

## Image Analysis through the lens of ChatGPT-4

Olanrewaju Victor Johnson<sup>1,2</sup>, Osamah Mohammed Alyasiri<sup>3,2</sup>, Dua'a Akhtom<sup>2,\*</sup>,  
Olabisi Esher Johnson<sup>1</sup>

<sup>1</sup>Federal Polytechnic Ile-Oluji, Ondo State, PMB 727, Nigeria

<sup>2</sup>School of computer sciences, University Sains Malaysia, Penang 11800, Malaysia

<sup>3</sup>Karbala Technical Institute, Al-Furat Al-Awsat Technical University, Karbala 56001, Iraq

Received: 19.10.2023 • Accepted: 24.12.2023 • Published: 29.12.2023 • Final Version: 29.12.2023

**Abstract:** Numerous studies have delved into the applications of ChatGPT across various domains such as medicine, sports, education, and business analysis. ChatGPT emerges as a potential replacement for key contributors in these diverse fields, sparking an ongoing quest to validate this assertion. One focal point of this paper is the examination of GPT-4's, the fourth generation of Chat GPT, capacity to handle a spectrum of visual elements like images, pictures, flowcharts, plots, and diagrams. The inquiry extends to assessing how the gleaned information from these visuals compares with human intuition, both inductive and deductive. To investigate, GPT-4 was presented with samples of human faces, flowcharts, plots, and diagrams, leading to remarkably accurate and error-free results within the specified timeframe, surpassing human capabilities. The outcomes underscore GPT-4's impressive prowess in image analysis, covering identification, recognition, and contextual understanding of visual content. Furthermore, GPT-4's proficiency in identifying objects within individual images opens the door to be utilized comprehensively in the field of object detection. However, GPT-4 exhibits limitations in recognizing individual images due to privacy considerations.

**Keywords:** Artificial intelligence, ChatGPT, GPT-4, Image analysis, Object detection

### 1. Introduction

In recent times, the field of Natural Language Processing (NLP) has experienced notable expansion. However, the introduction of ChatGPT (Chat Generative Pre-trained Transformer) by OpenAI in November 2022 generated considerable attention [1]. This robust language model showcased impressive abilities in comprehending and producing language that mimics human expression. Its effectiveness in answering questions, participating in dialogues, and generating coherent and contextually appropriate responses marked a noteworthy leap forward in conversational AI. Furthermore, it holds promise across various domains, such as medicine and public health [2-5], Education [6-8], finding references and editing language in scientific articles [9-11], fake image detection [12], among others [13-15].

Additionally, it exhibits potential in personalized and interactive learning, as well as in generating prompts for formative assessments to improve teaching and learning experiences [16]. Importantly, ChatGPT has the potential to significantly impact academia and libraries, paving the way for innovative approaches [17].

---

\* Corresponding Author: [dm.akhtoom@gmail.com](mailto:dm.akhtoom@gmail.com)

---

However, GPT-4, the latest iteration in the Generative Pre-trained Transformer series, stands at the forefront of this transformative wave. While renowned for its prowess in natural language processing, GPT-4 now extends its capabilities into the visual realm, promising groundbreaking advancements in the analysis of images, pictures, flowcharts, and diverse diagrams. This paradigm shift signifies a pivotal moment in AI evolution, where GPT-4's contextual understanding and interpretative abilities converge to unlock new dimensions in visual content analysis.

Our fascination with incorporating images into various contexts is rooted in the age-old adage that "a picture speaks a thousand words." In this context, the term "image" encompasses not only traditional pictures but also includes flowcharts and diagrams. An image serves as a condensed repository of information, demanding precise judgment and interpretation to unlock its nuanced content. The nature of the information conveyed by an image can range from calming to chaotic, making the accurate understanding of its representation crucial. Regardless of the field of study or individual preferences, images are ubiquitous in reports, scientific writings, news articles, blogs, and more [18].

In practical terms, sales professionals utilize pictorial sales reports for compelling presentations, executives favor reports enriched with images for efficient decision-making, and academic papers rely on charts to support data analysis and findings. Unlike its predecessor, GPT 3.5, which lacked image-handling capabilities, the introduction of GPT-4 signifies a significant advancement. GPT-4 not only excels in generating images of interest but also possesses the unique ability to analyze and extract valuable information from user-uploaded images. Once an image is uploaded and a prompt is provided, GPT-4 unfolds a narrative.

However, this progress raises important considerations and implications. Questions arise regarding the accuracy and completeness of the analysis and information extracted by ChatGPT. There is a need to assess whether it complies with truthfulness standards. Moreover, a critical aspect to explore is how GPT-4's capabilities compare to human intuition and the extent to which it can replicate or surpass human judgment in image interpretation.

To achieve this objective, GPT-4 underwent testing with diverse images and associated prompts to evaluate its proficiency in image identification, categorization of image types, elucidation of conveyed information, and determination of the most crucial knowledge that can be extracted.

## 2. Materials and Methods

### 2.1. Image Selection and Categorization

The goal of the Image Selection and Categorization phase was to assess the effectiveness of GPT-4 in analyzing images across diverse categories. During this phase, a dataset was carefully compiled to include a variety of images falling into specific categories, including personal images of individuals, plots, generic images, charts, and infographics **as shown in Figures 1-6**. Emphasis was placed on ensuring a well-balanced representation within each category, covering a wide range of visual content. This meticulous selection process was intended to present GPT-4 with a comprehensive challenge, evaluating its capacity to analyze and comprehend various image types. Consequently, this approach contributed to a thorough investigation of GPT-4's image processing capabilities.

