Frequency List Generation

- Once the closest words are identified, the next step is to look up their definitions in a dictionary. This helps understand the semantic context of the words. Words that are stop words or that lack significant meaning are removed. The remaining words are termed ”definition words”.
- The final output of the algorithm is the term frequency list, which contains the top-t words based on their occurrence frequency across the unknown images in the class. This list represents the most likely attributes of the unknown image.
- Example: If the word ”feline” appears most frequently, it would be ranked as the top word, followed by ”pet” and ”kitten.”

3.10 Generative AI Enhancement in GIRTARA

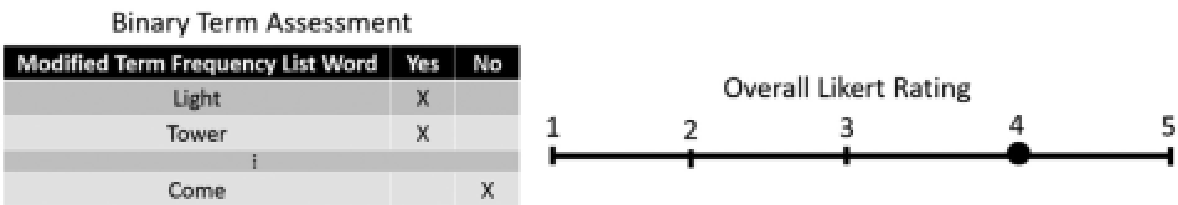

Figure 3.5: Example survey from Combs et al.

In GIRTARA, before the final evaluation, the top-5 words from the term frequency list are passed to a Generative AI model. This model refines the interpretations of the words by adding semantic context and improving the accuracy of the results. Example: If the words "feline" and "pet" are in the list, the Generative AI model might provide a more precise description such as "This is an image of a cat." The GIRTARA framework improves upon IRTARA by leveraging the power of Generative AI to refine the understanding of image data. By integrating word embeddings, analogical reasoning, and Generative AI, GIRTARA provides a more accurate and semantically rich interpretation of visual unknowns, overcoming the limitations of traditional image recognition models that rely solely on term frequency lists.This detailed methodology section explains each step of the process in GIRTARA and compares it to its predecessor, IRTARA, showing how it enhances the results.

Chapter 4

RESULTS AND DISCUSSION

4.1 Experimental Setup

Of significant interest to artificial intelligence (AI) research is to accurately interpret and describe out-of-library (true unknown) objects(Situ et al., 2016). Analogical reasoning has been proposed to assist with this end goal in both textual and visual scenarios; however,there has been limited research conducted regarding its application in computer vision problems. To answer our first research question, we present a literature review of image-based analogical reasoning and generative AI in Section 2. Related to research question2, we present the Image Recognition Through Analogical Reasoning Algorithm (IR-TARA), which integrates standard image classification methods from computer vision with the semantic meaning and interpretation from an analogical reasoning algorithm and dictionary initially described in Combs et al. (2023a,b).To answer research question 3, we proposed the"generative AI" version of the (GIRTARA), which is built upon the earlier IRTARA model but adds a generative AI module that takes the term frequency list and identifies a single object that the words

Table 4.1: Experimental environment setup

Resource	**Specification**
CPU	Intel® Core™ i-3-1005G1 @ 1.20 GHz
RAM	12 GB
GPU	Intel® UHD Graphics
Development Tool	Jupyter Notebook

4.2 Performance Evaluation

4.2.1 Performance Metrics

The classifiers were evaluated using standard metrics:

- Accuracy: $\frac{TP+TN}{TP+TN+FP+FN}$
- Precision: $\frac{TP}{TP+FP}$
- Recall: $\frac{TP}{TP+FN}$

- F1-score: $2 \times \frac{Precision \times Recall}{Precision + Recall}$
- Error Rate: $\frac{FP+FN}{TP+TN+FP+FN}$

  Positives (TP): Correct stroke predictions, Negatives (TN): Correct non-stroke predictions, False Positives (FP): Incorrect stroke predictions, False Negatives (FN): Missed stroke cases,

4.2.2 Pre-Optimization Results

Initial model performance before hyperparameter tuning:

Table 4.2: Types of generative AI models.

