

CAMERA CHATBOT



ABOUT THE PROJECT

The Smart Classroom Assistant is a desktop application designed to enhance classroom interactivity using AI technologies. It combines real-time object detection, speech-to-text conversion, and an AI-powered chatbot within an intuitive GUI.

Students or instructors can interact with the assistant using their voice, and the assistant will:

- Recognize speech in real-time and display the transcribed text.
- Respond using a large language model (e.g., Mistral via Ollama).
- Detect physical objects or gestures in the classroom using a webcam and a deep learning-based detection model (YOLOv5).

This tool runs entirely on a local PC, making it suitable for educational settings where privacy and offline functionality are important.



YOLOV5

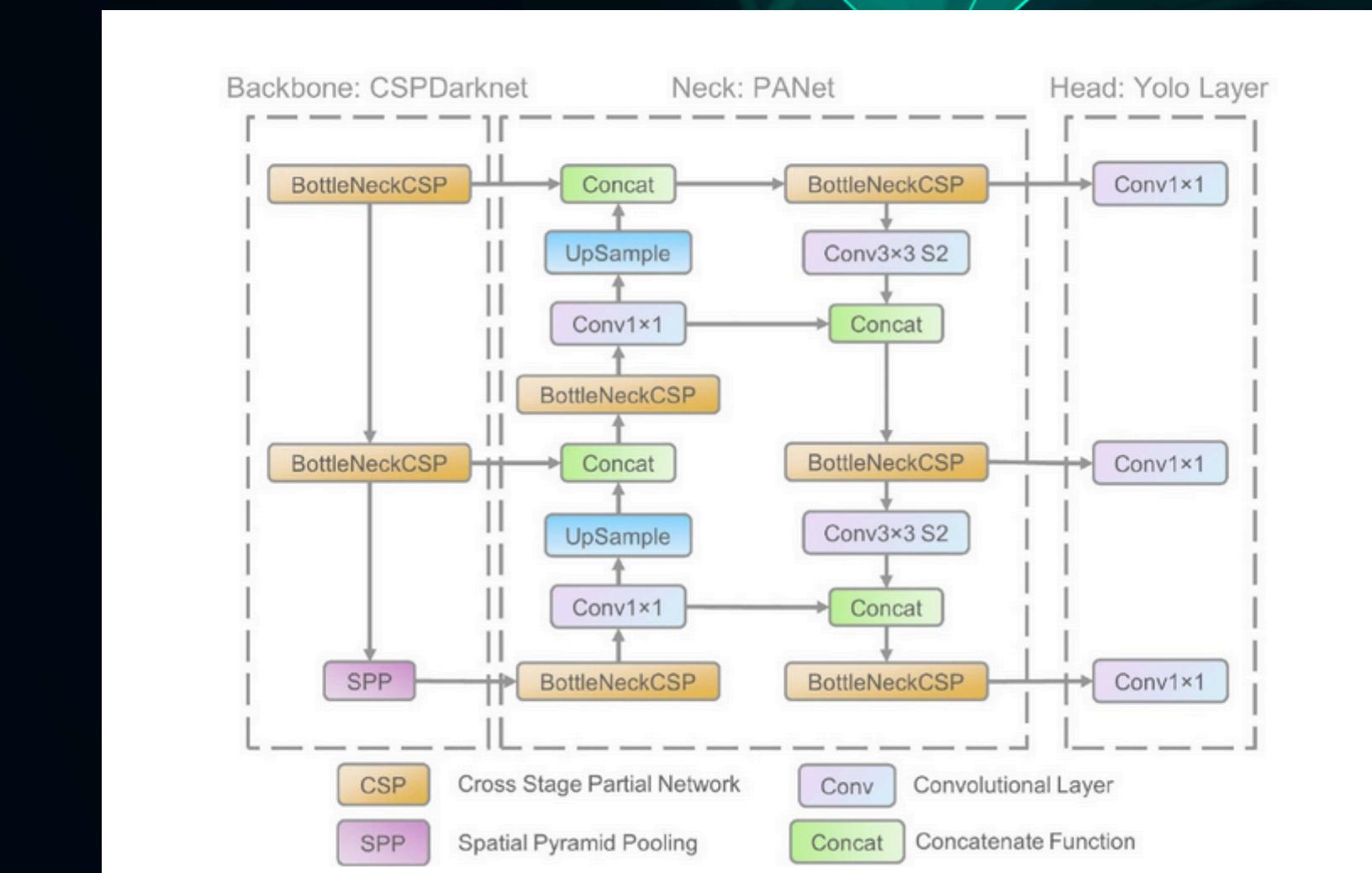
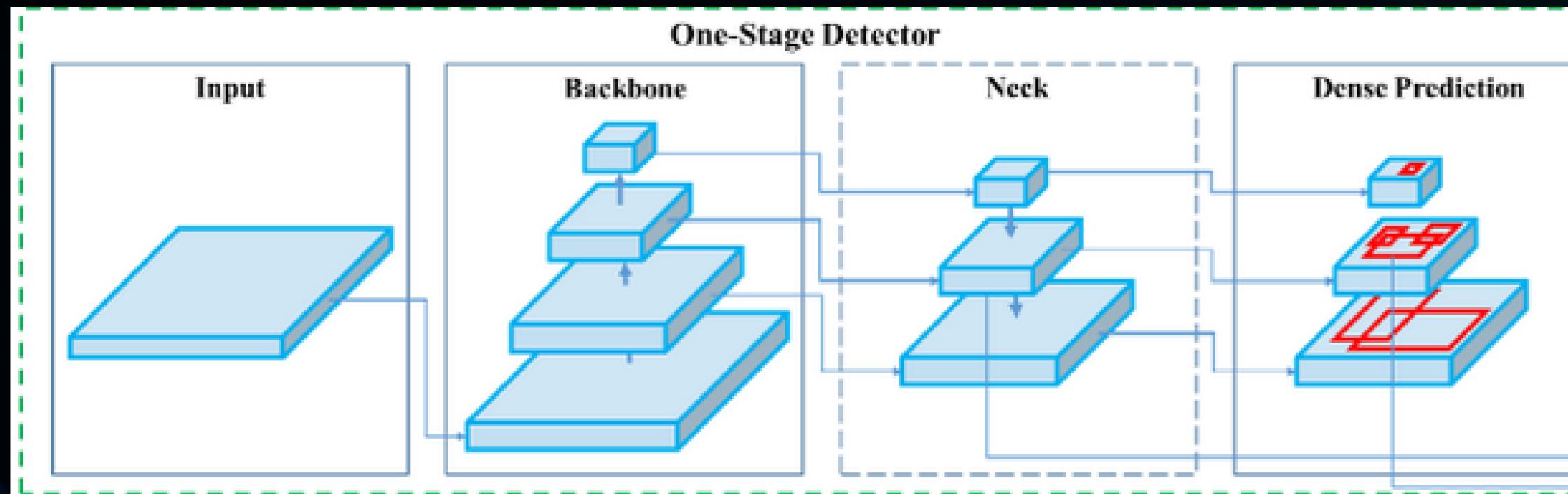


