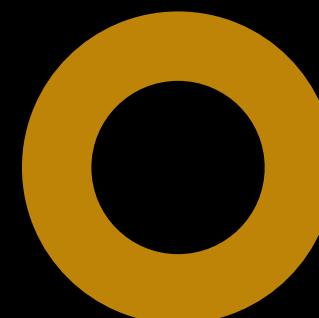
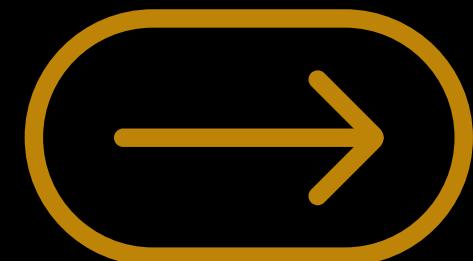




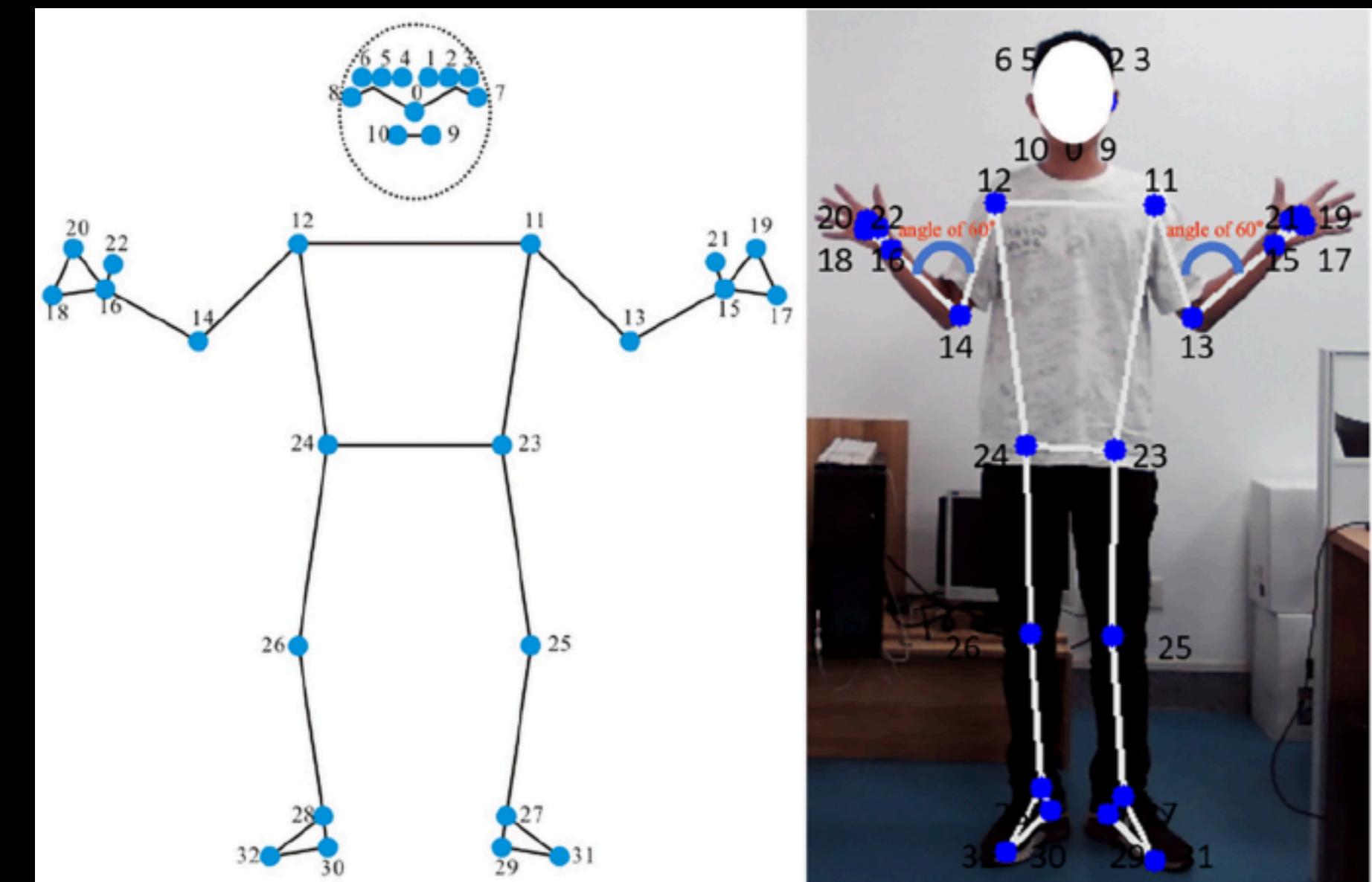
MEDIPIPE

Kalybek Aruzhan



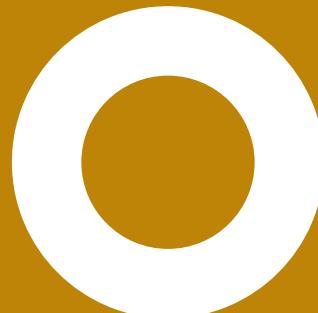
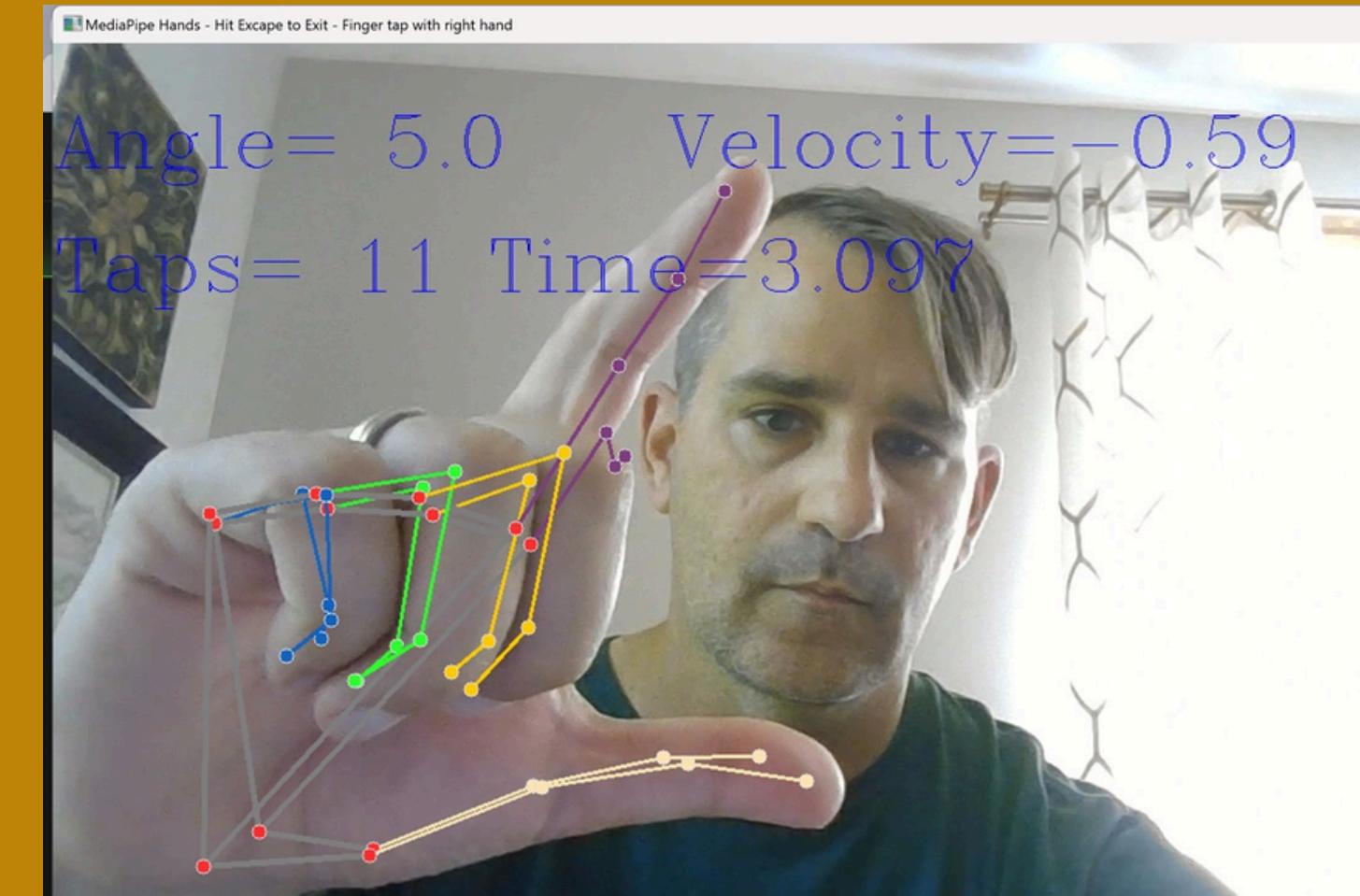
What is Mediapipe?