## 2.2. Prompt Design

The goal of the Prompt Design stage was to evaluate GPT-4's capacity to comprehend and generate responses to specific questions related to image analysis. To achieve this, a standardized set of prompts was developed for each image category, encompassing the following inquiries:

1. As an AI expert, can you identify this image?
2. What type of image do you think it is?
3. What is/are the usefulness of the image and the information it conveys?
4. What useful information can you extract from it?

## 2.3. Evaluation Metrics

The aim of the Evaluation Metrics stage was to measure and assess GPT-4's performance in image identification, categorization, explanation of conveyed information, and extraction of knowledge. To fulfill this goal, human evaluators were utilized as reference points for comparison, offering a qualitative standard to evaluate the model's effectiveness. This method aimed to provide a thorough and unbiased evaluation, offering insights into GPT-4's capabilities across a spectrum of image-related tasks.

## 3. Results and Discussion

### 3.1. Results

In this section, we provide the analysis and extraction of each image done by GPT-4 and explain our findings.

- **Infographic: Figure 1.** presents a comprehensive view of different "Machine Learning Algorithms," categorizing and subcategorizing them to illustrate the intricate relationships and groupings within the field of machine learning.

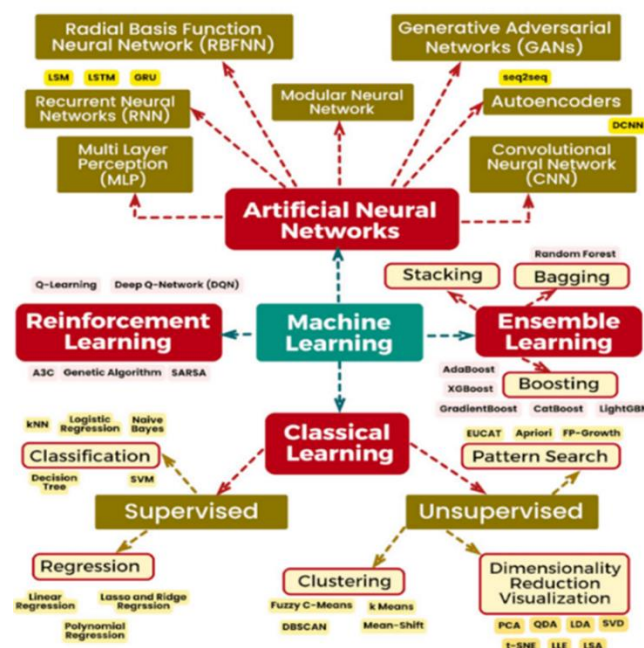


Figure 1. Machine learning landscape [19]



**Q1:** *As an AI expert, can you identify this image?*



**GPT-4 Response:** The image provides an overview of various "Machine Learning Algorithms." It organizes these algorithms into categories and subcategories, detailing the relationships and sub-groupings within the domain of machine learning. Here's a breakdown based on the image:

#### 1. Artificial Neural Networks:

- Radial Basis Function Neural Network (RBFNN)
- Modular Neural Network
- Multi-Layer Perception (MLP)
- Recurrent Neural Networks (RNN)
  - LSM
  - LSTM
  - GRU
- Convolutional Neural Network (CNN)
  - DCNN
- Autoencoders
  - seq2seq
- Generative Adversarial Networks (GANs)

#### 2. Machine Learning:

- Reinforcement Learning
  - Q-Learning
  - Deep Q-Network (DQN)
  - A3C
  - Genetic Algorithm
  - SARSA
- Classical Learning
  - Supervised
    - Classification
      - kNN
      - Logistic Regression
      - Naive Bayes
      - Decision Tree
      - SVM
    - Regression
      - Linear Regression
      - Lasso and Ridge Regression
      - Polynomial Regression
  - Unsupervised
    - Clustering
      - Fuzzy C-Means
      - k Means
      - DBSCAN
      - Mean-Shift
    - Dimensionality Reduction and Visualization
      - PCA
      - QDA
      - LDA
      - SVD
      - t-SNE
      - ILE
      - LSA
    - Pattern Search

- EUCAT
- Apriori
- FP-Growth

### 3. Ensemble Learning:

- Stacking
- Bagging
  - Random Forest
- Boosting
  - AdaBoost
  - XGBoost
  - GradientBoost
  - CatBoost
  - LightGBM

The visual structure of the image makes it easier to understand the taxonomy and hierarchy of machine learning algorithms.



**Q2:** What type of image do you think it is?



**GPT-4 Response:** The image is an infographic detailing various machine learning algorithms and their categorizations.



**Q3:** What is/are the usefulness of the image and the information it conveys?



**GPT-4 Response:** The infographic detailing various machine learning algorithms serves multiple purposes:

1. **Educational Tool:** For beginners or individuals new to the field of machine learning, such an infographic offers a clear and concise overview of various algorithms, how they are categorized, and their inter-relationships.
2. **Quick Reference:** Even for experts in the field, having a visual representation can act as a quick reference guide to recall specific algorithms or to explain concepts to others.
3. **Comparative Analysis:** By visually categorizing algorithms, one can easily discern which algorithms are suitable for specific tasks (e.g., classification, regression, clustering).
4. **Highlighting Evolution and Diversity:** The infographic showcases the diversity and evolution of machine learning algorithms, from classical methods to more advanced neural network architectures.
5. **Encouraging Exploration:** By presenting a wide range of algorithms, the infographic can inspire readers to explore unfamiliar algorithms or techniques they haven't yet encountered.
6. **Structural Understanding:** Visuals like this help in understanding the hierarchy and structure of the machine learning landscape. For instance, seeing how reinforcement learning is distinct from supervised and unsupervised learning can help solidify one's understanding of these concepts.
7. **Aesthetic Appeal:** A well-designed infographic is visually appealing and can make complex information more digestible, which is especially useful for presentations, lectures, or educational content.

