FAST-14 Projects: How Large Language Models Work

FAST-14 How Large Language Models Work · 7formations · Instructor: Wadood Chaudhary · All companies and data are fictional

No coding is required for any project. Use the course's fictional files, or describe a task from your own job in general terms (no employer names, client data, personal information, or confidential details; your manager approves). Scoring: Not yet / Meets / Strong. The capstone needs Meets on every criterion; one resubmission is allowed. Templates are in lab-guide.md (W3 and W4).

Capstone: LLM Solution Brief with Model Selection Evidence (W3 + W4)

What you produce: the one-page W4 brief for a real task type, backed by a W3-style scorecard on at least 10 test items, presented in 3 minutes in session 4.

Rubric

Criterion	Not yet	Meets	Strong
Mechanism	Explanation wrong (for example, "it looks things up")	Explains behavior through tokens, next-token prediction, context, and training stage	Predicts a failure from the mechanism and tests it
Approach	No alternatives considered	Prompt vs. RAG vs. fine-tuning vs. tool use compared; choice justified	Cost and maintenance of each option estimated
Evidence	Leaderboard or vendor claims only	Own 10-item test with the answer key written first; 3 models scored	Long-context or "not in source" items included; reruns checked for variability
Cost and settings	No formula	Monthly cost shown with input and output tokens separated; temperature or reasoning setting justified	Caching or batch options compared
Data	Sensitive data unaddressed	MA personal information (G.L. c. 93H, 201 CMR 17.00) kept out or protected; vendor terms recorded	Local vs. hosted trade-off argued
Risks	Fewer than three	Three OWASP LLM risks, each with a matching control and a human review point	Injection tested on practice data

Example (excerpt): Cape Harbor Inn guest FAQ assistant (fictional)

> **1. Task and user.** Answer guest questions on the website from the inn's FAQ; about 3,000 questions a month.
>
> **2. Approach.** Prompt with the whole FAQ in context (about 1,950 tokens). RAG isn't needed yet because the FAQ is short; revisit if it grows past a few pages. Fine-tuning rejected: facts change each season and would need retraining.

3. Model choice. Small hosted model: 10/10 on our 10-question test (including 2 "not in the FAQ" items), the same as the frontier model, at a fraction of the cost.

4. Settings and cost. Temperature 0. Per question: 2,050 input tokens × $1/M + 150 output × $4/M ≈ $0.00265; ≈ $7.95 a month at practice prices, less with cached input.

5. Data. No reservations or payment data in the assistant; it can't see bookings. Vendor plan: inputs not used for training; 30-day retention.

6. Risks and controls. LLM01 Prompt Injection (a guest message says "ignore your rules"): treat guest text as data; answers limited to FAQ topics. LLM09 Misinformation (wrong pet policy): "not in the FAQ" rule plus a monthly spot-check. LLM02 Sensitive Information Disclosure (a guest pastes a card number): instruction to refuse, plus a pattern filter that masks card numbers.

7. Measures. 20-question test rerun after every FAQ change; target 19/20; the front-desk manager reviews 25 real questions a month.

Mini-project 1: Explain-It-Back Video or Memo (sessions 1-2)

What you produce: a 3-minute recording or a one-page memo explaining one mechanism (tokens, attention, RLHF, or why models hallucinate) to coworkers, with one example from the course files.

Criterion	Not yet	Meets	Strong
Accuracy	Contains a misconception from the course list	Correct, in plain language	Names one common misconception and corrects it
Example	None	One example from the fictional files	A live demo (probabilities or temperature)
Workplace link	None	One implication for how coworkers use AI	A specific practice change

Example (excerpt): "Why did the assistant say Bayside was founded in 1987? With no fact sheet, the model predicts text that sounds like a company history, and a confident guess looks just like a real answer. Training rewards plausible text, and it often rewards guessing over saying 'I don't know'. With the fact sheet attached and a 'not in source' rule, the invented answers in our lab dropped sharply..."

Mini-project 2: Model Card Comparison (session 3)

What you produce: a one-page table comparing the model cards and data terms of three models your employer could use (one closed, one small hosted, one open-weight), using only public vendor pages.

Criterion	Not yet	Meets	Strong
Coverage	Fewer than three models, or fields missing	License, context window, modalities, training-data statement, known limits, data use, retention, processing location	Evaluation results and safety testing summarized

Criterion	Not yet	Meets	Strong
Sources	No links or dates	Every cell cites a vendor page with the date checked	Changes since the last version noted
Recommendation	None	Which model fits which task type, and why	Conditions that would change the recommendation

Example row: "License: open-weight, Apache 2.0, so commercial use is allowed. Data use: runs locally, so no vendor retention, but 201 CMR 17.00 safeguards still apply on our own hardware. Source: vendor model card, checked October 29, 2026."