YOLOv5 (You Only Look Once version 5) is a state-of-the-art, real-time object detection model developed by Ultralytics. It uses deep learning to detect multiple objects in an image or video frame in a single forward pass, making it extremely fast and efficient.

YOLOv5 is built on convolutional neural networks (CNNs) and is known for its balance of speed, accuracy, and small model size, making it suitable for real-time applications like surveillance, robotics, and classroom assistants.

It supports PyTorch, has pre-trained models (like `yolov5s.pt`, `yolov5m.pt`, etc.), and can be easily customized for detecting specific objects.

YOLOV5 Architecture



MISTRAL - OLAMA

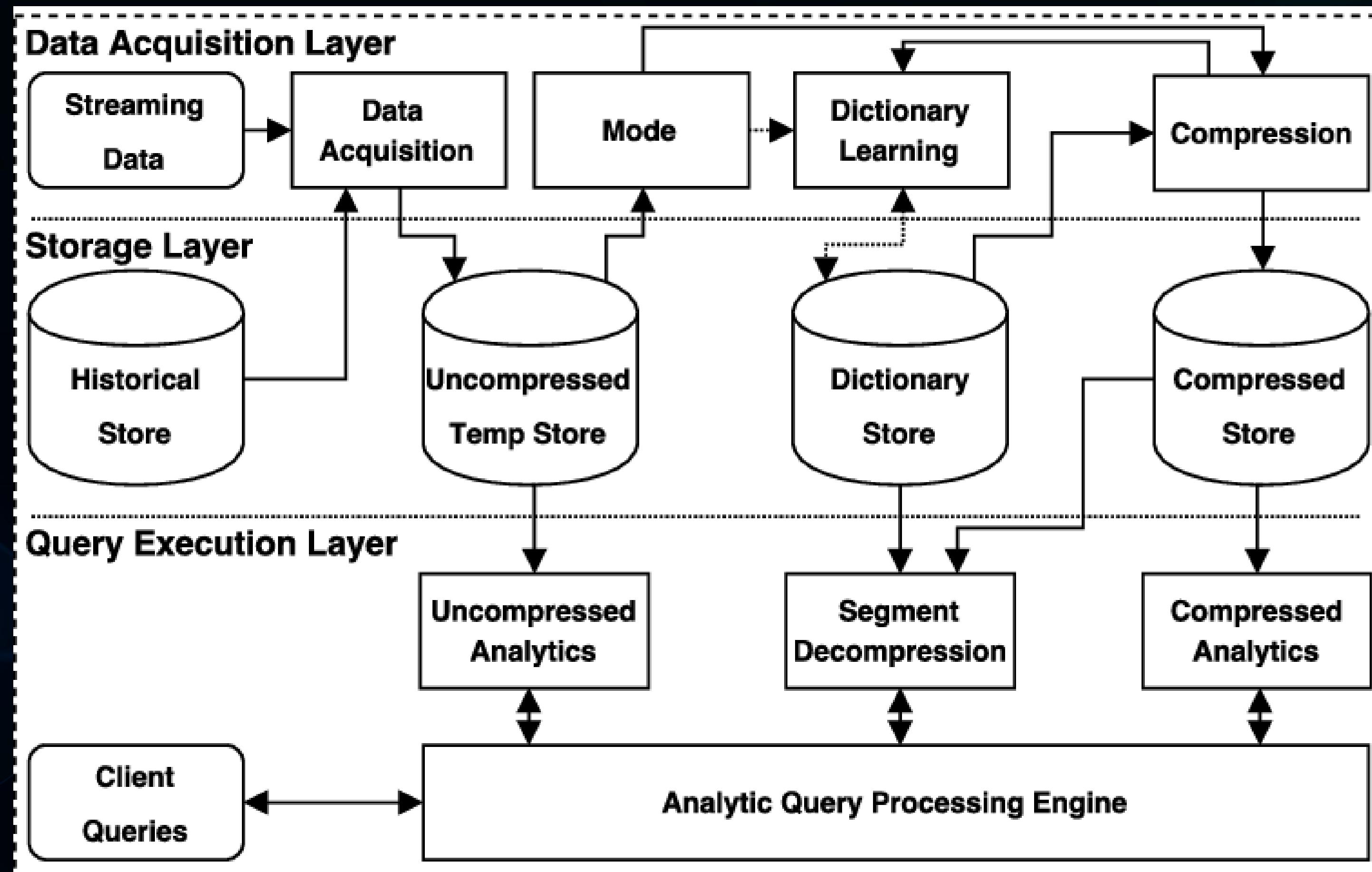


Mistral is a lightweight, open-source large language model (LLM) developed to deliver fast and efficient performance while still generating high-quality text. It's based on the Transformer architecture, the same powerful design used in models like GPT.

Mistral is trained on a large amount of internet text and is capable of:

