# LLM launch checklist for a mobile app

Release: _____ Date: _____ Owner: _____ Sign-off: _____

## 1. Scope

- ☐ The job is written down: what the feature must do, must never do, and who it is for
- ☐ Success measures and launch limits agreed with the product owner

## 2. Architecture

- ☐ Model calls run on our server, not from the app
- ☐ Provider keys and full prompts are not in the app package
- ☐ Prompts and settings are server-side configuration
- ☐ Model sits behind an abstraction so it can be swapped
- ☐ Fallback is defined for when the model is down

## 3. Evaluation

- ☐ Fixed test set of real questions (a common starting point is 50 to 100)
- ☐ Rubric written down: correct, grounded, right tone, refuses when it should
- ☐ Sample reviewed by people, not only automated scoring
- ☐ Pass marks agreed in advance and met
- ☐ Test set re-run after every prompt, content or model change

## 4. Guardrails

- ☐ Input checks: length, banned content, instruction override attempts
- ☐ Output checks: format, safety, topic limits
- ☐ Model output is not passed to other systems unchecked
- ☐ Actions need user confirmation and use least privilege
- ☐ Users can report a bad answer and reach a person
- ☐ App attestation in place (Apple App Attest, Google Play Integrity) as one anti-abuse signal

## 5. Latency

- ☐ First words appear quickly (aim for about 1 second); progress is shown from the start
- ☐ Long requests can be cancelled or run in the background (about 10 seconds is the attention limit)
- ☐ Tested on real phones and real networks
- ☐ Timeouts and retries set

## 6. Cost

- ☐ Message, history and reply length capped
- ☐ Rate limits per user and per day
- ☐ Spend alerts on the provider account
- ☐ Cost per conversation measured in testing and within budget

## 7. Privacy and law

- ☐ Lawful basis and privacy notice cover the feature
- ☐ Processor contract in place with any outside provider
- ☐ Data transfers outside the UK checked
- ☐ Impact assessment considered; decisions about people checked for automated decision-making rules
- ☐ EU users: AI transparency duties checked
- ☐ Logs minimised, with a retention period

## 8. Release and monitoring

- ☐ Staged rollout with a server-side kill switch tested in staging
- ☐ Model version pinned where the provider allows
- ☐ Weekly sample of live conversations scored
- ☐ Incident plan: who is told, who switches it off

General information, not legal advice. Nordstone, nordstone.co.uk