Session 3 Handout: Applications of Foundation Models (Domain 3, 28%)

FAST-12 AI Practitioner Certification Prep · 7formations · Instructor: Wadood Chaudhary · All examples fictional · Exam guide v1.2

Why this session matters

Domain 3 is the largest domain. It tests design choices: which model, which customization approach, how to prompt, how to evaluate, and what agents are.

Choosing a customization approach

Approach	Use when	Cost and effort
Prompt engineering	Task fits the model's general ability	Lowest
RAG (Bedrock Knowledge Bases)	Answers must use your current documents, with citations	Low to medium; update by re-syncing
Fine-tuning	Need a consistent style or narrow task the base model does poorly	Medium to high; labeled data needed
Continued pre-training	Model needs a whole new domain vocabulary	Highest; large unlabeled data

Facts that change often → RAG, not fine-tuning.

Model selection factors

Task fit, modality, quality on your evaluation set, latency, context window, cost, languages, customization support, licensing.

Prompting techniques

Zero-shot, few-shot (examples), chain-of-thought (reason step by step), role and context, output format, constraints ("say 'not in file' if missing"). Risks: **prompt injection**, jailbreaking, leaking the system prompt, poisoning of retrieved content.

RAG basics

Documents → chunks → embeddings → vector store (for example, OpenSearch, Aurora PostgreSQL with pgvector) → retrieve top matches → add to prompt → answer with citations. Most RAG errors are retrieval errors: chunking, metadata, number of results.

Evaluating

Human evaluation; automatic metrics (ROUGE for summaries, BLEU for translation, BERTScore for meaning); LLM-as-a-judge; business metrics (task completion, deflection rate). **Amazon Bedrock Evaluations** supports these for models

and knowledge bases.

Agents

An agent uses an FM to plan steps and call tools to complete a task. **Bedrock AgentCore** helps run agents securely at scale; **Strands Agents** is an open-source SDK for building them. Agents need limits on tools and human approval for risky actions.

Worked example (fictional)

Bayside's SOP assistant answered "safety glasses" for VF-4 PPE but missed hearing protection: the PPE table was split across chunks. Fix: larger chunks with overlap, then re-test the evaluation set.

Checklist

- ☐ I can choose prompt vs RAG vs fine-tune from a scenario.
- ☐ I can name three prompt-injection defenses.
- ☐ I can match ROUGE, BLEU, and BERTScore to tasks.
- ☐ I can explain what an agent adds over a single prompt.