

LLM Requirements Checklist

Section A: Business Use Cases

A1. Primary Use Cases

Describe the top use cases for the LLM and who they are for.

- **Use case 1:** *e.g., Customer support assistant*
- **Primary users:** *e.g., Support agents / Customers*
- **Problem today:** _____
- **What success looks like:** _____
- **Use case 2:** *e.g., Document summarisation for analysts*
- **Primary users:** _____
- **Problem today:** _____
- **What success looks like:** _____
- **Use case 3 (optional):** _____
- **Primary users:** _____
- **Problem today:** _____
- **What success looks like:** _____

A2. Success Metrics

Define how you will measure success for each use case.

Quantitative metrics (pick a few)

- Target accuracy / resolution rate: _____
- Time saved per task or ticket: _____
- Reduction in human escalations: _____
- Cost per conversation / request target: _____
- Other metric (e.g., conversion / engagement): _____

Qualitative metrics

- User satisfaction goal (e.g., CSAT / NPS): _____
- Perceived trust and usefulness notes: _____
- "Would users be disappointed if this tool disappeared?" yes / no

Section B: User Experience & Workload

B1. Interaction Model

Clarify how people will interact with the LLM.

- Interaction type:
 - Real-time chat / assistant
 - Async / batch processing

- Internal sidekick / copilot in another app
- Ideal response time (seconds): _____
- Maximum acceptable response time (seconds): _____

B2. Volume & Scale

- Expected requests per day (initial): _____
- Expected peak requests per second (RPS): _____
- Projected 6–12 month growth (e.g., ×2, ×5): _____
- Critical peak periods (e.g., business hours, launches): _____

B3. Context Needs

- Maximum input size (e.g., pages / tokens / length): _____
- Need long multi-turn chat history? yes / no
- Using RAG or external tools that increase context? yes / no

Section C: Constraints & Readiness

C1. Budget and Timeline

- Approximate monthly budget for LLM (all-in): _____
- Target date for first production launch: _____
- Project type:
 - Pilot / experiment
 - Business-critical production system

C2. Security, Privacy & Compliance

- Industry (e.g., healthcare, finance, SaaS): _____
- Contains PII? yes / no
- Contains PHI / financial / legal data? yes / no
- Data residency requirements (regions): _____
- Data must not be used to train external models? yes / no
- Need audit logs for prompts / outputs? yes / no
- Deployment preference:
 - Public cloud API
 - Private VPC
 - On-prem / self-hosted

C3. Team Skills & Operations

- Backend / API engineering capacity: low / medium / high
- MLOps / data engineering capacity: low / medium / high
- Security / governance support: low / medium / high
- Capacity for ongoing prompt / system iteration? yes / no
- Capacity for monitoring quality and cost? yes / no
- Plan for user training and change management? yes / no