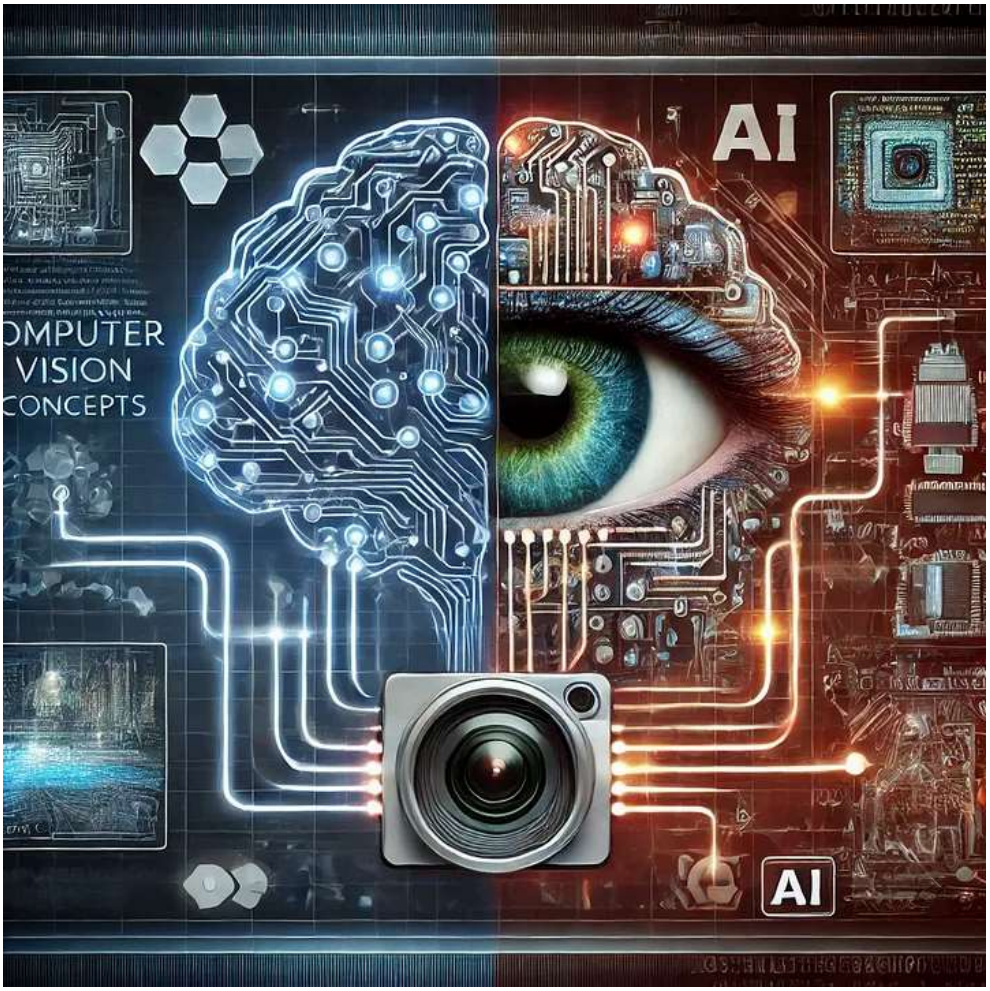
Member-only story

Ace Your 2024 Computer Vision Interview: Key Questions & Expert Answers

Ritesh Gupta · Follow
Published in Artificial Intelligence in Plain English · 16 min read · Oct 4, 2024

Common Computer Vision Interview Questions & Answers (2024)

Computer Vision (CV) is a field of artificial intelligence that enables machines to interpret and make decisions based on visual data. If you're preparing for a Computer Vision interview in 2024, it's crucial to be familiar with both the theoretical and practical aspects of this domain. Below, we've compiled a list of common interview questions and their answers to help you prepare.