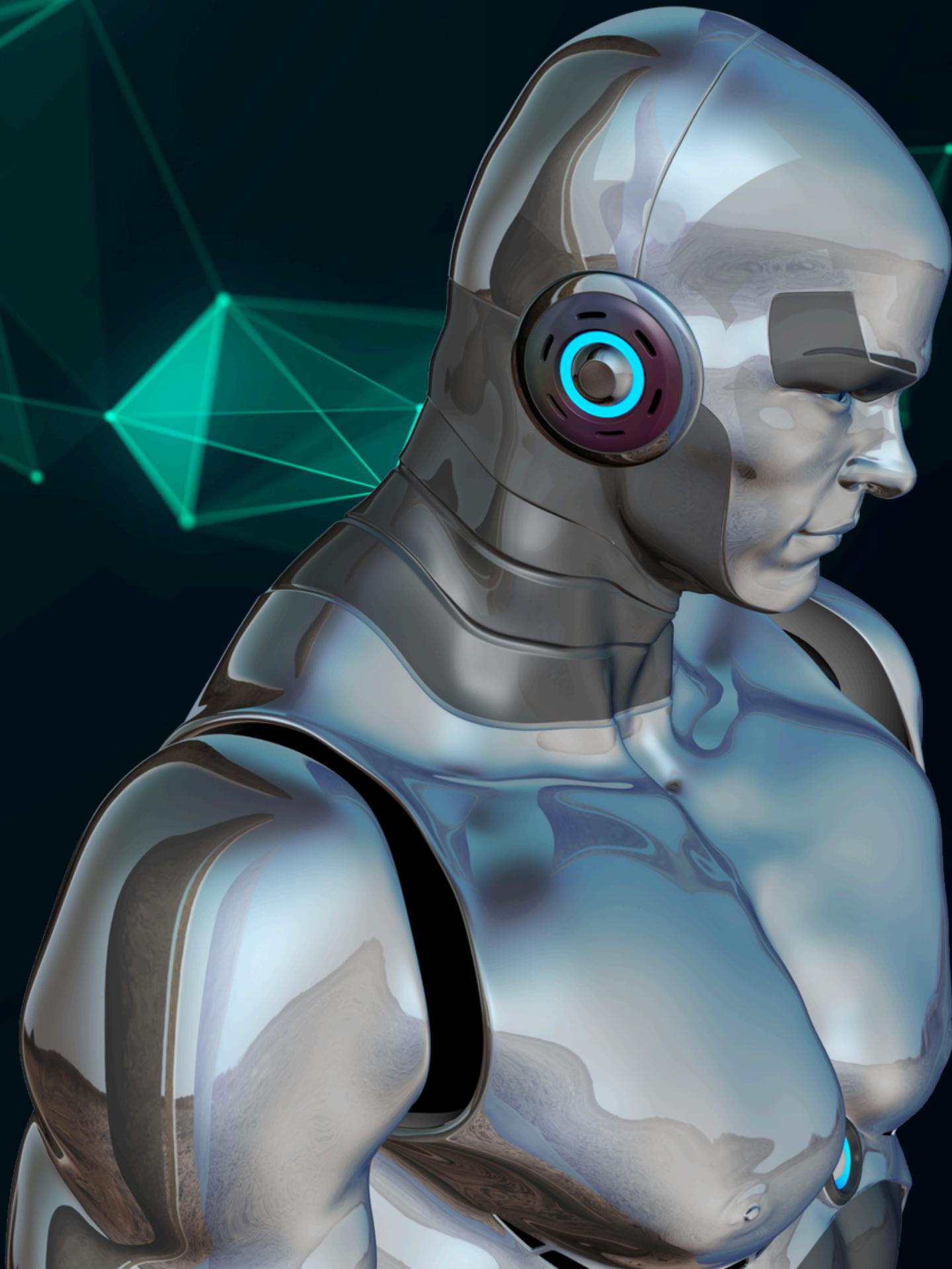
- Understanding and generating human-like responses
- Answering questions
- Summarizing content
- Holding contextual conversations

In this project, Mistral is used via Ollama, which allows you to run the model locally on your PC—no cloud connection needed. This makes it perfect for private, offline AI chatbot interactions in classrooms.



NATURAL LANGUAGE PROCESSING (CHATBOT WITH MISTRAL VIA OLLAMA)

- Concept: A language model (LLM) that generates human-like responses.
- Library/Tool: ollama, running Mistral model locally.
- What it does: Converts user input (spoken) into a chatbot response.
- DL Principle: Uses the transformer architecture, trained on large-scale text data to understand and generate meaningful text.



SPEECH-TO-TEXT (STT) CONVERSION



- Concept: Converts spoken audio into text in real-time.
- Library/Tool: speech_recognition + sounddevice
- What it does: Captures live audio, converts it into text using Google's STT engine.
- DL Principle: Google's backend model uses recurrent or transformer-based deep learning models trained on audio-linguistic data.

