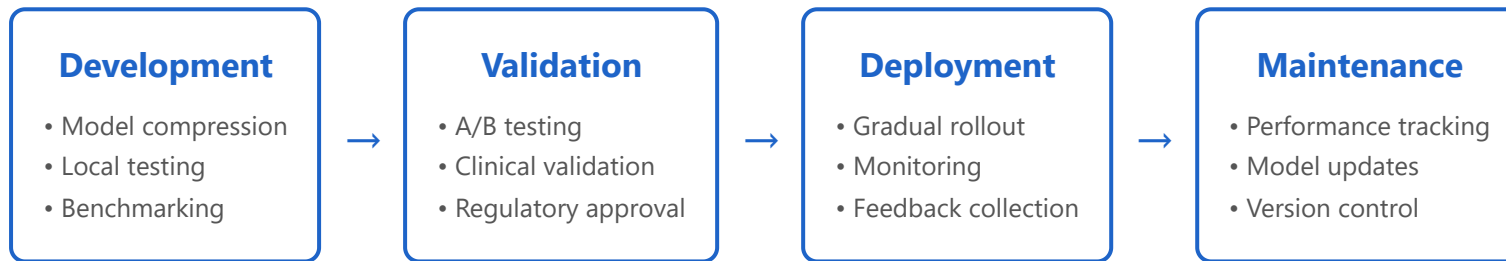


# Deployment Strategies

## Deployment Strategies and Checklist

Strategies for successful production deployment of compressed models

### Deployment Workflow



### Pre-Deployment Checklist

#### Performance

- ☐ Accuracy target achieved
- ☐ Latency requirements met
- ☐ Within memory constraints

#### Reliability

- ☐ Edge case testing
- ☐ Error handling
- ☐ Fallback mechanism

#### Regulatory

- ☐ FDA/CE certification (if required)
- ☐ Privacy protection
- ☐ Explainability

#### Operations

- ☐ Logging/monitoring
- ☐ OTA updates
- ☐ User training

# Recommended Strategies by Platform

Platform	Compression Method	Framework
iOS	Distillation + INT8	Core ML
Android	Distillation + INT8	TF Lite
Web	Lightweight compression	TensorFlow.js
Embedded	INT4 + Pruning	TF Lite Micro