1. What is Computer Vision? How does it differ from Image Processing?

Answer: Computer Vision is a field of AI where computers are trained to interpret and understand visual information from the world, similar to how humans do. The aim is to make machines capable of identifying and processing objects, people, and environments in images or videos to take appropriate actions.

Difference:

- **Computer Vision** deals with extracting high-level information from images (like identifying objects or faces).
- **Image Processing** focuses more on manipulating or enhancing images, such as removing noise or improving brightness.

For example, image processing might improve the clarity of a picture, while computer vision would focus on recognizing a face in the image.

💡 **Tip:** Think of image processing as *how* the image is handled, and computer vision as *what* the image is telling you.

2. What are the common techniques used in Computer Vision?

Answer: Here are some common techniques used in computer vision:

- **Edge Detection:** Used to identify the boundaries of objects within images. Methods include Sobel, Prewitt, and Canny edge detectors.
- **Feature Detection & Matching:** Algorithms like SIFT (Scale-Invariant Feature Transform) and ORB (Oriented FAST and Rotated BRIEF) are used to detect and match key points in different images.
- **Object Detection:** Techniques like YOLO (You Only Look Once) and SSD (Single Shot Detector) help in detecting objects in real-time.
- **Convolutional Neural Networks (CNNs):** A class of deep neural networks, CNNs are widely used for image classification and object detection tasks.

Example: You might use a CNN model to identify cats in pictures or recognize handwritten digits.

🤖 **Real-world application:** Self-driving cars use object detection to avoid obstacles and ensure safe navigation.

3. Explain how Convolutional Neural Networks (CNN) work.

Answer: CNNs are a type of deep learning algorithm designed specifically for image processing tasks. They use a combination of layers that work together to extract features from input images:

- **Convolution Layer:** This layer applies a filter (or kernel) to the input image, extracting features like edges, textures, or patterns.
- **Pooling Layer:** This layer reduces the dimensionality of the data while retaining important information, making the model more computationally efficient.
- **Fully Connected Layer:** Once the convolution and pooling layers have extracted features, these are passed to fully connected layers that make predictions or classifications.

🔍 **Example:** Let's say we want to classify images of cats and dogs. The CNN will first learn simple patterns (like edges) in initial layers and gradually

recognize more complex structures like eyes, ears, or fur patterns in deeper layers.

4. What are Generative Adversarial Networks (GANs) and how do they work?

Answer: Generative Adversarial Networks (GANs) consist of two neural networks: a **generator** and a **discriminator**, which work against each other in a game-like setting.

- **Generator:** This network tries to generate fake data that looks real (for example, fake images).
- **Discriminator:** This network tries to distinguish between real and fake data.

Over time, the generator gets better at producing realistic data, while the discriminator improves in identifying fakes.

Example: GANs can be used to generate realistic-looking images of humans that do not exist.

👹 **Fun Fact:** GANs are widely used in applications like image generation, video game character design, and even enhancing low-quality photos.

5. What is Image Segmentation and how is it different from Object Detection?

Answer:

- **Image Segmentation** involves dividing an image into multiple segments (or regions) to simplify its analysis. It assigns a label to every pixel in the image, where similar pixels (like those forming an object) are grouped together.
- **Object Detection** focuses on identifying and locating objects within an image, usually by drawing bounding boxes around detected objects.

Example: In a medical image, segmentation can be used to identify and highlight areas of a tumor. Object detection, on the other hand, would place a box around the tumor to locate it.

💡 **Key Difference:** Segmentation is more detailed since it labels individual pixels, whereas object detection is about localizing objects within the image.

6. What are some challenges in Computer Vision?

Answer: Some of the key challenges in Computer Vision include:

- **Variability in Images:** Objects can appear different due to changes in lighting, occlusion, or angle.
- **Data Scarcity:** Large datasets are often required to train deep learning models effectively.
- **Real-Time Processing:** Processing visual data in real-time for applications like autonomous driving requires high computational power.
- **Generalization:** Models trained on specific datasets may not generalize well to different environments or tasks.