Type	**Creator**	**Model Name**	**Release date**	**Sources**
Language Models	OpenAI	GPT-1	Jun, 2018	Vaswani (2017)
Language Models	OpenAI	GPT-2	Nov, 2019	Radford et al. (2019)
Language Models	OpenAI	GPT-3	May, 2020	Brown et al. (2020)
Language Models	OpenAI	GPT-3.5	Mar, 2022	OpenAI (2022a,b)
Language Models	OpenAI	GPT-4	Mar, 2023	OpenAI (2023a,b)
Language Models	Google	LaMDA	May, 2021	Collins (2021)
Language Models	Google	LaMDA 2	May, 2022	Pichai (2022)
Language Models	Google	PaLM	Mar, 2023	Narang (2022)
Language Models	Google	PaLM 2	May, 2023	Google (2023)
Language Models	Meta AI	LLaMA 1	Feb, 2023	Meta AI (2023)
Language Models	Meta AI	LLaMA 2	Jul, 2023	Meta (2023)
Language Models	You.com	You	May, 2023	You.com
Language Models	Inflection AI	Infection-1	Jun, 2023	Inflection AI (2023)
Image Creators	OpenAI	DALL-E	Jan, 2021	OpenAI (2021)
Image Creators	OpenAI	DALL-E 2	Apr, 2022	OpenAI (2022a,b)
Image Creators	OpenAI	DALL-E 3	Oct, 2023	OpenAI
Image Creators	Craiyon	Craiyon	Jul, 2021	Dayma et al(2023)
Image Creators	Midjourney	Midjourney	Feb, 2022	Midjourney (2022)
Image Creators	Stability AI	Stable Diffusion	Aug, 2022	Stability.AI (2022)
Speech Recognition	OpenAI	Whisper	Sept, 2022	Radford et al. (2022)
Code Generators	OpenAI	Codex	Aug, 2021	Zaremba (2021)
Multimodal	OpenAI	ChatGPTPro	Oct, 2023	OpenAI
Multimodal	Google	Bard with Gemini	May, 2023	Gemini Team (2023)

4.2.3 Post-Optimization Results

Performance after hyperparameter tuning with GridSearchCV:

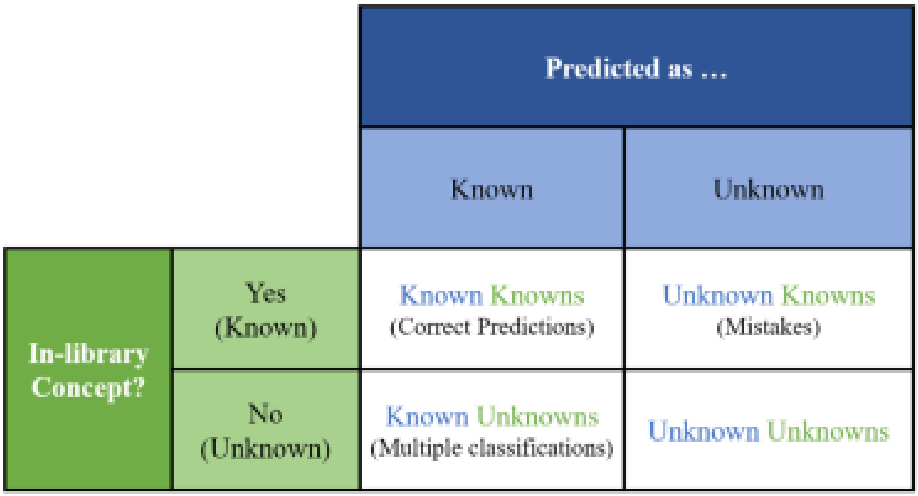

Figure 4.1: Known Unknown Evaluation

Table 4.3: Classifier performance after optimization

Model	Precision	Recall	F1	Accuracy	Error
SVM	99.99%	99.99%	99.99%	99.99%	0.00001%
Random Forest	99.85%	99.88%	99.86%	99.87%	0.001%
KNN	98.66%	98.97%	98.81%	98.82%	0.01%
Decision Tree	96.63%	97.27%	96.86%	96.90%	0.03%
Naïve Bayes	74.26%	74.39%	74.32%	74.77%	0.25%
Logistic Regression	77.07%	77.32%	77.17%	77.53%	0.22%
AdaBoost	78.11%	78.39%	78.21%	78.55%	0.21%
Gradient Boosting	80.82%	81.29%	80.95%	81.18%	0.19%
MLP	79.58%	80.05%	79.71%	79.94%	0.20%
Nearest Centroid	67.78%	66.21%	66.36%	68.22%	0.32%
Voting Classifier	92.33%	87.95%	89.00%	89.67%	0.10%