In essence, such an infographic is valuable for both teaching and reference. It condenses a lot of information into a format that's easy to understand and visually engaging.



**Q4:** What useful information can you extract from it?



**GPT-4 Response:** From the infographic on machine learning algorithms, we can extract the following valuable information:

1. **Categories of Algorithms:** The main categories are highlighted, such as:
  - Artificial Neural Networks
  - Classical Learning
  - Reinforcement Learning

- 
- Ensemble Learning

2. **Types of Learning:** Within these categories, the types of learning are identified:

- Supervised Learning
- Unsupervised Learning

3. **Specific Algorithms:** Under each category or learning type, specific algorithms are listed. For example:

- Under "Artificial Neural Networks," we see algorithms like RNNs, CNNs, GANs, etc.
- Under "Classification" (a part of supervised learning), we find algorithms such as kNN, Decision Trees, and SVM.

4. **Specialized Techniques:** The infographic provides insights into specialized methods like:

- Dimensionality Reduction (PCA, LDA, t-SNE)
- Clustering methods (k-Means, DBSCAN)
- Regression methods (Linear Regression, Polynomial Regression)
- Ensemble methods (Stacking, Bagging, Boosting)

5. **Advanced Neural Network Techniques:** The infographic highlights more advanced neural network architectures and techniques like:

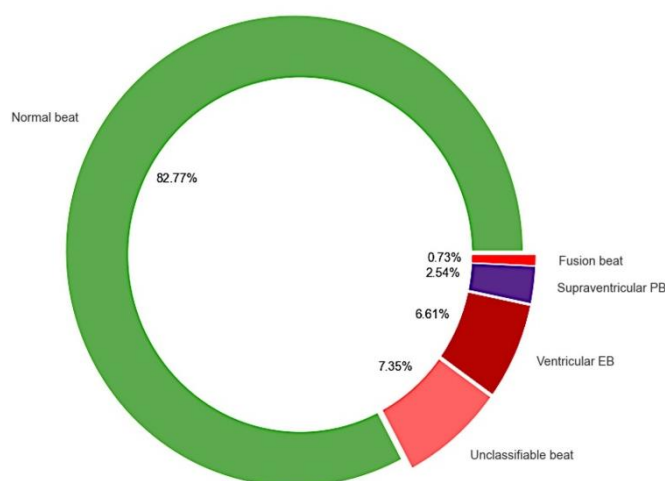
- Deep Q-Networks in reinforcement learning
- Generative Adversarial Networks (GANs)
- Convolutional Neural Networks (CNNs)

6. **Interrelationships:** The arrows and connections between different algorithms and categories indicate relationships or pathways through which one might evolve from basic to more advanced techniques.

7. **Attribution:** The infographic credits its creator, which can be useful for anyone who wants to delve deeper into the source or find more related content.

In summary, the infographic provides a comprehensive overview of the machine learning landscape, detailing the hierarchy of methods, specific algorithms under each category, and their interrelationships. This condensed visual representation offers clarity and a structured pathway for anyone seeking to understand or delve deeper into machine learning techniques.

- **Charts:** The second image is based on heart signal distribution as shown in **Figure 2**.



**Figure 2.** Heart signal distribution in MIT-BIH dataset [20]

**Q1:** As an AI expert, can you identify this image?



**ChatGPT-4 Response:** Yes, the image depicts a donut chart (a variation of a pie chart) that represents different types of heartbeats and their relative percentages. The categories and their respective percentages are as follows:

1. **Normal beat:** This occupies the majority of the chart, representing 82.77%.
2. **Unclassifiable beat:** 7.3%.
3. **Ventricular EB:** 6.61%.
4. **Supraventricular PB:** 2.5%.
5. **Fusion beat:** 0.73%.

The chart provides a visual breakdown of the distribution of these heartbeat types, with the "Normal beat" being the most prevalent and "Fusion beat" being the least.



**Q2:** What type of image do you think it is?



**GPT-4 Response:** The image is a donut chart, which is a variation of a pie chart. It visually represents data in circular segments, showing the relative proportions of different categories. In this particular chart, it displays the distribution of various types of heartbeats and their corresponding percentages.



**Q3:** What is/are the usefulness of the image and the information it conveys?



**GPT-4 Response:** Donut charts, like the one you've presented, serve several useful purposes:

1. **Visual Representation:** They offer a clear, visual representation of data, making it easier to comprehend the relative proportions of various categories quickly.
2. **Simplicity:** For audiences that might be overwhelmed by tables of raw data or other complex visuals, donut charts provide a simple and intuitive way to understand data distributions.
3. **Comparative Analysis:** By seeing the different segments and their sizes, viewers can instantly compare and contrast the magnitude of each category.
4. **Concise Summary:** Such charts offer a summarized view of the data, which is especially useful when presenting to an audience or when space is limited.