7. What is Transfer Learning in Computer Vision?

Answer: Transfer learning is a technique where a pre-trained model (usually trained on a large dataset) is used as a starting point for a new task, instead of training from scratch. This is particularly useful in computer vision, as training from scratch requires a lot of data and computational power.

Example: If you have a pre-trained model that was trained to recognize animals, you can use that model's learned features to identify specific breeds of dogs by fine-tuning the last few layers.

🤝 **Benefit:** Transfer learning saves time and resources, as it reuses knowledge from previous models.

8. Can you explain Optical Flow?

Answer: Optical flow is the pattern of apparent motion of objects in a visual scene caused by the relative motion between the camera and the objects. It captures the movement of pixels between two consecutive frames in a video sequence.

Example: Optical flow is used in video compression, motion detection, and video stabilization.

9. What is the role of Augmented Reality (AR) in Computer Vision?

Answer: Augmented Reality (AR) uses computer vision techniques to overlay digital content onto the real-world environment. By recognizing objects or markers in the real world, AR applications can project images, animations, or information on top of them.

Example: Popular apps like Pokémon GO use AR to show virtual creatures in the real world using your phone's camera.

10. What are the commonly used datasets in Computer Vision?

Answer: Here are some widely-used datasets in computer vision:

- **ImageNet:** A large visual database designed for object recognition tasks.
- **COCO (Common Objects in Context):** Provides labeled images for object detection, segmentation, and keypoint detection.
- **MNIST:** Contains images of handwritten digits used for image classification tasks.
- **Pascal VOC:** Another dataset for object detection, segmentation, and image classification.

📊 **Pro Tip:** Familiarity with datasets like these can help you answer practical questions during an interview.

11. What are the advantages and limitations of using CNNs for Computer Vision tasks?

Answer:

Advantages:

- **Automatic Feature Extraction:** CNNs automatically learn the features needed for a task (like edges, textures, or shapes) through convolutional layers, removing the need for manual feature engineering.
- **Parameter Sharing:** Filters (kernels) in CNNs are shared across the image, which drastically reduces the number of parameters compared to fully connected networks, making the model more efficient.
- **Spatial Hierarchy:** CNNs capture hierarchical patterns, from low-level features (edges) to high-level features (complex shapes), making them ideal for image classification and object detection tasks.

- **Translation Invariance:** CNNs can recognize objects in different locations within an image due to pooling layers, making them robust to variations in object positioning.

Limitations:

- **Need for Large Datasets:** CNNs perform best with large amounts of labeled data. Without enough data, they may overfit or not learn meaningful patterns.
- **Computationally Expensive:** Training large CNN models requires high computational power, especially for tasks like object detection in real-time applications.
- **Lack of Explainability:** It can be difficult to interpret what features CNNs are learning, leading to a lack of transparency in decision-making.

🔍 **Example:** CNNs might outperform traditional machine learning models in tasks like recognizing handwritten digits from the MNIST dataset. However, if applied to a small dataset, the model might overfit, learning details specific to the training data but failing on new images.

12. What is the difference between Semantic and Instance Segmentation?

Answer:

- **Semantic Segmentation:** This technique classifies each pixel of an image into a predefined class. It treats multiple instances of the same object class as one, grouping them together. For example, all cars in an image would be labeled as “car” without distinguishing between individual cars.
- **Instance Segmentation:** This method goes a step further by not only classifying pixels but also distinguishing between different instances of the same object class. In the case of cars, it would label each car separately, even if they belong to the same class.

Example:

- In a street scene image, **semantic segmentation** might label the entire road as one “road” class, while **instance segmentation** would label each pedestrian and car separately, even though they’re all pedestrians or cars.

🚗 **Use case:** Autonomous vehicles often use instance segmentation to differentiate between multiple cars, pedestrians, and other objects on the road.

13. How does Image Augmentation help in improving the performance of Computer Vision models?

Answer: Image augmentation refers to artificially increasing the size of the dataset by applying transformations such as rotation, flipping, zooming, and more to the original images. This helps in improving the robustness and generalization of computer vision models.