MediaPipe is an open-source framework developed by Google for building cross-platform machine learning pipelines, primarily focused on real-time computer vision and AI-based applications. It provides pre-built solutions for tasks like hand tracking, face detection, object tracking, and more.



Why is it so important?

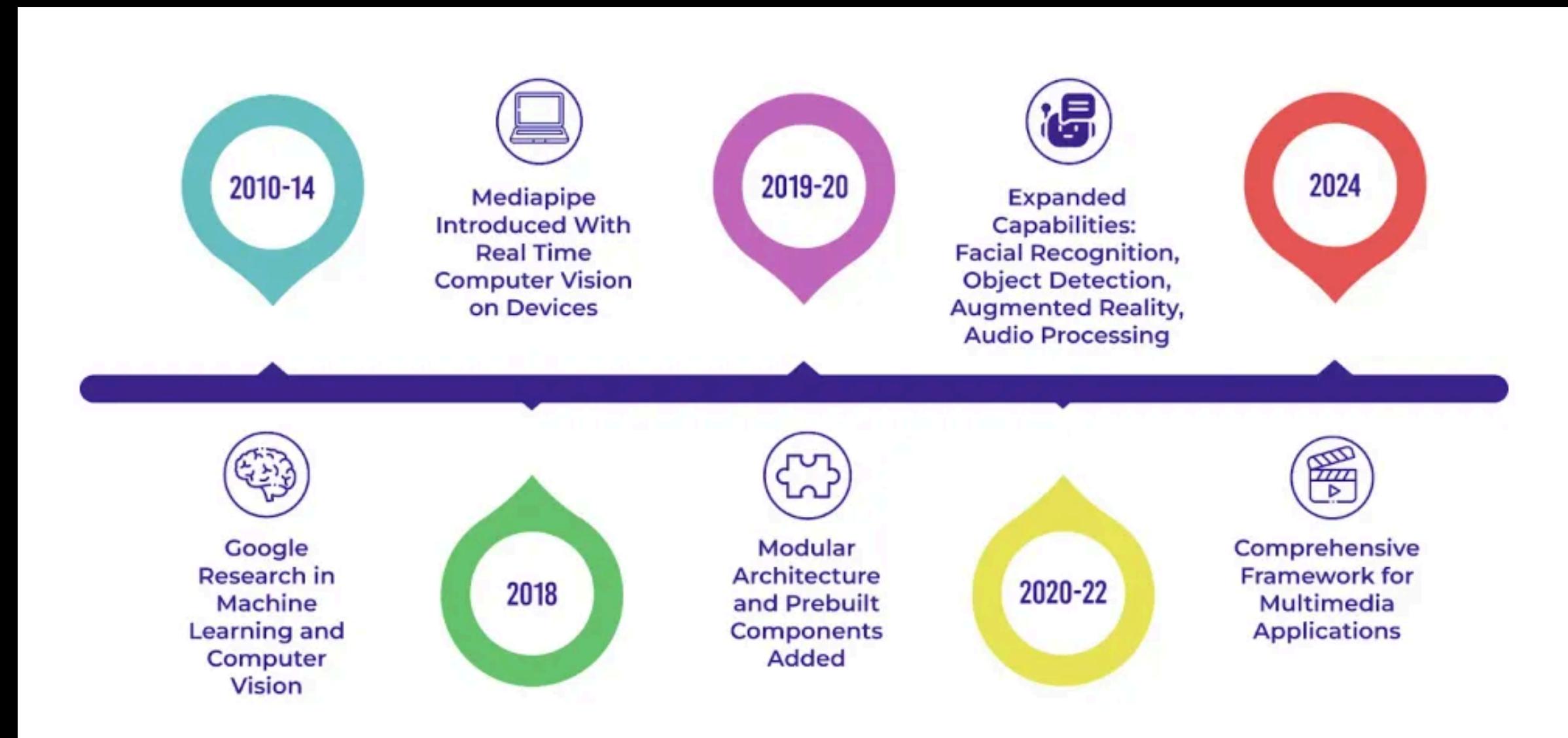
MediaPipe is a powerful tool for computer vision, offering pre-built, optimized solutions that simplify development and improve processing efficiency. Its modular architecture allows for easy customization and seamless integration with various technologies, making it ideal for augmented reality, healthcare, and other AI-driven applications.



MediaPipe Over the Years

MediaPipe's roots go back to the early 2010s when Google was working on improving machine learning and computer vision technologies. It was first used in 2012 to analyze video and audio in real-time on YouTube.