4.3 Limitations and Future Work

- Current limitations:
  - Bias in Generative Models: Bias: Generative models can reflect biases from their training data, leading to harmful stereotypes.
  - Data Dependency: The models require vast, high-quality data, which may not always be available.
  - Large models require substantial computational power, making them less accessible.
- Future improvements:

Table 4.4: Top-5 words produced by IRTARA for each true unknown

Unknown	**1**	**2**	**3**	**4**	**5**
AK-47	Long	Small	Move	Person	Played
Chandelier	Small	Observe	Person	Determine	Light
Fireworks	Large	Small	Long	Cloud	Light
Floppy disk	Small	Ball	Body	Long	Device
Frog	Large	Body	Fungi	Small	Edible
Galaxy	Planet	Sun	Mythology	Tha	Small
Iguana	Long	Small	Large	Coat	Genus
Mars	Brain	Skull	Nervous	Ability	Planet
People	Large	Body	Ball	Move	Small
Rainbow	Light	Little	Illumination	Fire	United
Sheet Music	Small	Rectangular	Area	Glass	Box
Skyscraper	Light	Tower	Small	Building	Little
Swiss Army Knife	Ball	Instrument	Body	Device	Ball
T-shirt	Light	Game	Face	Small	Fungi
Waterfall	Large	Fleshy	Body	Edible	

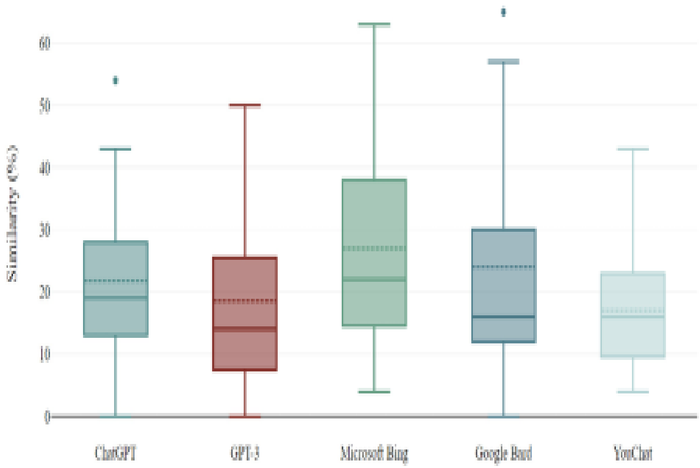

Figure 4.2: Similarity box and whisker plots

- Bias Mitigation: Develop methods to reduce bias and create more diverse training datasets.

- Domain-Specific Fine-Tuning: Focus on fine-tuning models for specialized fields like law or healthcare.

- Explainability: Improve model transparency and make outputs more understandable.

- Human-AI Collaboration: Enable better interaction between humans and AI for improved content generation.

Chapter 5

CONCLUSION

In summary, generative AI is a powerful tool that effectively creates context and transforms data into insights. Unlike traditional methods, the key advantage of IRTARA and GIRTARA lies in their ability to provide a semantic understanding of visual unknowns through analogical reasoning, without requiring prior training on observed concepts. The evaluation of IRTARA and GIRTARA using the Caltech-256 dataset demonstrated that the analogical reasoning-based evaluation method showed a higher correlation with human judgment compared to the definition-based evaluation method.

Additionally, the highest cosine similarity between a prediction and a true unknown class was 65highlighting the potential of generative AI in identifying unknown objects. Several directions can further enhance the effectiveness of generative AI models in this domain. First, considering multiple runs of experiments rather than a one-time test may provide a more accurate assessment of model performance. Additionally, different prompt variations can yield different results, so refining prompts for generative AI models can improve predictions. Expanding the number of generative AI models analyzed is another avenue, as new models continue to emerge and improve. Moreover, tweaking the IRTARA and GIRTARA parameters could lead to more representative term frequency lists, improving generative AI predictions.

Finally, integrating a vision model alongside text-based natural language processing could enhance both IRTARA's image-based reasoning and GIRTARA's textual predictions, making the system more robust in handling multimodal problems.