Benefits:

- **Prevents Overfitting:** Augmentation introduces variations in the training data, helping the model generalize better to unseen data, and reducing overfitting.
- **Expands Data:** It creates new samples from the existing dataset, which is especially helpful when there is a shortage of labeled data.
- **Improves Robustness:** Models become more adaptable to changes like rotations, shifts, or brightness variations in real-world images.

🔄 **Example:** In a facial recognition task, flipping and rotating face images during training can help the model identify faces even when they're not perfectly aligned in the real world.

14. Explain HOG (Histogram of Oriented Gradients). How is it used in Computer Vision?

Answer: HOG is a feature descriptor used for object detection. It works by dividing the image into small connected regions (cells) and calculating the gradient direction or intensity for each cell. The final feature vector represents the distribution of gradients (oriented edges) in the image, which can be used to detect objects like pedestrians, vehicles, or animals.

Steps in HOG:

- **Gradient Computation:** For each pixel, the gradient (direction and magnitude) is calculated.
- **Orientation Binning:** Gradients are binned based on their orientation into a fixed number of bins.

- **Block Normalization:** Blocks of cells are normalized to account for variations in lighting and contrast.
- **Feature Vector:** The HOG descriptors are concatenated to form the final feature vector.

🕵️ **Example:** HOG is often used for detecting humans in surveillance videos or images by analyzing the shape and outline of a person's body.

15. What is the role of Feature Pyramid Networks (FPN) in Object Detection?

Answer: Feature Pyramid Networks (FPN) are designed to handle multi-scale object detection. In object detection, objects in an image can appear in different sizes. FPNs improve detection performance by creating a pyramid of features at different resolutions. This allows the network to detect both large and small objects more effectively.

How FPN works:

- FPN extracts features at multiple scales by combining high-resolution and low-resolution feature maps.
- It allows the network to detect objects of various sizes by providing richer information across scales.

📊 **Example:** FPNs are widely used in state-of-the-art object detection frameworks like Faster R-CNN and Mask R-CNN, improving accuracy for both small and large objects.

16. How is Computer Vision used in Healthcare?

Answer: Computer Vision has significant applications in healthcare, improving diagnostics, treatment, and patient care. Some key uses include:

- **Medical Imaging:** CV models help in analyzing X-rays, MRIs, and CT scans to detect diseases like tumors, fractures, or infections.
- **Surgical Assistance:** Vision systems can guide robotic surgeries by providing real-time feedback and precision during procedures.
- **Telemedicine:** CV is used in remote diagnostics where visual examination is required, such as skin disease detection.

- **Patient Monitoring:** Cameras with CV technology can track patients' movements and alert medical staff in case of an emergency (e.g., if a patient falls).

👨‍⚕️ **Example:** A deep learning model can be trained to detect lung cancer by analyzing patterns in CT scan images, providing early detection and improving patient outcomes.

17. What is the significance of OpenCV in Computer Vision?

Answer: OpenCV (Open Source Computer Vision Library) is a widely-used open-source library that contains many tools and functions for building computer vision applications. It provides modules for image processing, object detection, video capture, and more.

Key Features:

- Supports multiple programming languages like Python, C++, and Java.
- Contains optimized algorithms for real-time applications.
- Extensive community support and documentation.

💻 **Example:** You can use OpenCV to build a real-time face detection system using the built-in functions for detecting facial landmarks and objects.

18. How do Self-Supervised Learning techniques contribute to Computer Vision?

Answer: Self-supervised learning (SSL) is an approach where models learn representations from unlabeled data by predicting parts of the data from other parts. It bridges the gap between supervised learning (which requires large amounts of labeled data) and unsupervised learning. In the context of computer vision, SSL can help models learn useful image representations without needing large, annotated datasets.