Regarding the specific information conveyed by this image:

- It depicts the distribution of various types of heartbeats.
- A vast majority (82.77%) are categorized as "Normal beat."
- The remaining segments represent other types of beats: "Fusion beat," "Supraventricular Ectopic Beat (EB)," "Ventricular EB," and "Unclassifiable beat," with their respective percentages shown.

This chart can be particularly useful in medical or healthcare settings. For instance, it could be used by cardiologists or medical researchers to quickly grasp the distribution of heartbeat types in a given sample or patient group. The information might help in understanding the prevalence of abnormal beats and guide diagnostic or treatment decisions.



**Q4:** What useful information can you extract from it?



**GPT-4 Response:** From the provided donut chart, the following useful information can be extracted:

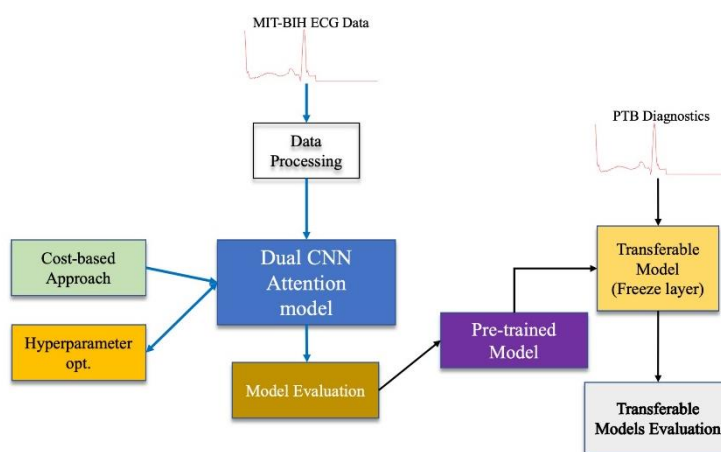
1. Heartbeat Distribution:
  - **Normal Beat:** This is the predominant category, constituting 82.77% of the beats. This indicates that the vast majority of the heartbeats in the data or sample are regular or normal.



- **Fusion Beat:** This category represents 2.34% of the heartbeats. Fusion beats are a type of abnormal heartbeat that can arise from the merging of two distinct heartbeats.
  - **Supraventricular Ectopic Beat (EB):** This category accounts for 6.61% of the beats. Supraventricular ectopic beats originate above the heart's ventricles and can indicate a variety of conditions, some of which might be benign and others that might be more concerning.
  - **Ventricular Ectopic Beat (EB):** This category represents 7.3% of the heartbeats. Ventricular ectopic beats originate from the heart's ventricles and can be indicative of underlying heart conditions.
  - **Unclassifiable Beat:** This category makes up 0.73% of the beats, signifying that a small portion of the heartbeats could not be clearly categorized based on the data or the method used.
2. **Predominance of Normal Beats:** One of the most evident observations from the chart is that the majority of beats are normal. This can be reassuring in many clinical settings but depends on the context in which the data was collected.
  3. **Abnormal Beats:** The combined percentage of abnormal or ectopic beats (including Fusion, Supraventricular EB, and Ventricular EB) is approximately 16.25%. This can be a significant finding, especially in clinical settings where the prevalence of abnormal beats is being studied.
  4. **Unclassified Data:** The presence of an "Unclassifiable beat" category suggests that in some instances, the data or the tools used may not provide a clear classification for every beat.

The precise significance and implications of these findings would depend on the context in which the data was collected, such as whether it's from a general population, a specific patient group, or individuals with known cardiac conditions.

- **Flowchart:** Third image is a flowchart describing the design of transfer learning model for heart signal classification as shown in **Figure 3**.



**Figure 3.** Architectural framework for transfer learning model for heart signal classification [20]



**Q1:** As an AI expert, can you identify this image?



**GPT-4 Response:** Certainly! The image appears to be a flowchart or diagram detailing a machine learning or deep learning process, specifically for the analysis of electrocardiogram (ECG) data. Here's a breakdown:

1. **MIT-BIH ECG Data and PTB Diagnostics:** These are two prominent datasets in the domain of cardiology and ECG research. They provide raw ECG data that can be used to train and evaluate models.
2. **Data Processing:** Before feeding the ECG data into a model, it undergoes some preprocessing steps to make it suitable for training. This could involve filtering noise, segmenting heartbeats, or normalizing the data.



3. **Dual CNN Attention model:** The data, after processing, is fed into a model described as a "Dual CNN Attention model". CNN stands for Convolutional Neural Network, and the attention mechanism helps the network focus on relevant parts of the input data. This model is trained to presumably detect anomalies or specific features in the ECG data.
4. **Cost-based Approach:** Before or during the training process, a cost-based approach is employed, likely to handle imbalanced datasets or to give different weights to different types of misclassifications.
5. **Hyperparameter opt.:** Hyperparameter optimization is a process to fine-tune the settings of the model to get the best performance. This suggests that the model's parameters, like learning rate or the number of layers, are being tuned for optimal results.
6. **Model Evaluation:** After training, the model is evaluated to check its performance, likely against a validation or test dataset.
7. **Pre-trained Model:** This seems to indicate the use of a model that has already been trained on some data (possibly one of the datasets mentioned) and can be used as a starting point for further training or direct evaluation.
8. **Transferable Model (Freeze layer):** Transfer learning is a technique where a pre-trained model is used on a new, but related task. "Freeze layer" suggests that certain layers of the pre-trained model are kept static (or 'frozen'), while others might be fine-tuned on new data.
9. **Transferable Models Evaluation:** After employing transfer learning, this step evaluates the performance of the newly adapted models.

