What You Need to Know About LLM Training

Cutting Through the Hype to Understand AI Capabilities

Franck Albinet

franckalbinet@gmail.com

Independent Data Science & AI Consultant

October 8, 2025

Connecting to Yesterday

Why Learn About Training?

Yesterday you experienced the Three Gulfs with real prompting challenges.

Today we'll understand:

- **What's actually happening** when you prompt an AI
- **Why** certain approaches work better than others
- **How** training stages create specific capabilities

This knowledge will make you a more effective prompt engineer

Why This Matters

The Challenge

Demystifying AI Terminology

You've heard terms like "reasoning," "intelligence," and "understanding" applied to AI systems.

The reality: These terms are a mix of:

- Marketing language to make AI sound impressive
- Engineers needing names for technical processes
- The persistent myth that AI mimics human cognition

The Challenge (cont.)

What Researchers Actually Need

Focus on what matters:

- Understanding what these systems actually do
- Setting appropriate expectations
- Learning how to interact with them effectively

Key insight: AI capabilities are specific, measurable processes—not mysterious intelligence

The Four Stages of LLM Training

The Four Stages of LLM Training

Our Journey Today

1. **Self-Supervised Pre-Training** - Building the Foundation
2. **Supervised Fine-Tuning** - Learning to Converse
3. **Preference Fine-Tuning** - Learning What Humans Want
4. **Reasoning Fine-Tuning** - Teaching Step-by-Step Analysis

Each stage builds specific capabilities that inform how we should interact with AI

The Four Stages of LLM Training

Our Journey Today

1. **Self-Supervised Pre-Training** - Building the Foundation ← *Current*
2. **Supervised Fine-Tuning** - Learning to Converse
3. **Preference Fine-Tuning** - Learning What Humans Want
4. **Reasoning Fine-Tuning** - Teaching Step-by-Step Analysis

Stage 1: Self-Supervised Pre-Training

Building the Foundation

The task: Learn to predict what word comes next

- Model reads massive amounts of text (books, websites, articles)
- Parts are intentionally hidden: `"The cat sat on the ___"`
- System learns `"mat"` appears more often than `"elephant"`

This is self-supervised learning - no human labels needed, just pattern recognition from existing text

What you get: A system that understands language structure but can't hold conversations

The Four Stages of LLM Training

Our Journey Today

1. ~~**Self-Supervised Pre-Training**~~ - Building the Foundation ✓
2. **Supervised Fine-Tuning** - Learning to Converse ← *Current*
3. **Preferred Fine-Tuning** - Learning What Humans Want
4. **Reasoning Fine-Tuning** - Teaching Step-by-Step Analysis

Stage 2: Supervised Fine-Tuning

Learning to Converse

The problem: Text completion ≠ helpful assistant

- Pre-trained model might continue: `"This question often comes up..."`
- Instead of actually answering your question

Example: Ask “What are key success factors?” → gets academic discussion instead of actionable answer

Stage 2: Supervised Fine-Tuning (cont.)

The Solution

Show thousands of human-AI conversations

- Learn dialogue structure
- Recognize questions and provide direct answers
- Package existing knowledge into useful responses

What you get: A conversational system, but it may still produce confident-sounding wrong answers

The Four Stages of LLM Training

Our Journey Today

1. ~~**Self-Supervised Pre-Training**~~ - Building the Foundation ✓
2. ~~**Supervised Fine-Tuning**~~ - Learning to Converse ✓
3. **Preferred Fine-Tuning** - Learning What Humans Want ← *Current*
4. **Reasoning Fine-Tuning** - Teaching Step-by-Step Analysis

Stage 3: Preference Fine-Tuning

Learning What Humans Actually Want

The challenge: How do you train on subjective tasks?

- No single “right” way to write a compelling proposal
- Can’t easily create training examples for complex judgment calls

Example: What makes one research summary “better” than another? It’s often a matter of judgment, not facts

Stage 3: Preference Fine-Tuning (cont.)

The Approach

Human feedback on response quality

- Show model multiple responses to same question
- Learn from human preferences about what's better
- Also handles safety: refusing harmful requests

What you get: Responses aligned with human preferences, but may struggle with complex reasoning

The Four Stages of LLM Training

Our Journey Today

1. ~~**Self-Supervised Pre-Training**~~ - Building the Foundation ✓
2. ~~**Supervised Fine-Tuning**~~ - Learning to Converse ✓
3. ~~**Preferred Fine-Tuning**~~ - Learning What Humans Want ✓
4. **Reasoning Fine-Tuning** - Teaching Step-by-Step Analysis ← *Current*

Stage 4: Reasoning Fine-Tuning

Teaching Step-by-Step Analysis

The gap: Models jump to conclusions without showing work

- Ask for budget evaluation → immediate conclusion
- No systematic breakdown of the problem

The challenge: How do we get systematic analysis instead of quick answers?