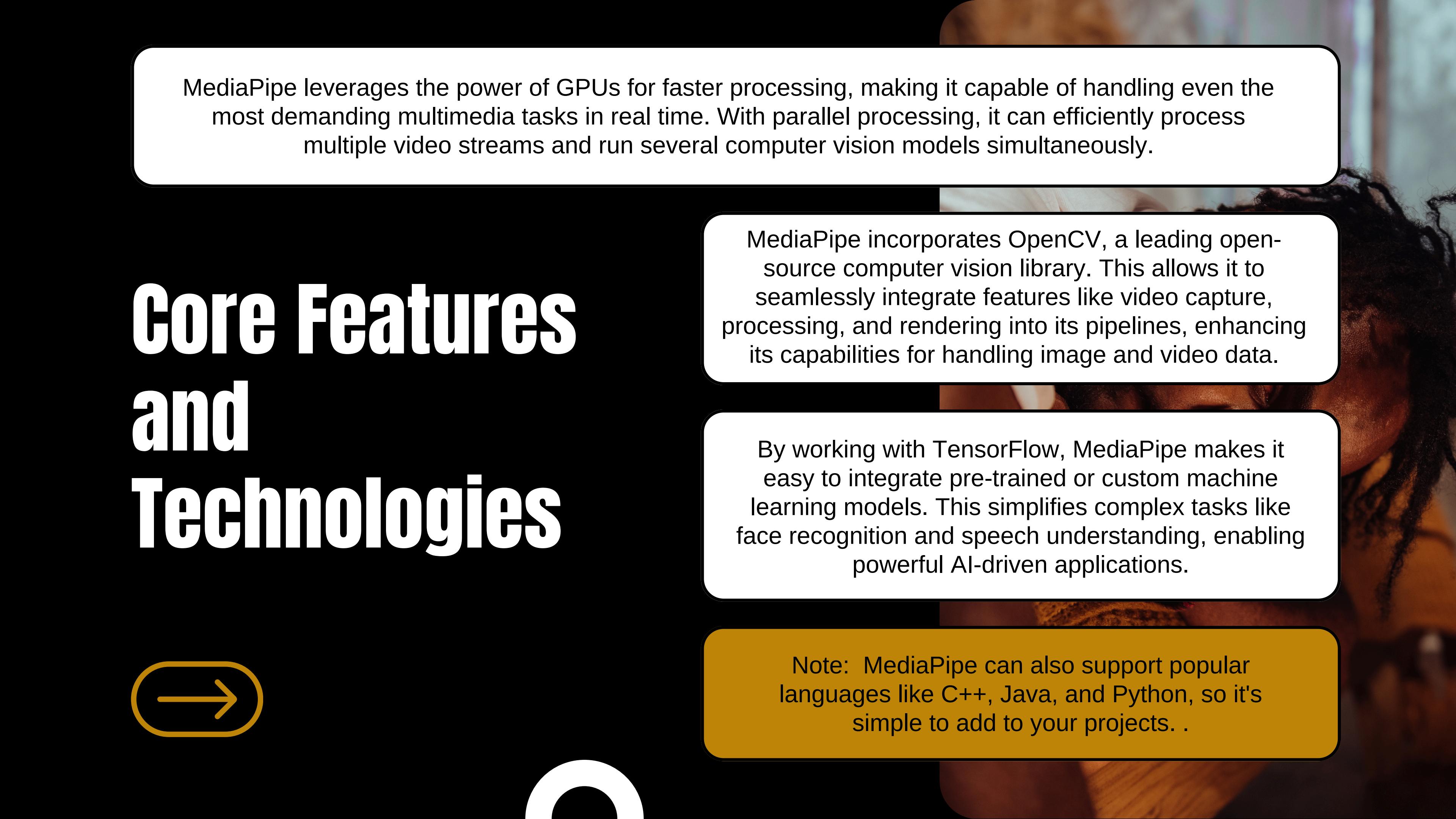
In 2018, MediaPipe began solving problems related to using complex computer vision models on devices like smartphones and small computers. By 2020, there was a growing need for a fast and efficient way to process multimedia, so Mediapipe was updated with the same. Now, Mediapipe remains a strong framework for developers who want to create innovative multimedia apps that work really well.



Core Features and Technologies

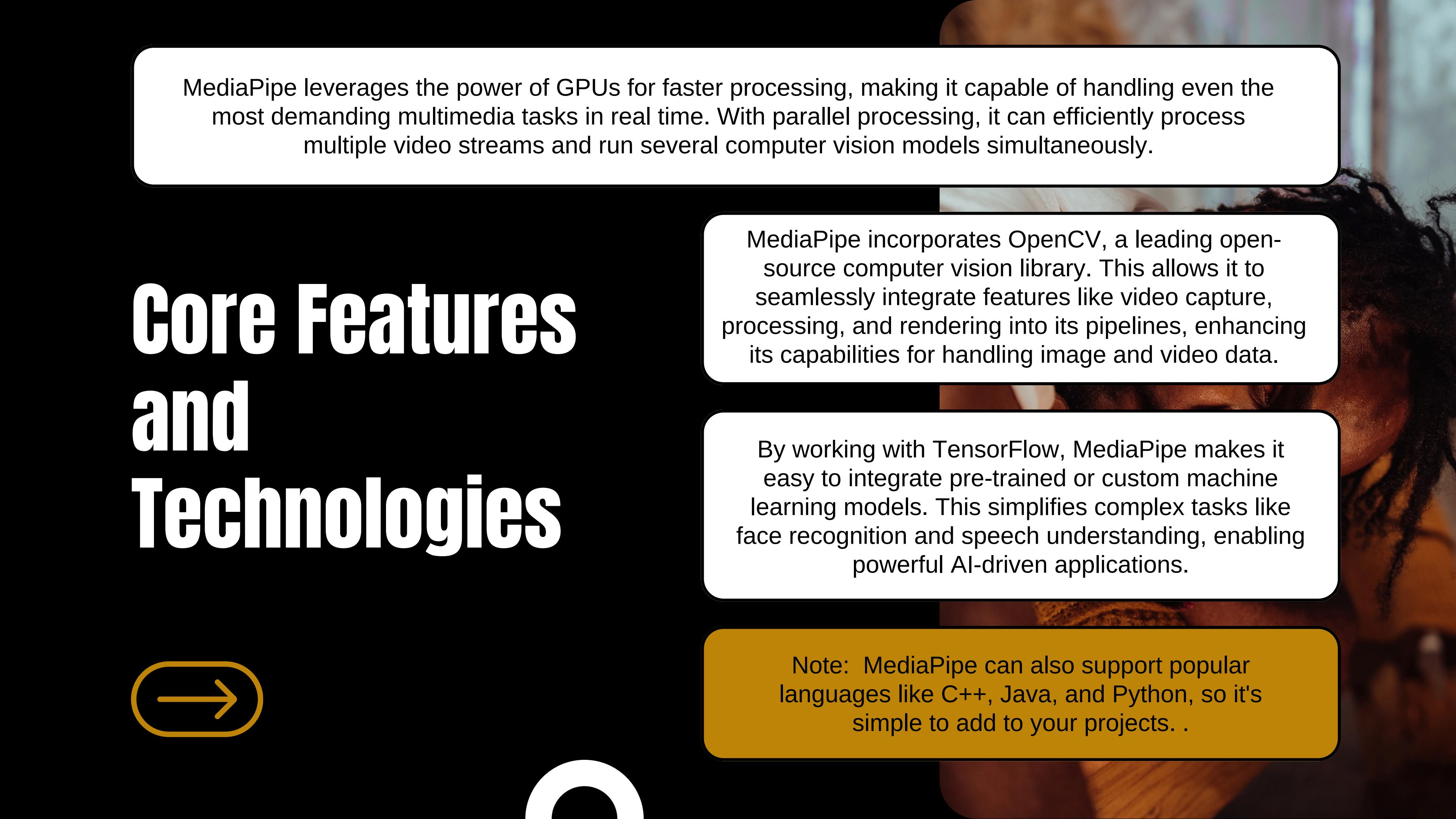


MediaPipe leverages the power of GPUs for faster processing, making it capable of handling even the most demanding multimedia tasks in real time. With parallel processing, it can efficiently process multiple video streams and run several computer vision models simultaneously.



