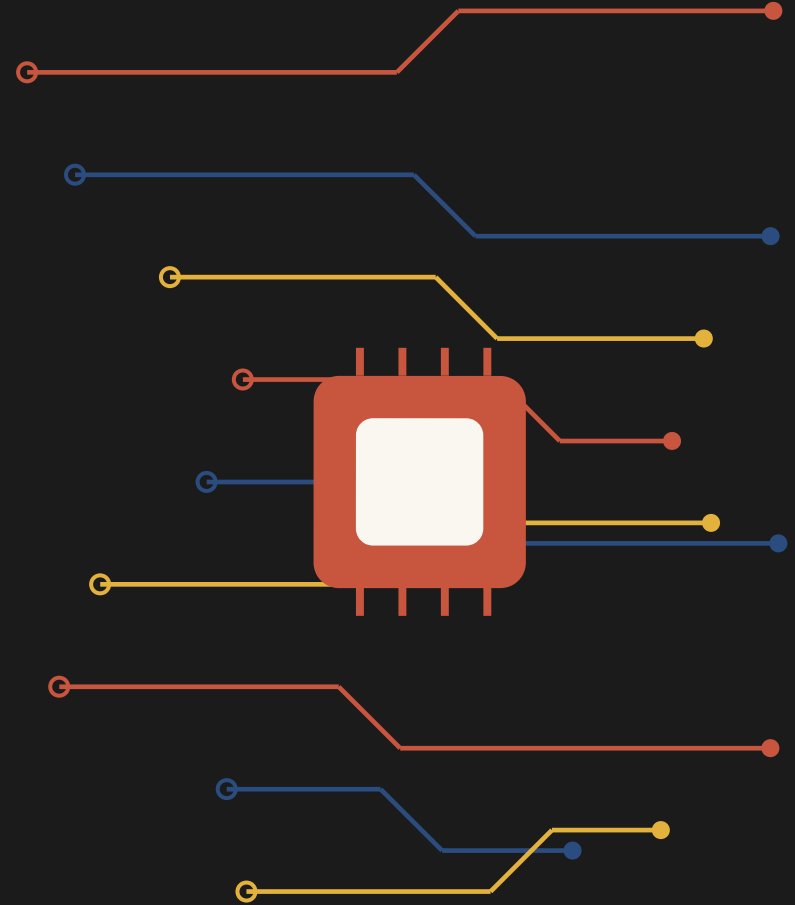


TECH TALK

Edge AI on Devices

Running AI models on phones and sensors

Presenter Name · 2026 · hello@yourdomain.in



Agenda

01 Introduction

02 Architecture

03 Trends

04 Adoption





**The fastest AI is the one
that never leaves your
device.**

Why it matters

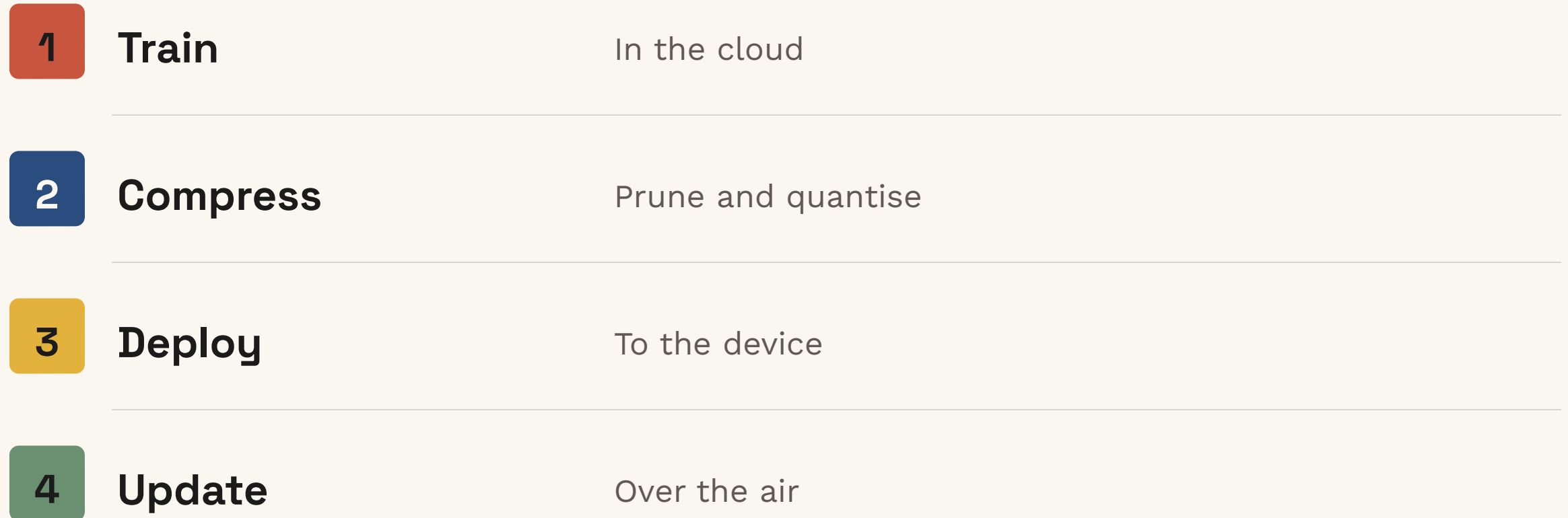
Overview

- Edge AI runs models directly on devices
- It cuts latency and keeps data private
- Used in phones, cameras, cars and factories
- Small, efficient models make it possible