Stage 4: Reasoning Fine-Tuning (cont.)

The Breakthrough

The solution: Teach models to “think” step-by-step

- Break complex tasks into steps
- Consider multiple approaches
- Show the analytical process

What you get: Systems that can tackle multi-step problems and show their work

Recent Innovations

Reinforcement Learning

The DeepSeek Innovation

Traditional approach: Expensive human evaluation of every response

DeepSeek's insight: Focus on verifiable problems

- Math and coding have clear right/wrong answers
- Success can be measured automatically
- Dramatically reduced training costs

Impact: Made sophisticated reasoning capabilities more accessible

The Next Wave: GEPA

Reflective Prompt Evolution (July 2025)

New insight: Since these are language models, why not teach them using language?

- Natural language feedback instead of numerical rewards
- Models learn to critique their own work
- Self-reflection and improvement through conversation

Results: 10-20% better performance with 35× fewer computational resources

Timeline: How We Got Here

The Early Foundation (2017-2019)

Stage 1: Self-Supervised Pre-Training

2017: Transformer architecture (Google)

- **“Attention is All You Need” paper** revolutionizes NLP

2018: First breakthrough models

- GPT-1 (OpenAI) - generative pre-training
- BERT (Google) - bidirectional encoder

The foundation was set - but these models couldn’t hold conversations

The Conversation Breakthrough (2020-2021)

Stage 2: Supervised Fine-Tuning

2020: GPT-3 (OpenAI)

- First model that could follow instructions
- Showed emergent abilities at scale

2019-2022: Google's contributions

- T5 (2019) - text-to-text transfer transformer
- PaLM (2022) - 540B parameter breakthrough

Models could now converse - but responses weren't always helpful or safe

The Alignment Revolution (2022)

Stage 3: Preference Fine-Tuning

November 2022: ChatGPT launch (OpenAI)

- RLHF (Reinforcement Learning from Human Feedback)
- Suddenly, everyone had access to AI assistance

Rapid competitive response:

- Claude (Anthropic) - Constitutional AI approach
- Mistral AI emerges with open-source alternatives

The AI race intensifies - every company rushes to catch up

The Reasoning Era (2024-2025)

Stage 4: Reasoning Fine-Tuning

2024: OpenAI's o1 series

- Chain-of-thought reasoning capabilities
- Models that "think" step-by-step

2024-2025: Efficiency breakthroughs

- DeepSeek-R1 - reasoning at lower cost
- GEPA (July 2025) - 35× more efficient training

Key insight: Each breakthrough is adopted and improved by competitors within months, not years

Implications for Research

Understanding Capabilities

What This Means for Literature Review

Stage 1 (Pre-training):

- Vast knowledge of scientific literature
- Understanding of domain terminology and concepts

Stage 2 (Instruction following):

- Can respond to specific research queries
- Formats information appropriately

Understanding Capabilities (cont.)

What This Means for Literature Review

Stage 3 (Preference alignment):

- Produces responses researchers find useful
- Refuses to fabricate citations or make unsupported claims

Stage 4 (Reasoning):

- Can analyze complex research questions systematically
- Shows analytical steps for verification

Critical insight: Each stage enables different capabilities—understanding this helps you design better prompts

Practical Applications

How This Informs Your AI Use

Better prompting:

- Ask for step-by-step analysis (leverages Stage 4)
- Specify output format (leverages Stage 2)
- Request evidence and reasoning (leverages all stages)

Example: “Analyze this research question step-by-step and show your reasoning” vs. “What do you think about this?”

Practical Applications (cont.)

Better Evaluation

Check the AI's work systematically:

- Check if reasoning steps make sense
- Verify factual claims independently
- Look for systematic vs. superficial analysis

Remember: Understanding the training helps you work *with* the AI's strengths rather than against its limitations

Questions & Discussion

Key Takeaways

Moving Forward

- AI capabilities result from specific training stages
- Each stage builds particular abilities
- Understanding this helps you use AI more effectively
- The field continues evolving rapidly

Next steps: How does this change how you think about using AI in your research?

Further Reading

- What Decision-Makers Need to Know About AI Training
- Constitutional AI