- **Reduction in Label Dependency:** With self-supervised learning, models learn from unlabeled images, reducing the need for extensive manual labeling, which can be expensive and time-consuming.
- **Better Pretraining:** SSL techniques can be used for pretraining models, allowing them to learn general features from a large amount of unlabeled data. This pretraining can be followed by supervised fine-tuning on a smaller labeled dataset for a specific task.

- **Improved Performance:** SSL-based models can achieve better performance in tasks like image classification, object detection, and segmentation, as they learn rich representations from large unlabeled datasets.

Popular SSL Techniques in CV:

- **Contrastive Learning:** Models like SimCLR and MoCo aim to learn representations by pulling similar images together (e.g., different views of the same image) and pushing dissimilar images apart.
- **Masked Image Modeling (MIM):** Similar to how BERT works in NLP, models like MAE (Masked Autoencoders) mask parts of an image and train the model to predict the missing regions.

🔍 **Example:** In an SSL task, the model might take an image, apply random augmentations (like rotating or cropping), and try to predict the original or reconstruct the masked parts. This forces the model to learn meaningful image features.

19. What is the role of 3D Vision in Computer Vision?

Answer: 3D Vision enables machines to perceive and interpret depth information from images or video, allowing for a richer understanding of the world compared to standard 2D vision. It's crucial for tasks where understanding the geometry of objects is necessary, like robotics, autonomous vehicles, and AR/VR applications.

Applications of 3D Vision:

- **Depth Estimation:** Predicting the distance of objects from the camera by analyzing disparities between different views (stereo vision) or using depth sensors.
- **3D Object Detection:** Identifying objects in a 3D space rather than just 2D images. This is essential for applications like autonomous driving, where it's important to know the real-world dimensions and positions of obstacles.
- **3D Reconstruction:** Creating a 3D model of an object or scene from multiple 2D images. This is used in fields like medical imaging, architecture, and augmented reality.

Example: A robot using 3D vision can better navigate its environment by detecting obstacles' precise positions and shapes, rather than just detecting their presence.

🛠️ **Real-world use case:** Lidar and depth cameras in autonomous vehicles help create 3D maps of the surrounding environment to avoid collisions and navigate roads safely.

20. What is the difference between YOLO and Faster R-CNN in Object Detection?

Answer: Both YOLO (You Only Look Once) and Faster R-CNN are popular object detection algorithms, but they differ significantly in their approaches and use cases.

YOLO:

- **Speed:** YOLO is designed for real-time object detection, processing images extremely fast. It achieves this by framing object detection as a single regression problem, where the image is divided into grids, and each grid predicts the bounding box and class for an object.
- **Accuracy:** While YOLO is fast, its accuracy is lower compared to algorithms like Faster R-CNN, especially for small objects.

Faster R-CNN:

- **Accuracy:** Faster R-CNN uses a region proposal network (RPN) to propose object locations before classifying them, leading to higher accuracy.
- **Speed:** However, Faster R-CNN is slower because it processes images in multiple stages — first proposing regions of interest and then classifying them.

Example:

- If you need to detect objects in real-time, like in a video stream from a surveillance camera, **YOLO** would be the better choice.
- For more detailed image analysis, like in medical imaging, where accuracy is crucial, **Faster R-CNN** would be preferred.

🚀 **Fun fact:** YOLO is often used in real-time applications like drone navigation, while Faster R-CNN is used in applications where detection precision is more critical than speed.

21. What are the ethical concerns associated with Computer Vision?

Answer: As computer vision systems become more integrated into everyday life, ethical considerations around their usage are gaining prominence. Here are some major concerns:

- **Privacy Invasion:** Surveillance systems using computer vision can infringe on individuals' privacy, especially with widespread deployment of facial recognition technologies in public spaces.
- **Bias and Fairness:** If trained on biased datasets, computer vision models may exhibit discriminatory behavior. For example, facial recognition systems might perform poorly on certain ethnic groups if the training data doesn't represent them well.
- **Security and Misuse:** Computer vision technology, such as deepfakes (generated using GANs), can be used to manipulate videos and images for malicious purposes, leading to misinformation and identity fraud.
- **Autonomy in Decision-Making:** In critical areas like healthcare or law enforcement, over-reliance on computer vision systems could lead to errors or biases affecting people's lives.