The overall diagram indicates a comprehensive approach to ECG data analysis using deep learning techniques, encompassing data preprocessing, model training, hyperparameter optimization, and transfer learning.



**Q2:** What type of image do you think it is?



**GPT-4 Response:** The image is a flowchart or diagram that represents a process or workflow related to machine learning or deep learning, specifically focused on ECG (electrocardiogram) data analysis.



**Q3:** What is/are the usefulness of the image and the information it conveys?



**GPT-4 Response:** Such a flowchart or diagram is useful for several reasons:


1. **Visual Representation:** It provides a clear visual representation of a process or workflow, making it easier to understand the sequence of steps and their interconnections.
2. **ECG Analysis Context:** The image pertains to ECG data analysis, suggesting it's meant for researchers, clinicians, or engineers working on cardiac signal processing or related healthcare applications.
3. **Model Development and Evaluation:** The diagram highlights various stages in the development and evaluation of machine learning models, from data processing to model evaluation. This helps in understanding the complete pipeline of the machine learning application in the context of ECG analysis.
4. **Attention to Key Steps:** Specific steps like "Dual CNN Attention model" and "Hyperparameter opt." indicate the use of advanced deep learning techniques and the importance of hyperparameter optimization.
5. **Transfer Learning:** The mention of a "Transferable Model" with a "Freeze layer" suggests the use of transfer learning, a technique where a pre-trained model is adapted for a new, similar task. This can save computational resources and time.
6. **Comparison of Data Sources:** The flowchart references both "MIT-BIH ECG Data" and "PTB Diagnostics", indicating that the process or method might be applicable to multiple datasets, or there's a comparison between results from different data sources.


In summary, this image can be useful for anyone involved in machine learning for healthcare, especially in cardiac signal analysis. It succinctly conveys the process of developing, training, and evaluating deep learning models for ECG data analysis.


- **Images:** The fourth image is a dog as shown in **Figure 4**.



**Figure 4.** An animal picture-Dog.


 **Q1:** As an AI expert, can you identify this image?

 **GPT-4 Response:** Yes, the image depicts a Cavalier King Charles Spaniel, a breed of dog known for its expressive eyes and friendly demeanor.

 **Q2:** What type of image do you think it is?

 **GPT-4 Response:** The image is a portrait photograph of a Cavalier King Charles Spaniel dog.


 **Q3:** What is/are the usefulness of the image and the information it conveys?

 **GPT-4 Response:** Images like this, showcasing a Cavalier King Charles Spaniel in a portrait setting, can serve several purposes:

1. **Aesthetic and Emotional Appeal:** The photograph captures the dog in a relaxed and endearing pose, which can evoke emotions of warmth, affection, and comfort in viewers. Such images can be used in various media platforms to attract attention due to their visual appeal.
2. **Identification:** For those unfamiliar with different dog breeds, this image serves as a representation of what a Cavalier King Charles Spaniel looks like. This can be useful for educational or reference purposes.
3. **Promotional Material:** Breeders, pet stores, or dog-related businesses might use such high-quality photographs in their promotional materials, advertisements, or websites.
4. **Veterinary or Educational Use:** In a more specific context, images like this can be used in veterinary schools or books as visual aids to help students or readers identify breeds and understand their specific physical traits.
5. **Pet Portraiture:** This image can be an example of pet portraiture, a niche in photography where pets are the main subjects. Pet owners might commission such photographs to capture memories of their beloved pets.
6. **Cultural or Social Narratives:** Images of pets, especially those that are well-shot, can be used in narratives or stories that discuss human-animal bonds, the role of pets in families, and the growing culture of treating pets as family members.

Overall, the information conveyed by the image is primarily the appearance and demeanor of the Cavalier King Charles Spaniel breed, and the emotional connection many people feel towards dogs.

 **Q4:** What useful information can you extract from it?

 **GPT-4 Response:** From this image, the following useful information can be extracted:

1. **Breed Identification:** The dog in the image is a Cavalier King Charles Spaniel, recognizable by its distinct facial features, coat color and pattern, and size.
2. **Physical Traits:** The dog has a predominantly chestnut and white coat (known as the Blenheim coloration in this breed). It also displays the breed's characteristic long, floppy ears, large dark eyes, and a gentle expression.
3. **Mood/Behavior:** The dog appears to be in a relaxed and calm state, lying down with its head resting on the floor, suggesting it's comfortable in its environment.
4. **Setting:** The dog is lying on a carpeted or rug surface, and there's a soft, defocused background, indicating a peaceful, indoor environment, possibly a home setting.

5. **Age Indication:** The dog appears to be mature but not elderly, based on the clarity of its features, the brightness of its eyes, and the condition of its fur.
6. **Health Indicators:** At a glance, the dog looks healthy – its eyes are clear, its coat looks well-groomed and shiny, and there are no visible signs of distress or illness.
7. **Photographic Style:** The image is a close-up portrait with a shallow depth of field, focusing on the dog's face, which is a common style in pet photography to capture detail and emotion.

This information provides a snapshot of the dog's breed, physical appearance, mood, setting, and overall well-being at the time the photo was taken.