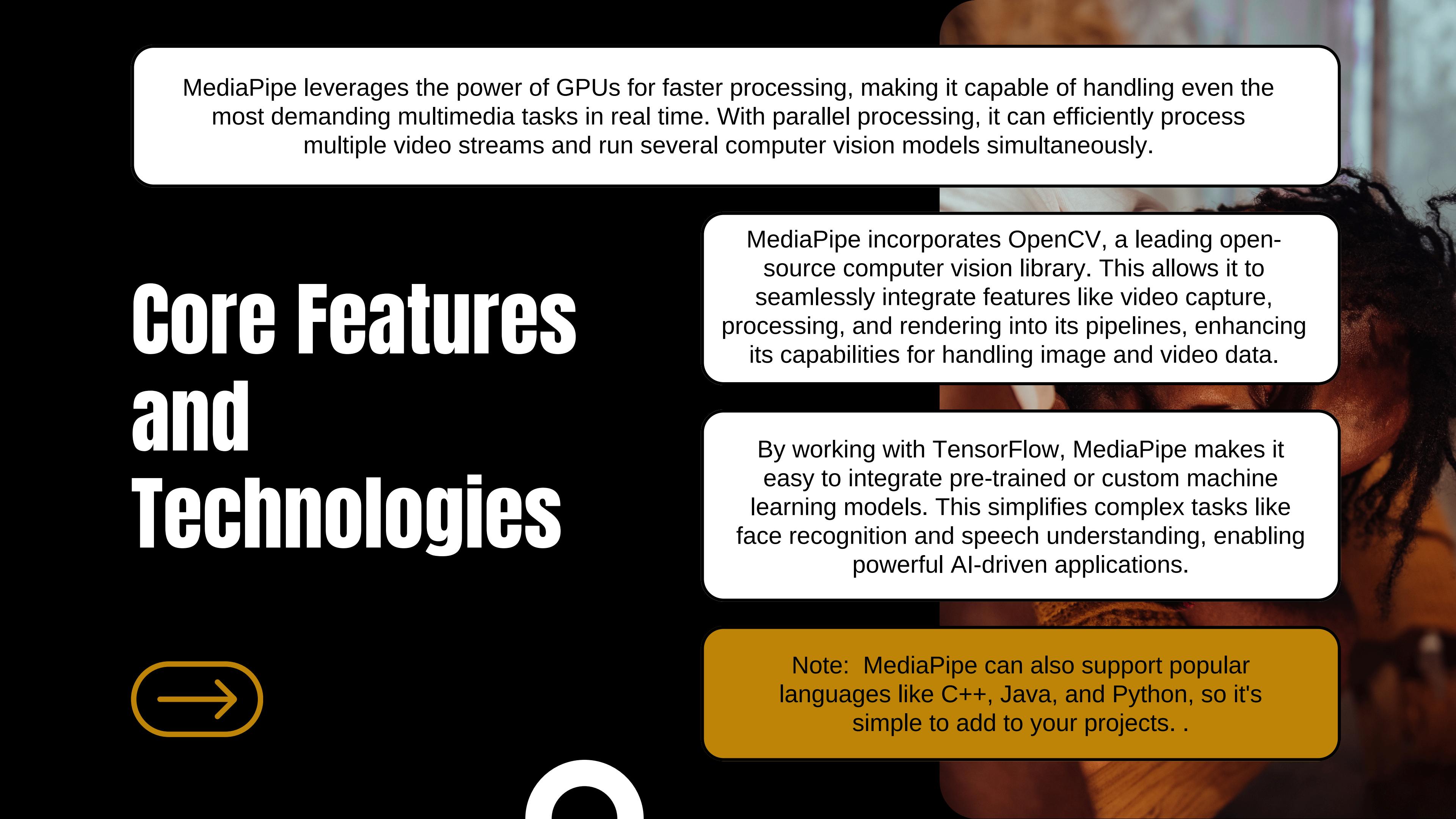
A close-up photograph of a person's face, showing their eyes and forehead. The lighting is warm and focused on the eyes, creating a sense of depth and engagement. The background is blurred, drawing attention to the subject's eyes.

MediaPipe incorporates OpenCV, a leading open-source computer vision library. This allows it to seamlessly integrate features like video capture, processing, and rendering into its pipelines, enhancing its capabilities for handling image and video data.



A close-up photograph of a person's face, showing their eyes and forehead. The lighting is warm and focused on the eyes, creating a sense of depth and engagement. The background is blurred, drawing attention to the subject's eyes.

By working with TensorFlow, MediaPipe makes it easy to integrate pre-trained or custom machine learning models. This simplifies complex tasks like face recognition and speech understanding, enabling powerful AI-driven applications.



A close-up photograph of a person's face, showing their eyes and forehead. The lighting is warm and focused on the eyes, creating a sense of depth and engagement. The background is blurred, drawing attention to the subject's eyes.

Note: MediaPipe can also support popular languages like C++, Java, and Python, so it's simple to add to your projects. .



Core Features



✓ **Pre-Trained Models** - Offers ready-to-run models to facilitate quick integration into applications

✓ **Evaluation and Benchmarking** - Aids in visualizing, evaluating, and benchmarking solutions directly in the browser

✓ **Efficient On-device Processing:** MediaPipe is optimized for on-device machine learning, ensuring real-time performance without relying on cloud processing.

✓ **Customization with MediaPipe Model Maker** - Allows tailoring models for solutions with specific data

Mediapipe consists of three main structural components: calculators, input/output packets, and computation graphs.

How MediaPipe Works?

