

Pose-Tracking Model Decision Worksheet

AI for Video Engineering · Lesson 2.7 — Five questions, three topologies, four pitfalls — pick the right pose model in ninety seconds.

Five-question decision tree

1. Is this a commercial product?

- ☐ **Yes** → do not ship OpenPose. License is non-commercial.
- ☐ **No** (research / internal) → any family is fine.

2. Where does the model run?

- ☐ Browser / user device → **MediaPipe Pose**.
- ☐ Edge box near camera → **RTMPose-m / -s**.
- ☐ Cloud server → continue to question 3.

3. How many concurrent streams?

- ☐ > 10 / server → **RTMPose-m on CPU** (90+ FPS).
- ☐ ≤ 10 / server, max accuracy → **ViTPose-L on GPU**.

4. Need whole-body keypoints?

- ☐ Hands + face landmarks → **RTMW** (133 keypoints).
- ☐ Just COCO 17 / BlazePose 33 → smaller model.

5. Real 3D measurements?

- ☐ Clinical-grade → **multi-camera triangulation rig**.
- ☐ UX feedback only → MediaPipe 3D / RTMW3D.

Deployment topology — default model

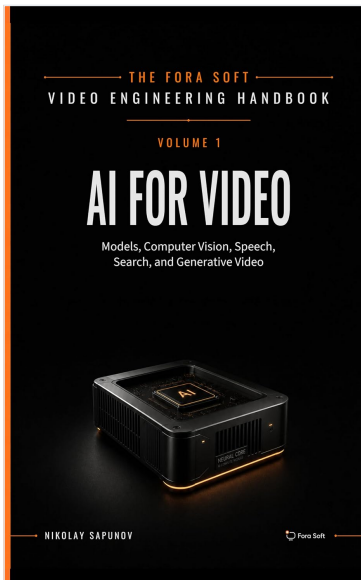
Topology	Model	Cost owner	Latency
Browser	MediaPipe Pose	User device	~16 ms
Edge	RTMPose-m / -s	Hardware	~50 ms
Server CPU	RTMPose-m	Cloud / stream	~30 ms
Server GPU	ViTPose-L	Cloud / stream	~20 ms

Four pitfalls — pre-launch checklist

- ☐ **License:** OpenPose *not* shipped in a paid product.
- ☐ **Smoothing:** One-Euro filter applied per keypoint.
- ☐ **Frame rate:** pose runs ≤ 10 FPS for slow-motion features.
- ☐ **3D:** single-camera 3D is UX feedback, not measurement.

Pipeline stages

- ☐ **1 • Detection** — RTMDet-tiny or YOLOv8-person.
- ☐ **2 • Pose** — one of the four families.
- ☐ **3 • Tracking** — ByteTrack on bounding boxes.
- ☐ **4 • Smoothing** — One-Euro filter per keypoint.



GO DEEPER · VOLUME 6 OF 7

AI for Video: A Complete Guide to Intelligence in Video Products

From Models and Inference to Computer Vision, Speech, Search, and Generative Video

By Nikolay Sapunov, CEO at Fora Soft · 1,444 pages · Kindle & paperback on Amazon

[Read it on Amazon →](#)
amazon.com/dp/B0HH6TKPZ1

This worksheet picks a pose model in five questions, maps deployment topologies and lists four pre-launch pitfalls. Volume 6 covers the computer vision behind it in depth: detection, tracking, segmentation and pose estimation, model artifacts and deployment topologies from browser to GPU server.

WHAT YOU WILL FIND INSIDE

- Pose estimation model families
- Edge, browser and server deployment
- Detection and tracking before pose
- Model formats and licensing checks

THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI

Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF

Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery

Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing

Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs

Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video

Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

Building fitness, rehab or sports features on pose tracking?

We will choose the pose model for your devices, build the detection, tracking and smoothing pipeline, and hit your latency target. 250+ video products shipped since 2005.

[Book a 30-min call →](#)
[WhatsApp →](#)
[Email us →](#)