🔍 **Example:** If a facial recognition system wrongly identifies an individual as a suspect in a criminal investigation, it can lead to wrongful accusations and serious consequences.

💡 **Key takeaway:** Ensuring fairness, transparency, and accountability in the development and deployment of computer vision systems is crucial to addressing these ethical challenges.

22. What are Capsule Networks, and how do they improve upon CNNs in Computer Vision?

Answer: Capsule Networks (CapsNets) were introduced to address some of the limitations of Convolutional Neural Networks (CNNs), particularly in handling hierarchical relationships between objects in an image. CapsNets are designed to better model the spatial relationships between different parts of an object, making them more robust to changes in orientation, scale, and perspective.

Key Concepts:

- **Capsules:** A capsule is a group of neurons that captures not only the presence of a feature but also its properties (like orientation and position).
- **Dynamic Routing:** Capsules use dynamic routing to communicate with each other, ensuring that relevant information is passed forward while irrelevant information is ignored. This reduces the risk of misclassifications caused by changes in an object's orientation.

Advantages over CNNs:

- **Better at Capturing Hierarchical Information:** CNNs rely on max-pooling to reduce dimensions, which can cause loss of important spatial relationships. Capsule networks preserve these relationships better by understanding the pose and arrangement of objects.
- **Improved Robustness:** CapsNets are more robust to variations in pose, rotation, and viewpoint. This makes them particularly useful in tasks where objects may appear in different orientations.

Example: A CapsNet would be able to differentiate between a rotated cat and a flipped cat, whereas a traditional CNN might struggle to recognize the cat if its orientation significantly changes.

🔄 **Real-world application:** Capsule networks could be useful in tasks like medical imaging, where the orientation and position of objects (like organs or tumors) vary significantly, and precise spatial relationships are critical for accurate diagnosis.

23. What is the significance of Multi-Task Learning (MTL) in Computer Vision?

Answer: Multi-Task Learning (MTL) is an approach where a single model is trained to perform multiple tasks simultaneously, such as object detection, segmentation, and classification. The idea is that learning different tasks in parallel allows the model to share representations and learn more efficiently.

Advantages:

- **Improved Generalization:** By learning from multiple tasks, the model can generalize better to new tasks or data, as the shared knowledge helps in capturing more useful features.
- **Efficient Training:** Instead of training separate models for each task, MTL allows for simultaneous training, reducing the time and computational cost.
- **Data Efficiency:** Some tasks may have limited labeled data, but by training jointly with tasks that have more data, the model can benefit from this additional information.

Example: In autonomous driving, a multi-task learning model might be trained to detect objects (cars, pedestrians), estimate depth, and segment road lanes — all in one go. This results in a more efficient and powerful system compared to training separate models for each task.

💡 **Key Insight:** MTL helps make the most of available data and computational resources, and is particularly useful in scenarios where related tasks can share useful information.

24. What are Attention Mechanisms, and how are they used in Computer Vision?

Answer: Attention mechanisms allow models to focus on the most relevant parts of the input data, selectively giving more weight to certain regions or features. Originally popularized in Natural Language Processing (NLP) with models like Transformers, attention mechanisms are now widely used in computer vision tasks to enhance the performance of convolutional networks and other architectures.

How Attention Works:

- Attention assigns different weights to different regions of an image, helping the model "attend" to the most important parts, like objects or features that contribute the most to a particular task (e.g., classification or detection).

Types of Attention in CV:

1. **Spatial Attention:** Focuses on specific areas in an image where important features are likely to be found.

2. **Channel Attention:** Focuses on different feature maps or channels within the image, emphasizing certain features more than others.