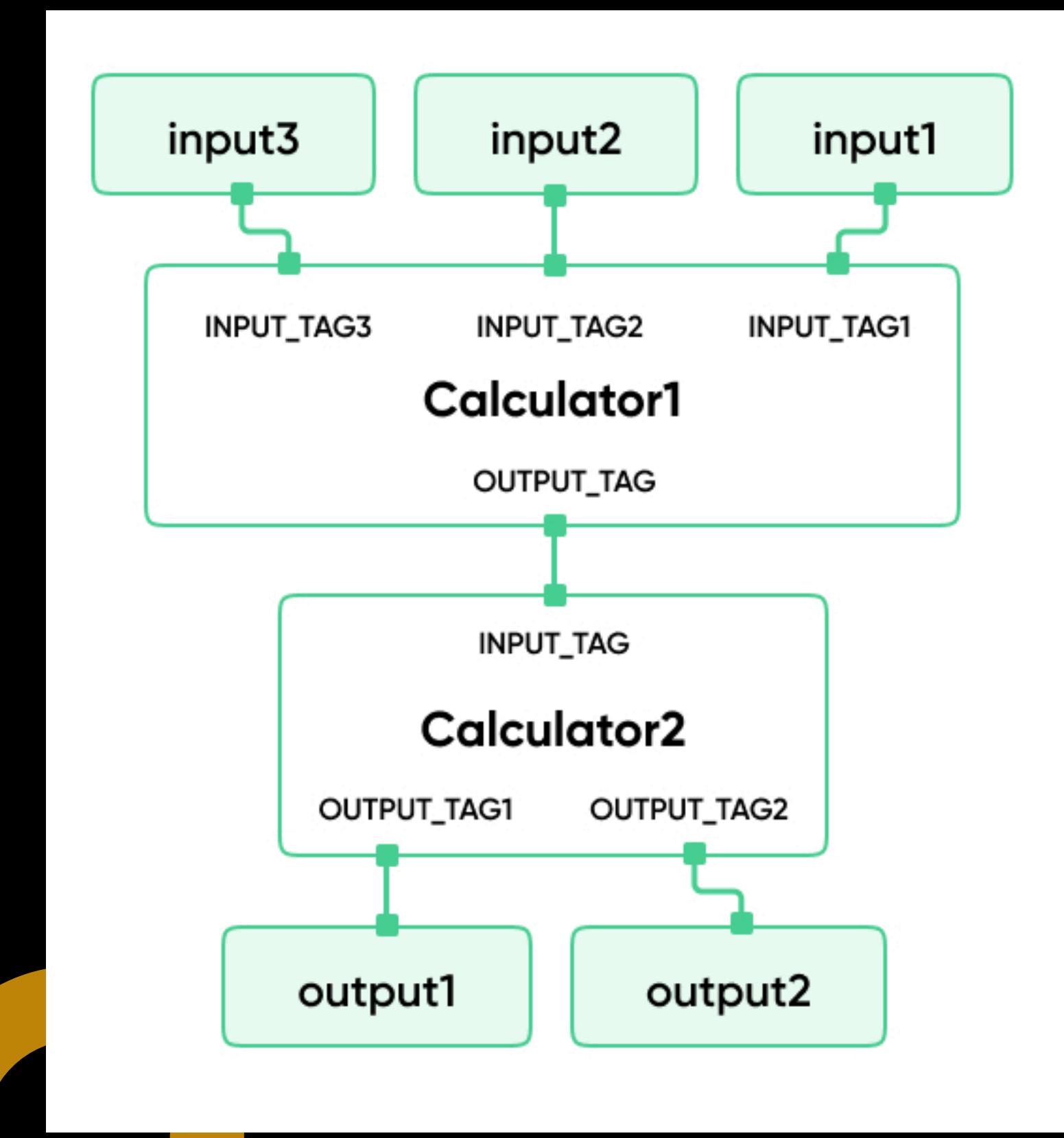
A **calculator** is a separate block that receives data, processes it, and passes it on. It performs a specific task, such as face recognition or motion tracking.

Example:

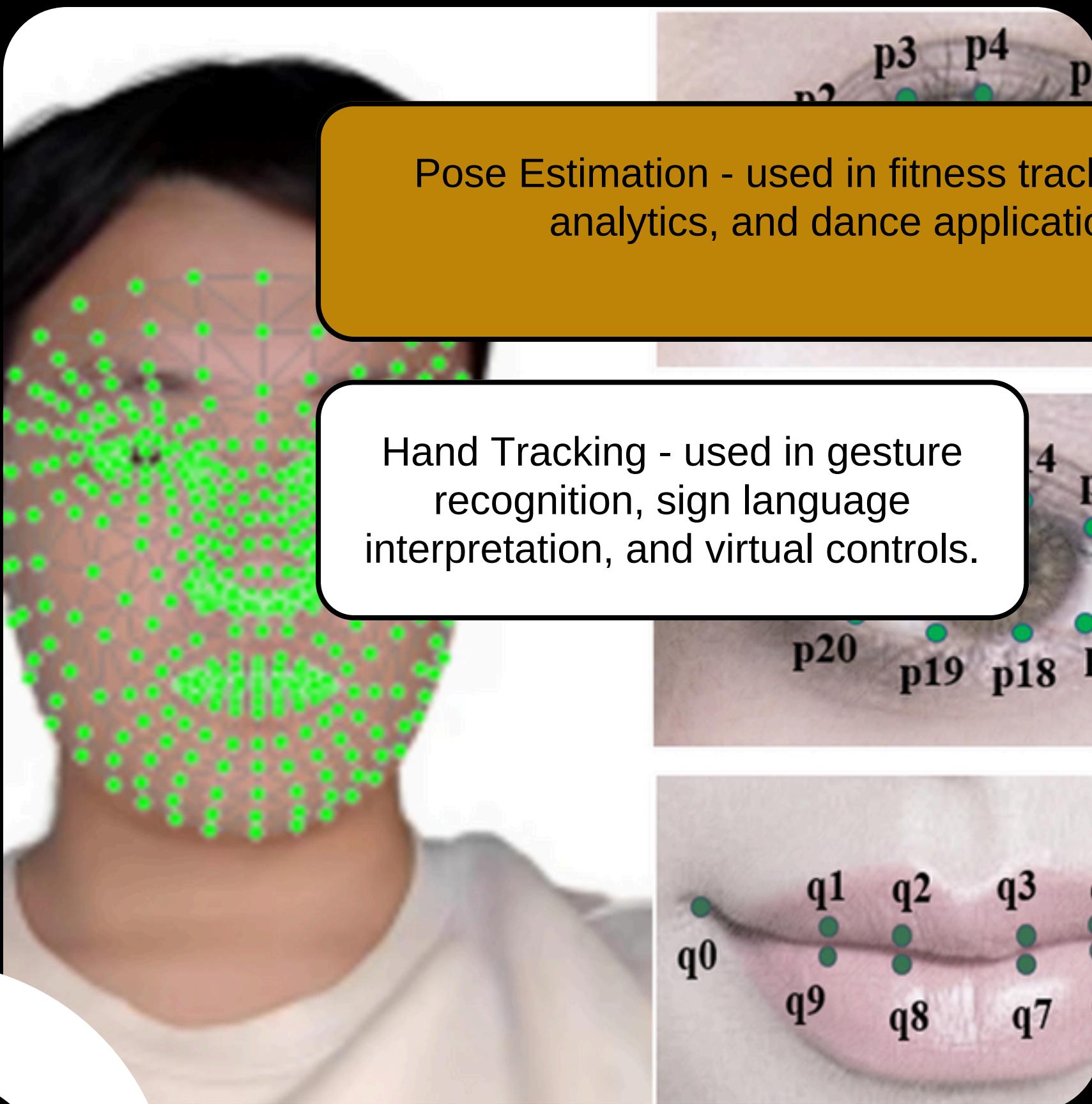
- In Open(), we load the model.
- In Process(), we take an image, run the model on it, get the result, and pass it further.
- In Close(), we free up memory.

A **computation graph** is a scheme that connects input data and calculators, creating a complete processing pipeline.

- It is written in TensorFlow Graph Text format (.pbtxt file).
- This graph defines which data goes where and which calculators process it.



Popular Solutions in MediaPipe



Face Detection - used in Face Mesh, Face Stylization, and AR Filters.

Hand Tracking - used in gesture recognition, sign language interpretation, and virtual controls.

