Session 2 Handout: How LLMs Are Trained and Why They Behave as They Do

FAST-14 How Large Language Models Work · 7formations · Instructor: Wadood Chaudhary · All examples fictional

Why this session matters

The same base network behaves very differently depending on how it was trained after pretraining. Knowing the pipeline tells you which model type to use and why models make things up.

The training pipeline

Stage	What it does	Resulting behavior
Pretraining (self-supervised)	Predict the next token on huge text corpora	**Base model:** continues text; doesn't reliably follow instructions
Supervised fine-tuning (SFT)	Train on example instructions and good answers	Follows instructions
Preference tuning (RLHF, DPO, constitutional methods / RLAIF)	People or AI feedback rank responses; a **reward model scores** them, or DPO learns directly from preference pairs	Helpful, polite, safer **instruct (chat) model**
Reasoning training (RL with verifiable rewards)	Reward correct answers on checkable problems (math, code)	**Reasoning model:** thinks step by step, uses more tokens and time (test-time compute)

The term "foundation model" dates from 2021 (Bommasani et al., Stanford). Chat products such as ChatGPT, Claude, and Gemini are instruction-tuned, not base models.

Scale

Scaling laws: loss falls predictably with more compute, data, and parameters; compute-optimal training balances data and model size. **Mixture-of-Experts** activates only part of the network per token. **Distillation** trains a small model to imitate a large one. "Emergent abilities" are debated: some apparent jumps are artifacts of the metric (Schaeffer et al., 2023).

Why LLMs hallucinate

- The model predicts **plausible** text; it has no built-in check against facts.
- Training and benchmarks often reward a confident guess over "I don't know".
- Facts that are rare, recent, or private (a fictional company's founding year) are not in its training data.

Reduce it: give a source (grounding), allow and require "not in source", ask for quotes or citations, lower temperature for facts, use tools for lookups and math, and have a person review high-stakes answers.

Worked example (fictional Bayside fact sheet; illustrative, your results will vary)

Question	No source	With sheet + "not in source" rule
Founded?	"1987" (invented)	"2009"
ISO registrar?	"Bureau Veritas" (invented)	"NorthMark Registrars"
Titanium?	"Yes, aerospace grades" (invented)	"No titanium"
CEO's name?	A plausible name (invented)	"Not in source"

Reference answers for Lab 2: 4.2562 in is **not** within 4.250 ± 0.005 (max 4.255); spindle-hours = 6 × 80 × 0.85 = **408**.

Checklist

- ☐ I can describe pretraining, SFT, preference tuning, and reasoning training.
- ☐ I can predict how base, instruct, and reasoning models respond to the same prompt.
- ☐ I can explain why hallucinations happen and name three controls.
- ☐ W2 log complete with at least one invented answer quoted.