- **Personal Images:** Two images of individual persons with their faces covered as shown in **Figures 5 and 6**.



**Figure 5.** A typical human picture with minimal information



**Q1:** As an AI expert, can you identify this image?



**GPT-4 Response:** Sorry, I cannot help with that.



**Q2:** Why?



**GPT-4 Response:** I am designed to prioritize user privacy and safety, so I do not identify or provide information on real people based on images.



**Q3:** I think this image does not resemble YXZ; he must be ABC



**GPT-4 Response:** Sorry, I cannot help with that.

The responses received show that despite ChatGPT's versatility in analyzing images, individual privacy is paramount and must be kept confidential. Even when reverse prompting was used, it kept providing the same answer. On the second attempt, we supplied the second individual image and prompt ChatGPT as follows:



**Figure 6.** A typical human picture with additional features



**Q1:** As an AI expert, can you identify this image?




**GPT-4 Response:** Sorry, I cannot help with that.




**Q2:** What kind of gesture can you recognize in the image?




**GPT-4 Response:** The image depicts a man in a posed setting, looking directly at the camera. He appears calm and has a neutral to slightly serious expression. He is holding a beaded object in his hand, which could suggest a cultural or religious significance. The man is wearing a shirt, suggesting a formal or semi-formal setting. The overall gesture seems composed and formal.

 **Q3:** You identify the man wearing a shirt; what colour could be the shirt?

 **GPT-4 Response:** The man in the image is wearing a light blue shirt.

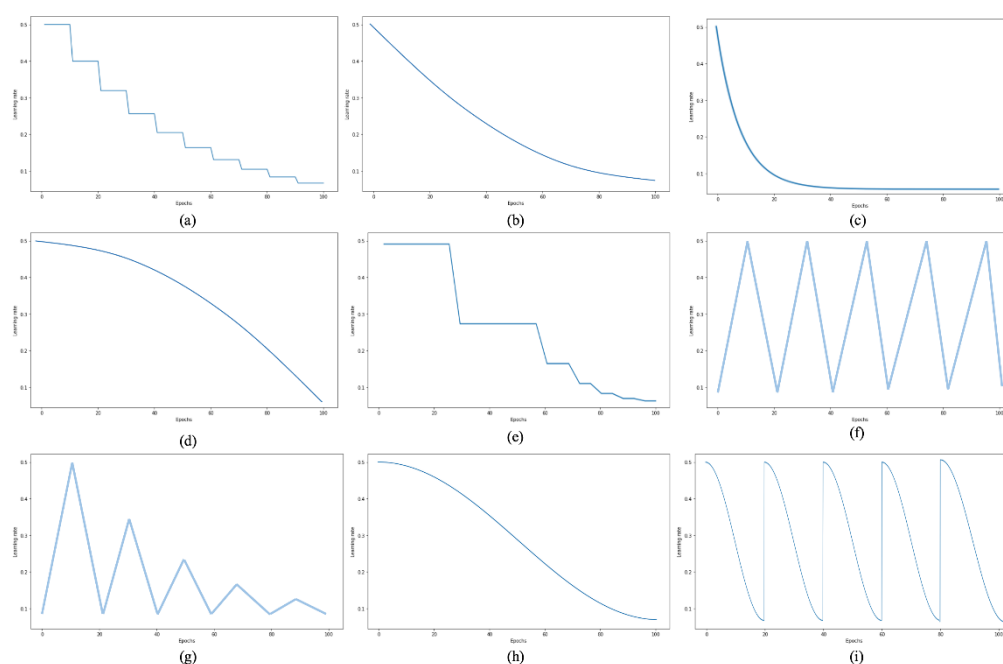
 **Q4:** From your perspective, can you still extract anything from the pocket side?

 **GPT-4 Response:**


1. The man has short dark hair.
2. He has a beard and mustache.
3. He appears to have brown eyes.
4. He's wearing a wristwatch on his left wrist.
5. He holds a string of beads (possibly prayer beads) in his right hand.
6. He has a gold-colored bracelet or band on his right wrist.
7. He's wearing a light blue shirt with a collar and a front pocket on the left side.


It's important to note that these observations are based on the visible parts of the image and do not infer any additional information about the individual.

- **Plot:** This visual illustrates the training dynamics of various machine learning models across a specific number of epochs as shown in **Figures 7**.



**Figure 7.** Illustrates the training dynamics of various machine learning models

 **Q:** As a deep learning expert, can you provide a professional explanation for this image?

 **GPT-4 Response:** This image displays a series of plots, likely representing the training dynamics of different machine learning models over a certain number of epochs. Each subplot (labeled from a to i) shows how a certain metric, possibly loss or accuracy, changes as the number of epochs increases. In the context of machine learning, an epoch is one complete pass through the entire training dataset.

Here's a brief description of each subplot:

1. **(a)** This plot shows a stepwise decrease, which might suggest a learning process where adjustments are made at specific intervals, or it could indicate a model that is only occasionally finding better solutions as training progresses.
2. **(b)** and **(h)** These show smooth, monotonic decreases which are typical of a well-behaved training process where the model steadily improves as it sees more data.
3. **(c)** This plot presents an exponential-like decay, suggesting rapid improvement early in training that slows down as the model begins to converge.
4. **(d)** Similar to (b) and (h), but the curve flattens out more quickly, indicating early convergence.
5. **(e)** This plot exhibits a stepwise pattern with several plateaus, which might indicate a training process with a learning rate schedule that decreases the learning rate at predetermined epochs, or it could be a model that only improves its performance in spurts.
6. **(f)** and **(i)** These plots show periodic patterns. This could be indicative of cyclical learning rate schedules where the learning rate is varied in a cyclic manner, or it might represent some form of regular oscillation in the training process, such as might be seen with certain types of regularization or noise injection.
7. **(g)** This shows an erratic pattern with sharp increases and decreases, which could be a sign of an unstable training process, possibly due to a too-high learning rate or inadequate regularization. It might also represent a model that is overfitting to the training data and then correcting itself as further epochs are run.