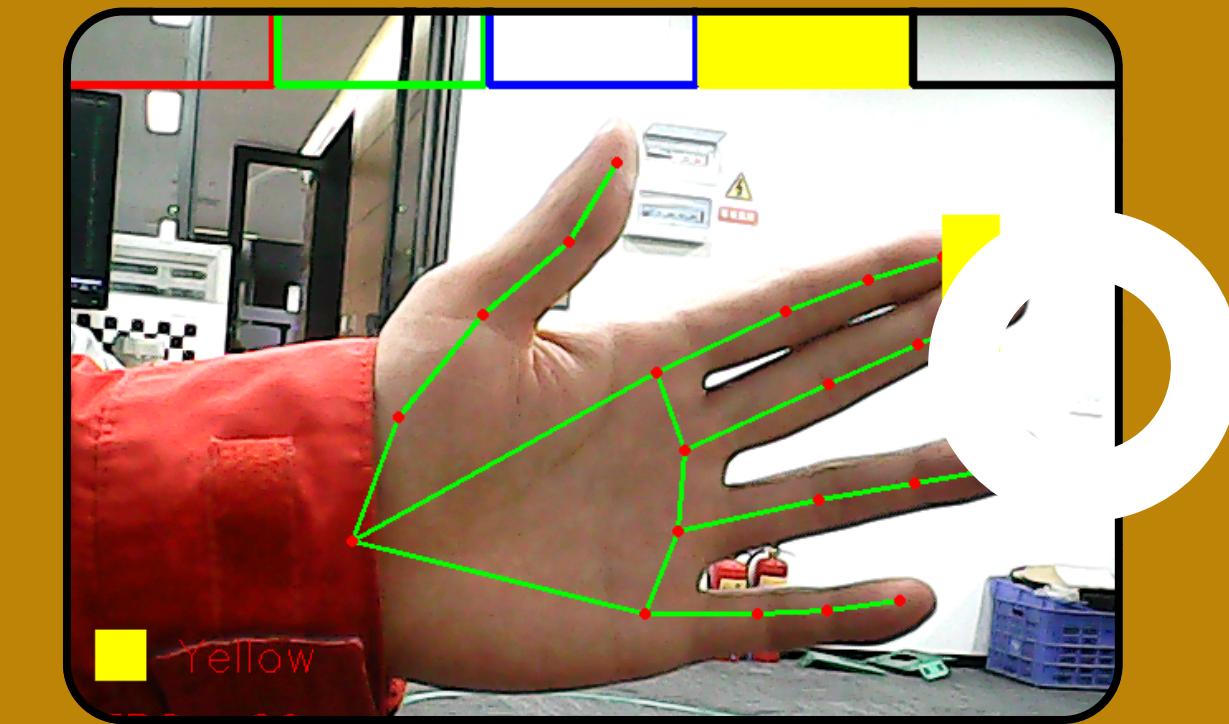
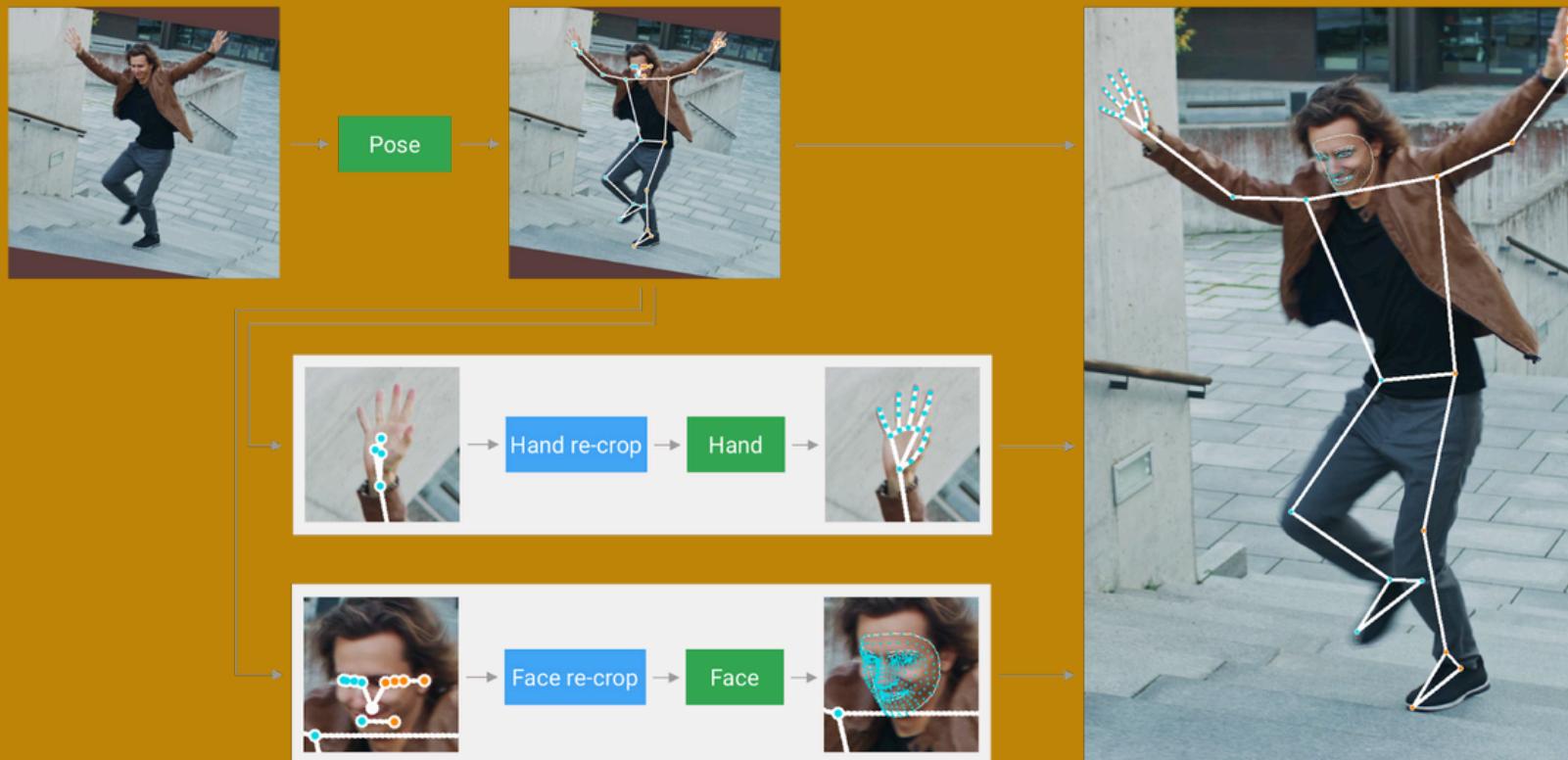
Pose Estimation - used in fitness tracking, sports analytics, and dance applications.