REFERENCES

[1] **Aonvolutional Neural Networks (CNN) in Image Processing**, "Convolutional Neural Networks (CNN) in Image Processing," 2023.

- CNNs significantly improve image recognition and processing tasks by extracting spatial hierarchies from images, but struggle with recognizing truly unknown objects.

[2] **Combs et al., 2022**, "Analogical Reasoning in AI," 2022.

- Explores the use of analogical reasoning in text-based domains and the evolution of methods from psychology-based models to neural network-based models like Word2Vec, GloVe, and fastText.

[3] **Ichien et al., 2020**, "Analogical Reasoning in AI," 2020.

- Investigates the role of analogical reasoning in AI, especially in text-based domains, and how neural network-based methods have evolved.

[4] **Gentner, 1983**, "Analogical Reasoning in AI," 1983.

- Focuses on mapping relations between words and analyzing sentence structures for relational meaning in analogical reasoning.

[5] **Holyoak and Thagard, 1989**, "Analogical Reasoning in AI," 1989.

- Discusses how analogical reasoning has been primarily explored in word-based relationships, with a focus on sentence structure and relational meaning.

[6] **Hofstadter and Mitchell, 2022**, "Analogical Reasoning in AI," 2022.

- Explores the theoretical aspects of analogical reasoning in AI and its application in text-based domains.

[7] **Evans, 1964**, "Visual Analogical Reasoning," 1964.

- Discusses early work on visual analogical reasoning, particularly in geometric pattern-based tasks like Raven's Progressive Matrices.

[8] **Raven and Court, 1938**, "Visual Analogical Reasoning," 1938.

- Investigates geometric patterns and their use in visual analogical reasoning, focusing on tasks like Raven's Progressive Matrices.

[9] **Sadeghi et al., 2015**, "Visual Analogical Reasoning," 2015.

- Focuses on algorithms like Visalogy that attempt to apply analogy reasoning to object attributes like color and shape.

[10] **Lu et al., 2019a**, "Visual Analogical Reasoning," 2019.

- Explores the challenges of applying visual analogical reasoning to action, orientation, or semantic context.

[11] **Bihl and Talbert, 2020**, "Generative AI in Various Domains," 2020.

- Examines the rapid expansion of generative AI, particularly in fields like healthcare, education, and research.

[12] **Doshi et al., 2023**, "Generative AI in Various Domains," 2023.

- Discusses the impact of generative AI in various domains, including its ability to assist in advanced contextual understanding and applications like diagnosis assistance and automated content generation.

[13] **Singh, 2023**, "Generative AI in Various Domains," 2023.

- Explores generative AI's potential in applications such as predictive modeling and diagnosis assistance.

[14] **Raimondi et al., 2023**, "Generative AI in Various Domains," 2023.

- Analyzes the role of generative AI in enhancing knowledge processing across various fields.

[15] **Eggmann et al., 2023**, "Generative AI in Various Domains," 2023.

- Discusses the expanding role of generative AI in diverse fields and its potential applications in knowledge processing.

[16] **Chen et al., 2021**, "Generative AI in Programming and Coding," 2021.

- Highlights the rise of AI-driven code generation tools like Codex and GitHub Copilot, which leverage large language models (LLMs) to assist developers in coding tasks.

[17] **Dohmke, 2023**, "Generative AI in Programming and Coding," 2023.

- Explores how AI-driven tools like Codex and GitHub Copilot are improving coding efficiency by providing autocomplete and refactoring suggestions.

[18] **OpenAI, 2023a,b,c,d,e,f**, "Generative AI in Programming and Coding," 2023.

- Discusses the evolution of Codex and other AI-driven code generation tools and their impact on coding tasks.

[19] **Webb et al., 2023**, "Large Language Models (LLMs) and Analogical Reasoning," 2023.

- Examines the analogical reasoning abilities of LLMs, identifying both their strengths and weaknesses in complex or multimodal contexts.

[20] **Yu et al., 2023**, "Large Language Models (LLMs) and Analogical Reasoning," 2023.

- Investigates the limitations of LLMs in performing deeper conceptual reasoning, especially in multimodal contexts.

[21] **Mitchell et al., 2023**, "Large Language Models (LLMs) and Analogical Reasoning," 2023.

- Discusses challenges faced by LLMs in integrating text and images for analogical reasoning.