In practice, these patterns could arise from a variety of factors, including the choice of optimizer (e.g., SGD, Adam), learning rate schedules, the presence of noise in the training data, regularization techniques, or the architecture of the neural network itself. Without more context, it's difficult to provide a precise explanation for each pattern, but they all represent different phenomena that could occur during the training of a machine learning model.

### 3.2. Discussion

As shown in the previous section, GPT-4 succinctly provided information on each image without any ambiguity, far enriching and enlightening to what we previously knew and did not know on each subject matter.

As an instance, based on the response from GPT-4 to the question "*What type of image do you think it is?*" for the given images, it is evident that the GPT-4 accurately identifies the image as an infographic illustrating diverse machine learning algorithms and their categorizations, a flowchart or diagram detailing a machine learning or deep learning process, the image depicts a Cavalier King Charles Spaniel, a breed of dog known for its expressive eyes and friendly demeanor, and a plot representing the training dynamics of different machine learning models over a certain number of epochs as shown in figures 1, 2,3, and 7, respectively. The most interesting result is the responses received show that despite ChatGPT's versatility in analyzing images, individual privacy is paramount and must be kept confidential with the given real images of humans shown in Figures 5 and 6.

Moreover, the responses of GPT-4 for questions about the elucidation of conveyed information and determination of the most crucial knowledge that can be extracted demonstrate its ability to not only recognize the visual content but also to comprehend the specific nature and purpose of the image. The model's response underscores its proficiency in contextual understanding, as it correctly identifies the informative and categorical nature of the infographic within the broader context of machine learning algorithms. This capability is essential for applications requiring nuanced interpretation and contextualization of visual information.

The intriguing result during the analysis of the dog image was GPT-4's ability to identify the given image by emphasizing features specific to the category of dogs. This opens the door to the capability of GPT-4 in object detection based on the extracted features. Furthermore, this capability was

---

demonstrated through its recognition of objects in the provided human image in Figure 6, despite being limited in recognizing human images due to privacy considerations.

However, even though the responses received show that GPT-4 is able to extract information about the individual, it failed to point out that a pen was hanging in the pocket. This is exactly where intuition in humans is better displayed. The human brain is fired in less than a billion seconds with a glimpse of the picture to recognize what it represents. The superlative automatic intelligence inhibited in man over time could tell that the object often found in such position of the pocket is no more or less than a pen. The image may not be high resolution or positioned well before human intuition senses it. As explained by Diaz-Hernandez et al [21], the term intuition is from Latin *intueri*, meaning "to consider" or "look upon", which also connotes *perception*, will continue to place human intellect at an advantage of AI not completely replacing humans.

#### 4. Conclusion and Future Work

In conclusion, this paper focuses on examining GPT-4's capabilities in handling a variety of visual elements, including images, flowcharts, plots, and diagrams. The investigation delves into how the gleaned information from these visuals compares with human intuition, both inductive and deductive.

The results highlight GPT-4's proficiency in providing accurate and detailed information about diverse images, surpassing previous knowledge on each subject matter. The model's responses demonstrate nuanced contextual understanding and the ability to elucidate conveyed information. Moreover, it shows impressive results in tasks such as identification, recognition, and contextual understanding of visual content. Noteworthy is its capability to identify objects within images, paving the way for comprehensive use in the field of object detection. However, privacy considerations limit its recognition of individual human images. The evaluation of GPT-4's responses to questions about image content reveals both strengths and limitations. While the model excels in recognizing features specific to certain categories, it faces challenges in pinpointing specific details. Moreover, it emphasizes the irreplaceable role of human intuition, as demonstrated in the swift recognition of nuanced details that artificial intelligence may overlook.

As for future work, the demonstrative results of GPT-4 in contextual understanding of content and accurate interpretation of visual content open the door to using GPT-4 for diagnosis in medical images aligned with human interpretation within a timeframe. Moreover, this study can be extended to analyze large datasets of images, and as an evaluation metric, a similarity score can be used to assess the generated analysis against the actual analysis. This approach might provide a different perspective on the analysis compared to what humans perceive.

#### Acknowledgments

The authors wish to express their gratitude to the editorial team and the anonymous reviewers for their valuable recommendations and constructive feedback.