Popular Solutions in MediaPipe

3D Object Detection - detects and tracks 3D objects like shoes, cups, and books

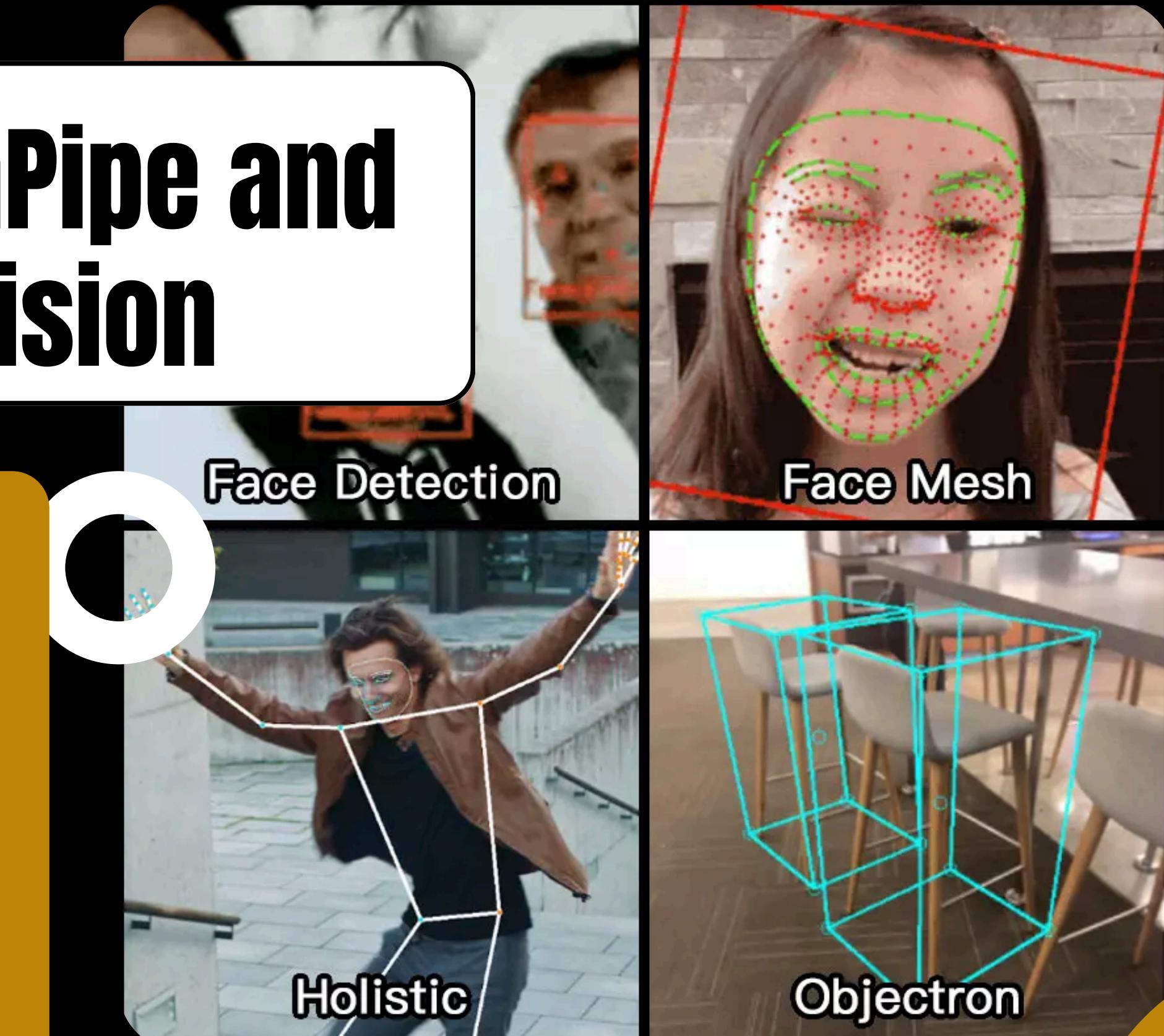
Selfie Segmentation - used for video conferencing effects, virtual backgrounds, and AR apps.

Text Detection - used in OCR (Optical Character Recognition) applications.