KEY IDEA

**Quantised models
can be four times
smaller.**

How it works

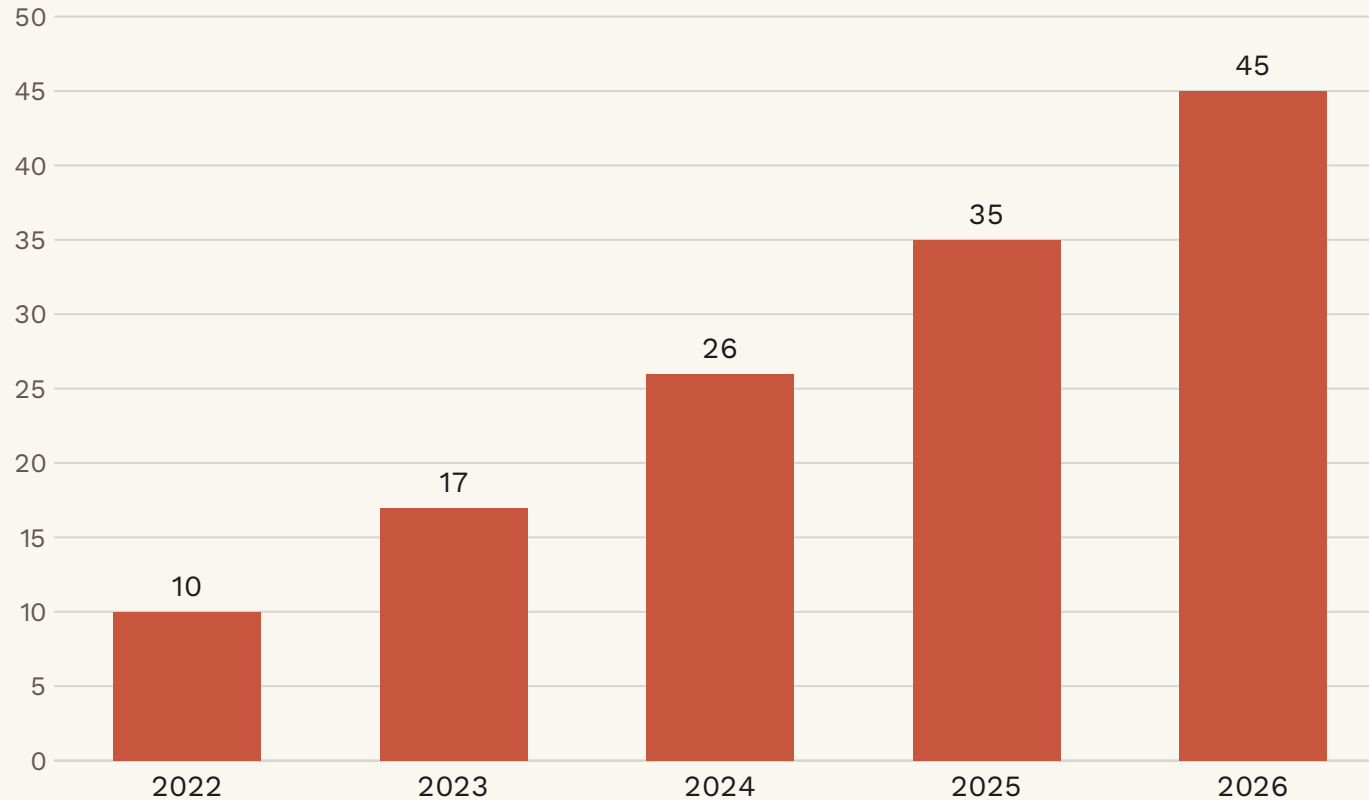


Tools and options

Layer	Popular choices	Notes
Runtimes	Mobile and embedded runtimes	Optimised inference
Hardware	NPU, GPU, MCU	Match to power budget
Techniques	Quantisation, pruning	Smaller, faster models

Generic examples; pick what fits your project.

Adoption over time



Sample data for illustration — replace with your own figures.

Adoption keeps rising as tools mature (sample data).

Use cases



Image 1

Smartphones

Photo and voice features



Image 2

Security

On-camera detection



Image 3

Industry

Defect inspection

Benefits and risks

Benefits

- Low latency
- Privacy by design
- Works offline

vs

Risks

- Limited compute
- Model updates are harder
- Battery impact

Best practices



Measure first

Profile on real devices.



Compress smartly

Check accuracy after.



Plan updates

Version your models.

Questions?

Thanks for your time.

Presenter Name

hello@yourdomain.in