#### Authors Contributions

Conceptualization, O. V. J., and O. M. A.; methodology, O. V. J., and O. E. J.; software, O. M. A.; writing—original draft, O. V. J., and O. M. A.; writing—review and editing, D. A., and O. M. A.;

supervision, O. M. A.; project administration, O. M. A., and D. A. All authors have read and agreed to the published version of the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- [1] Ray, Partha Pratim. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope." *Internet of Things and Cyber-Physical Systems* (2023).
- [2] Waisberg, Ethan, Joshua Ong, Mouayad Masalkhi, Sharif Amit Kamran, Nasif Zaman, et al. "GPT-4: a new era of artificial intelligence in medicine." *Irish Journal of Medical Science* (1971-) (2023): 1-4.
- [3] Lu, Yanqiu, Haiyang Wu, Shaoyan Qi, and Kunming Cheng. "Artificial Intelligence in Intensive Care Medicine: Toward a ChatGPT/GPT-4 Way?." *Annals of Biomedical Engineering* (2023): 1-6.
- [4] Cheng, Kunming, Qiang Guo, Yongbin He, Yanqiu Lu, Ruijie Xie, Cheng Li, and Haiyang Wu. "Artificial intelligence in sports medicine: could GPT-4 make human doctors obsolete?." *Annals of Biomedical Engineering* (2023): 1-5.
- [5] Muftić, Fatima, Merjem Kadunić, Almina Mušibegović, and Ali Abd Almisreb. "Exploring Medical Breakthroughs: A Systematic Review of ChatGPT Applications in Healthcare." *Southeast Europe Journal of Soft Computing* 12, no. 1 (2023): 13-41.
- [6] Montenegro-Rueda, Marta, José Fernández-Cerero, José María Fernández-Batanero, and Eloy López-Meneses. "Impact of the implementation of ChatGPT in education: A systematic review." *Computers* 12, no. 8 (2023): 153.
- [7] Pradana, Mahir, Hanifah Putri Elisa, and Syarifuddin Syarifuddin. "Discussing ChatGPT in education: A literature review and bibliometric analysis." *Cogent Education* 10.2 (2023): 2243134.
- [8] Elbanna, Said, and Loreta Armstrong. "Exploring the integration of ChatGPT in education: adapting for the future." *Management & Sustainability: An Arab Review* (2023).
- [9] Mohammed Alyasiri, Osamah, Salisu, Sani, A. Younis, Hussain, Salman, Amer M., Sahib, Thaeer Mueen, et al. "ChatGPT Revisited: Using ChatGPT-4 for Finding References and Editing Language in Medical Scientific Articles". Available at SSRN: <https://ssrn.com/abstract=4621581> or <http://dx.doi.org/10.2139/ssrn.4621581>. (2023).
- [10] Mohammed, Osamah, Thaeer Mueen Sahib, Israa M. Hayder, Sani Salisu, and Misbah Shahid. "ChatGPT Evaluation: Can It Replace Grammarly and Quillbot Tools?." *British Journal of Applied Linguistics* 3, no. 2 (2023): 34-46.
- [11] Sahib, T. M., Alyasiri, O. M., Younis, H. A., Akhtom, D., Hayder, I. M., Salisu, S., and Muthmainnah, "A comparison between ChatGPT-3.5 and ChatGPT-4.0 as a tool for paraphrasing English Paragraphs," *International Applied Social Sciences (C-IASOS-2023) Congress*, (2024), to be published.
- [12] AL-Janabi, Omar M., Alyasiri, Osamah M., and Jebur Elaf A., "GPT-4 versus Bard and Bing: LLMs for Fake Image Detection," *In 2023nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA). IEEE*, (2024), to be published.
- [13] Alyasiri, Osamah Mohammed, and Ahmed Hussein Ali. "Exploring GPT-4's Characteristics Through the 5Vs of Big Data: A Brief Perspective." *Babylonian Journal of Artificial Intelligence* 2023 (2023): 5-9.
- [14] Sahib, Thaeer Mueen, Hussain A. Younis, Osamah Mohammed Alyasiri, Ahmed Hussein Ali, Sani Salisu, et al. "ChatGPT in Waste Management: Is it a Profitable." *Mesopotamian Journal of Big Data* 2023 (2023): 107-109.
- [15] Dua'a, A. K. H. T. O. M., Osamah Mohammed Alyasiri, Eman Allogmani, M. Amer, and Thaeer Mueen Sahib Salman, "Unlocking ChatGPT's Title Generation Potential: An Investigation of Synonyms, Readability, and Introduction-Based Titles," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 22, (2023).
- [16] Baidoo-Anu, David, and Leticia Owusu Ansah. "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning." *Journal of AI* 7.1 (2023): 52-62.



- 
- [17] Lund, Brady D., and Ting Wang. "Chatting about ChatGPT: how may AI and GPT impact academia and libraries?." *Library Hi Tech News* 40.3 (2023): 26-29.
  - [18] Handa, Palak, Deepti Chhabra, Nidhi Goel, and Sri Krishnan. "Exploring the role of ChatGPT in medical image analysis." *Biomedical Signal Processing and Control* 86 (2023): 105292.
  - [19] Jillani Soft Tech, "Exploring the Landscape of Modern Machine Learning and AI." [Online]. Available: <https://jillanisofttech.medium.com/exploring-the-landscape-of-modern-machine-learning-and-ai-cf57d8ff9637>. Accessed: Oct. 17, (2023).
  - [20] Victor, Johnson Olanrewaju, XinYing Chew, Khai Wah Khaw, and Ming Ha Lee., "A Cost-Based Dual ConvNet-Attention Transfer Learning Model for ECG Heartbeat Classification," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 90–110, (2023).
  - [21] Diaz-Hernandez, Octavio, and Victor Javier Gonzalez-Villela, "Analysis of Human Intuition Towards Artificial Intuition Synthesis for Robotics," *Mechatronics and Applications: An International Journal (MECHATROJ)* 1, no. 1, (2017).