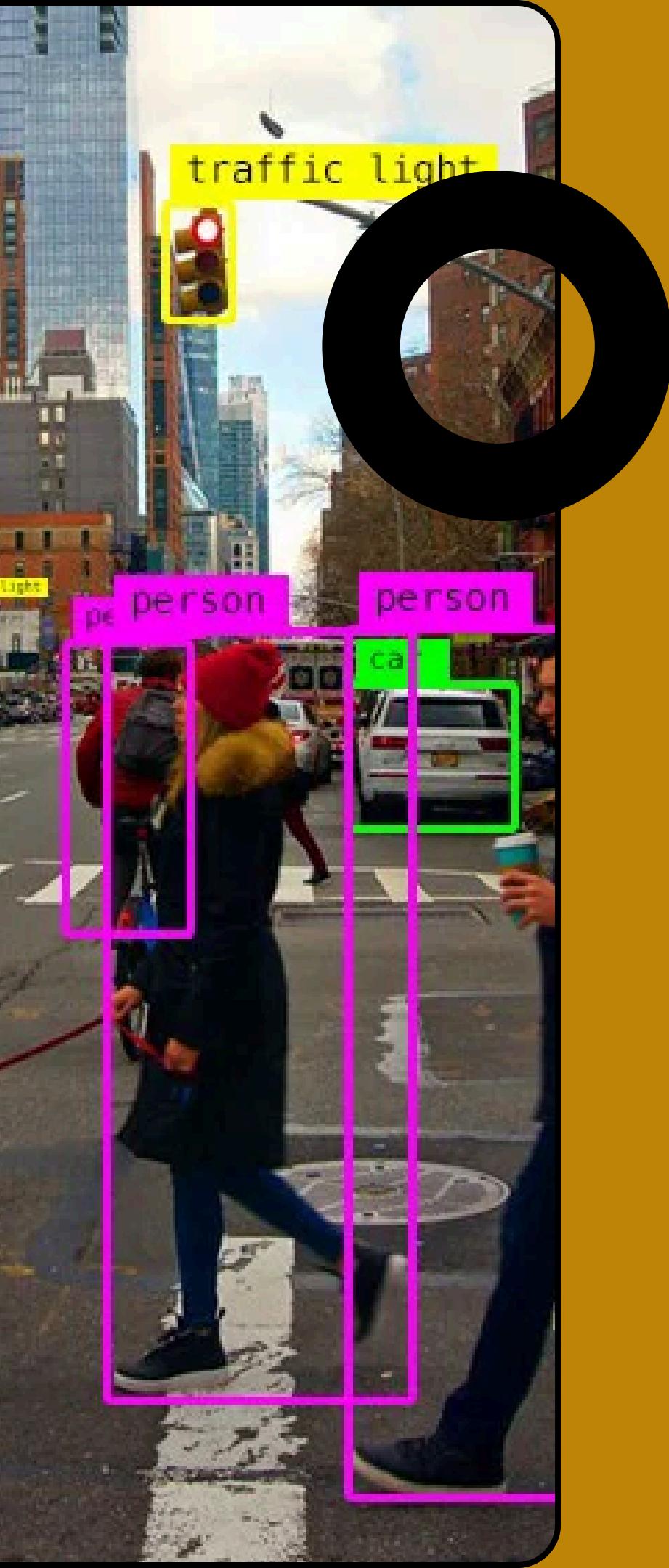
Future of MediaPipe and Computer Vision

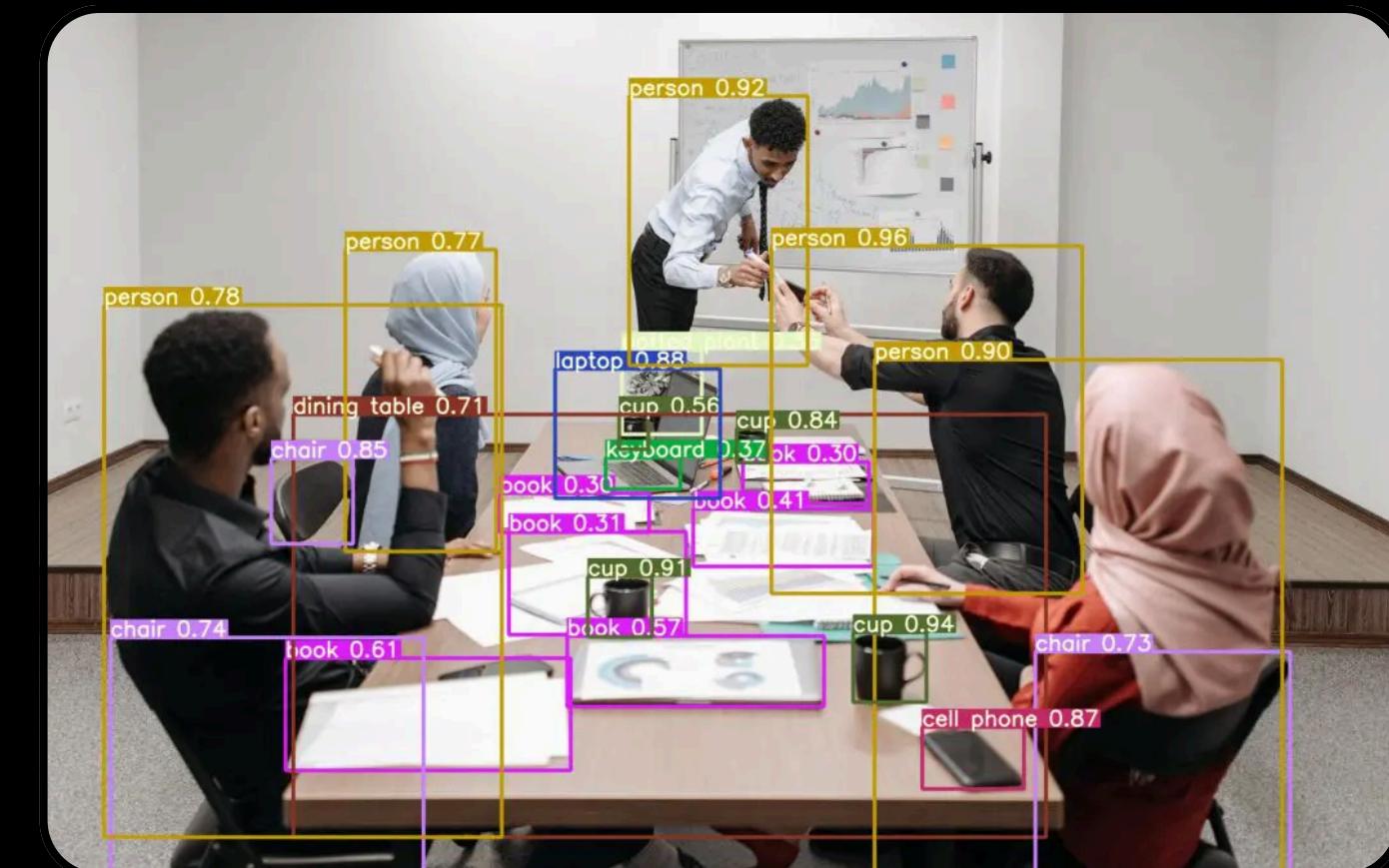
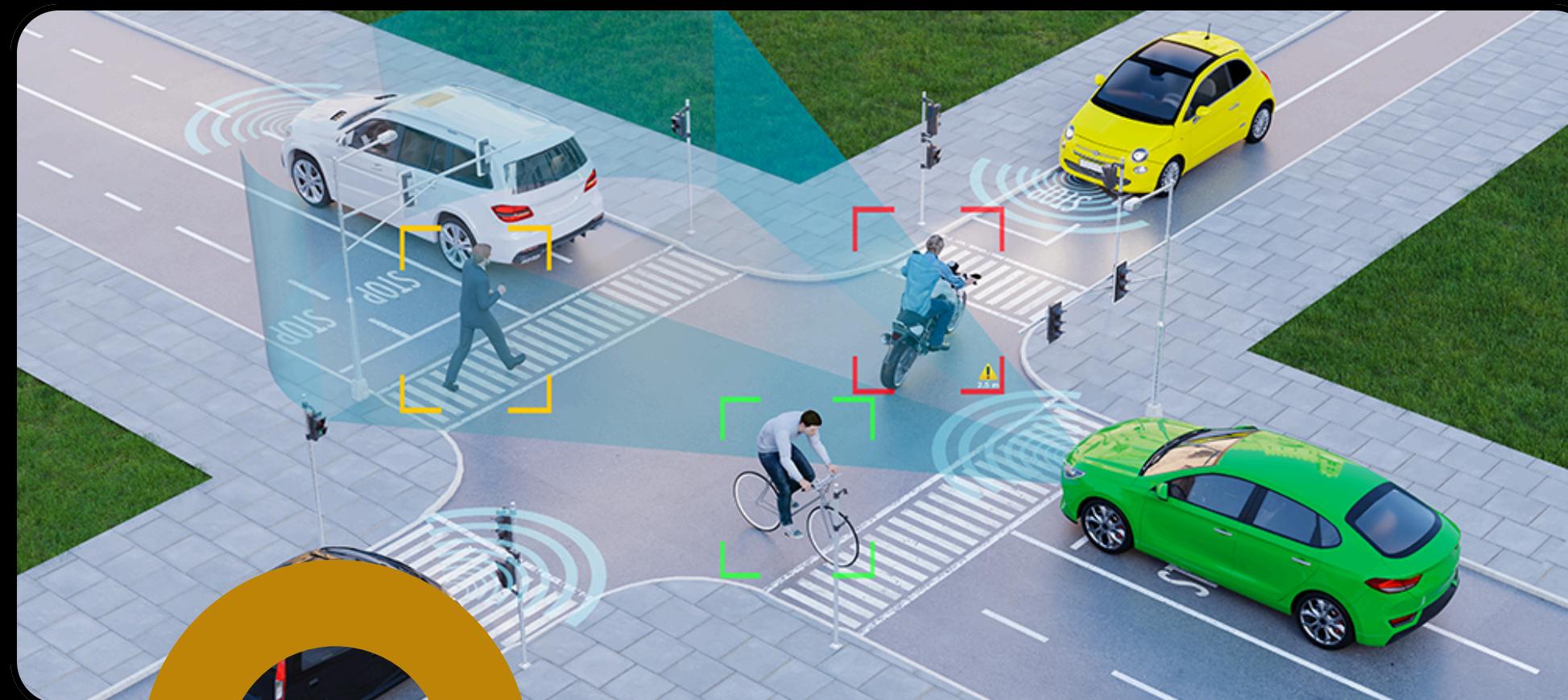
The future of MediaPipe and computer vision holds so much potential. With the rapid advancement of AI and increasingly powerful hardware, new innovations continue to emerge. For instance, as research in pose estimation progresses, we are witnessing its growing integration into MediaPipe. A notable example is the addition of 3D pose estimation in 2021 following the release of the BlazePose model. Given this trend, it is highly likely that we will see further enhancements and new features in the coming years.



Conclusion

Over the years, MediaPipe has significantly evolved, expanding its capabilities in multimedia processing. Whether it is gesture control, facial recognition, pose estimation, or object tracking, the framework provides powerful tools to bring these ideas to life. As technology continues to advance, MediaPipe will remain a valuable resource, enabling developers to explore new AI applications and push the boundaries of innovation.





Thank You