REAL-TIME VIDEO PROCESSING



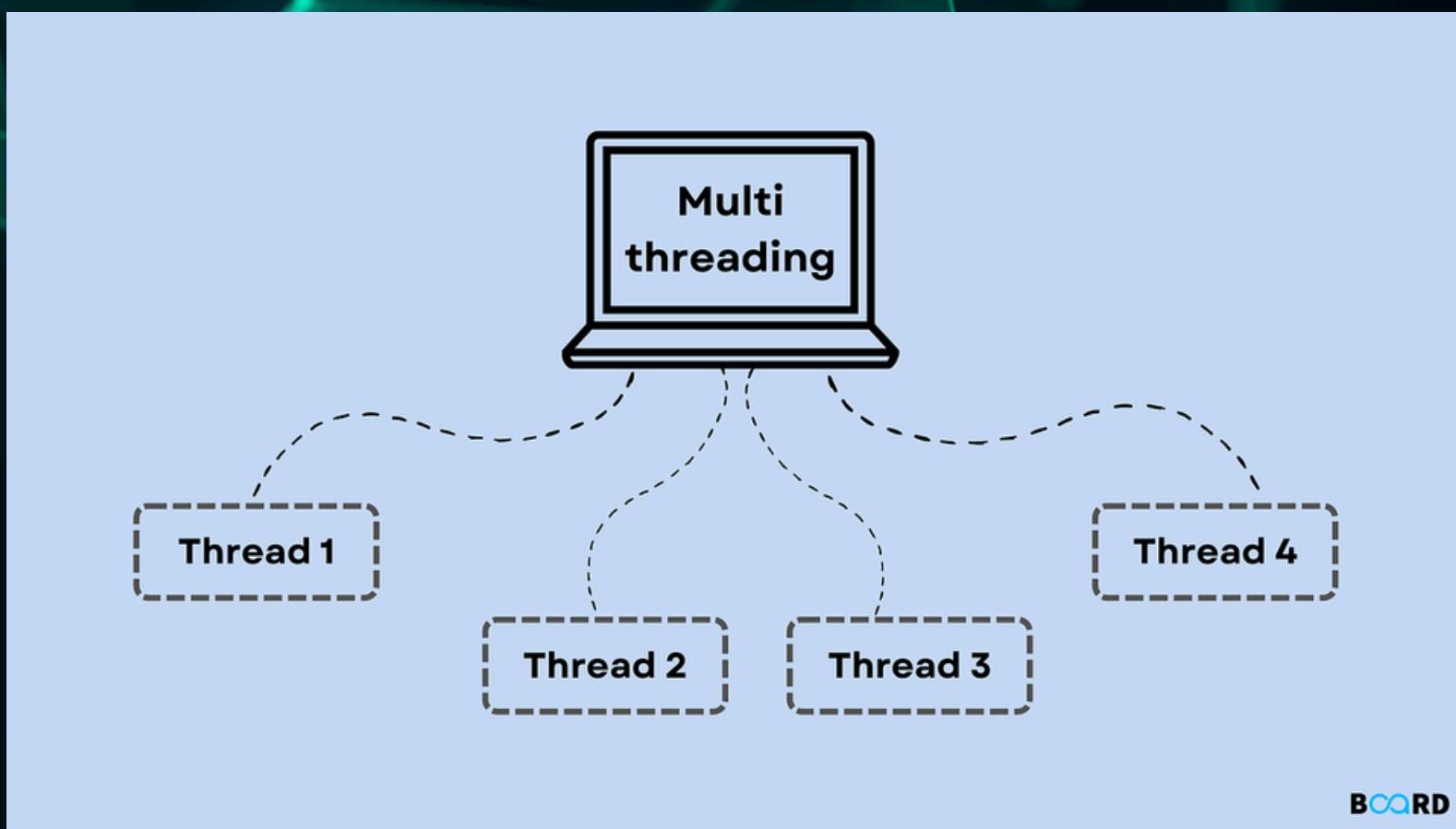
- Concept: Captures live webcam video and applies object detection.
- Library/Tool: OpenCV, YOLOv5
- What it does: Captures frames from the webcam, feeds them into YOLOv5, and overlays detection results in real-time

GUI DEVELOPMENT (USER INTERFACE)

- Concept: Interactive desktop interface for controlling audio/video/chat.
- Library/Tool: PyQt5
- What it does:
 - Displays video frames with detection boxes
 - Shows transcription text
 - Provides chatbot interaction
 - Buttons for audio control



MULTI-THREADING FOR REAL-TIME PERFORMANCE



- Concept: Running video and audio processing in parallel threads.
- Tool: QThread from PyQt5
- What it does:
- Keeps the GUI responsive while processing audio/video in the background.
- Avoids GUI freezing by isolating long-running tasks.

LOCAL EXECUTION & PRIVACY



- Concept: No cloud services needed.
- Tool: Ollama runs LLM locally using models like Mistral.
- Why it's important: Ensures data privacy and works without internet access—great for offline classrooms.

Smart Classroom Assistant - AI Chatbot & Real-Time Detection

person 0.67

cell phone 0.36

Live Transcription:
hello I am Sam what's your name

- Perception: AI systems can perceive the world through sensors and interpret the data collected by these sensors to understand the environment better.

- Language understanding: AI systems can process and generate human language, enabling them to communicate effectively with humans.

- Deep Learning: A subfield of machine learning (ML), deep learning uses artificial neural networks consisting of layers of interconnected nodes or neurons. The network learns from large amounts of data, allowing it to recognize patterns and make predictions based on that data.

It's important to note that AI is still in its early stages, with many challenges remaining, such as explaining how decisions are made (explainability) and ensuring ethical behavior (AI ethics). However, the rapid advancements in AI have led to breakthroughs in various fields, including healthcare, finance, transportation, and entertainment.

You: hello I am Sam what's your name

Assistant: Hello, Sam! I'm an assistant designed to help answer questions and engage in conversations. You can ask me all kinds of things, like trivia, explanations about complex topics, or just chat for fun. How can I assist you today?

Start Listening

Stop Listening

Exit

THANK YOU!

FOR YOUR ATTENTION