Example: In an image classification task, attention mechanisms can help the model focus more on the object of interest (like a cat) while ignoring background details (like trees or houses), improving both accuracy and efficiency.

📊 **Popular Models:** Vision Transformers (ViT) use attention mechanisms as their core architecture and have shown state-of-the-art performance in many computer vision tasks.

25. What is the role of Reinforcement Learning (RL) in Computer Vision?

Answer: Reinforcement Learning (RL) is a learning paradigm where an agent interacts with an environment and learns to make decisions by receiving feedback in the form of rewards or penalties. In computer vision, RL is particularly useful in applications where sequential decision-making is required.

Applications of RL in CV:

- **Object Tracking:** RL can be used for dynamic object tracking, where an agent learns to follow a moving object across frames by making decisions about where to focus in each frame.
- **Active Vision:** In scenarios where computational resources are limited (e.g., on a drone or robot), RL can guide the vision system to focus on the most relevant parts of the environment, instead of processing the entire scene at once.
- **Autonomous Navigation:** RL is used in robots and self-driving cars to navigate environments based on visual input, learning optimal actions to reach a goal while avoiding obstacles.

Example: In a drone navigation task, RL can teach the drone to learn the best path by interpreting its visual inputs and adjusting its trajectory to avoid collisions.

🤖 **Real-world use case:** In autonomous vehicles, RL is used in conjunction with computer vision to help the vehicle learn optimal driving strategies in complex traffic environments.

Final Thoughts

Computer vision is rapidly advancing, and staying updated with the latest concepts, techniques, and ethical considerations is essential for success in interviews and in the field. By understanding the questions discussed above, you can demonstrate both theoretical knowledge and practical expertise. As industries continue to adopt CV for real-world applications, having a solid grasp of the fundamentals, as well as the cutting-edge developments, will position you as a strong candidate in this ever-evolving domain.

Good luck on your journey to mastering computer vision! 🚀

In Plain English 🚀

*Thank you for being a part of the **In Plain English** community! Before you go:*

- Be sure to **clap** and **follow** the writer 👏
- Follow us: **X** | **LinkedIn** | **YouTube** | **Discord** | **Newsletter**
- Visit our other platforms: **CoFeed** | **Differ**
- More content at **PlainEnglish.io**

Computer Vision | Deep Learning | Machine Learning | Artificial Intelligence | Data Science

52

Published in Artificial Intelligence in Plain English

Follow

14K Followers · Last published 12 hours ago

New AI, ML and Data Science articles every day. Follow to join our 3.5M+ monthly readers.

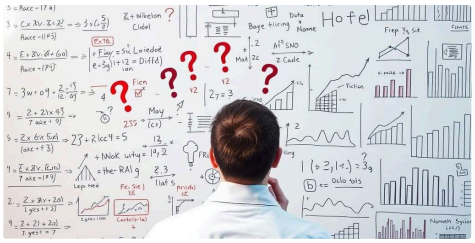
Written by Ritesh Gupta

Follow

3.6K Followers · 28 Following

Data Scientist, I write Article on Machine Learning| Deep Learning| NLP | Open CV | AI Lover ❤️

More from Ritesh Gupta and Artificial Intelligence in Plain English

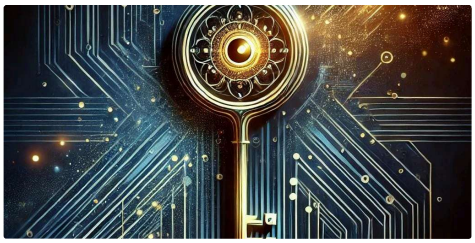


Ritesh Gupta

Can You Handle These 25 Toughest Data Science Interview Questions?

The role of a Data Scientist demands a unique blend of skills, including statistics, machine...

Sep 25 211 7



In Artificial Intelligence in Plain En... by Andrew B...

New KILLER ChatGPT Prompt—The “Playoff Method”

Super powerful prompt for ChatGPT—01 Preview

Sep 27 5K 96

In Artificial Intelligence in Plain ... by Antony Matt...

Only 1% Chat GPT users know these Secret Prompts

These can 10X the Quality of your Chat GPT Responses

Oct 16 2.3K 27

In Python in Plain English by Ritesh Gupta

7 GitHub Repos to Transform You into a Pro ML/AI Engineer

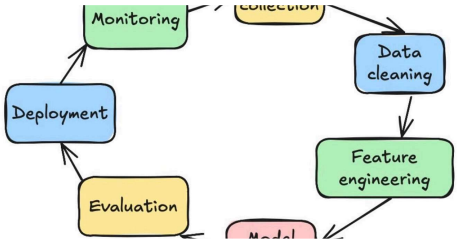
Hands-On Guides, Tools, and Frameworks to Fast-Track Your AI Journey

Nov 5 143 1

See all from Ritesh Gupta

See all from Artificial Intelligence in Plain English

Recommended from Medium

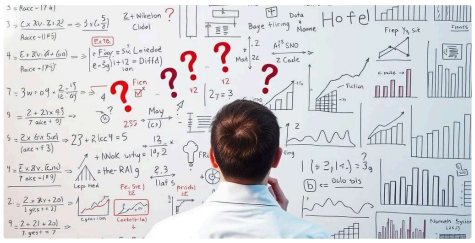


In Towards Data Science by Haden Pelletier

Every Step of the Machine Learning Life Cycle Simply Explained

A comprehensive guide to the ML life cycle with examples in Python

3d ago 193 3



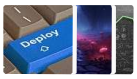
Ritesh Gupta

Can You Handle These 25 Toughest Data Science Interview Questions?

The role of a Data Scientist demands a unique blend of skills, including statistics, machine...

Sep 25 211 7

Lists

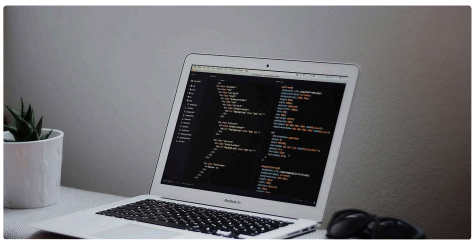


Predictive Modeling w/ Python
20 stories · 1691 saves

Natural Language Processing
1839 stories · 1461 saves

Practical Guides to Machine Learning
10 stories · 2057 saves

data science and AI
40 stories · 292 saves

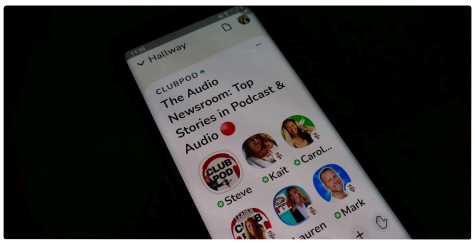


In Stackademic by Abdur Rahman

How Stopping Coding Turned Me Into a Truly Better Developer

What happened when I stopped learning coding — and how it made me more...

4d ago 1.1K 22

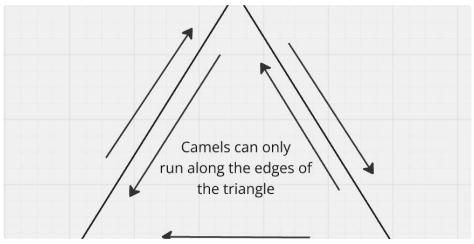


Pierre DeBois

Podcasts That Will Elevate Your Data Science Skills and Tech...

Like radio, podcasts offer a window into the experiences of others. Use that to build your...

4d ago 25



Lucas Samba

3 Probability Questions I was asked in Walmart Data Scientist Interview

Recently I got an opportunity to interview at Walmart for Data Scientist—3 position. All...

Aug 23 1.1K 32

In Towards AI by Talib

RAG from a Beginner to Advance—Introduction

Most of us have interacted with large language models (LLMs) like GPT-4, and the...

6d ago 195 3

See more recommendations