Extended Reading in Applied Deep Learning

Applied Deep Learning — Chapter 26

Yin Yang

Foundation Books Project

Roadmap

What This Chapter Is

Six Reading Paths

When to Promote a Topic to a Chapter

Closing

What This Chapter Is

The Right Ending

After the last technical chapter, the right ending is not another syllabus. It is a short map for the places where the book has had to stop.

- Not an assignment plan.
- Not a curriculum.
- Not a claim that the book is incomplete.
- A reader-facing note: an extended-reading map.

What the First Edition Already Covers

- Neural-network foundations.
- Computer vision.
- Language models and retrieval.
- Adaptation (LoRA, fine-tuning) and reinforcement learning.
- Multimodal models, image generation, speech.
- Advanced architecture and serving.

The readings ahead should be used selectively. A reference belongs here when it helps a reader build, measure, debug, govern, or explain.

The Practice Loop

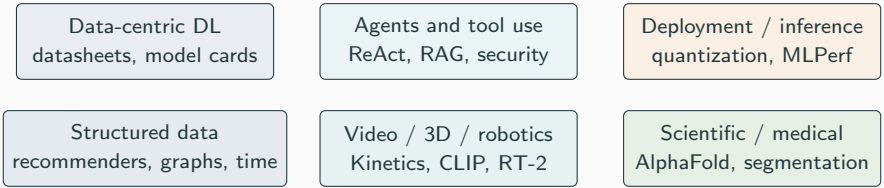
Treat extended reading as a habit, not a museum shelf.

1. Pick one path.
2. Write the applied question it could answer.
3. Choose one reference.
4. Ask: what data, metric, failure case, and runnable baseline would make the idea teachable?

Claims become useful when they can be measured, debugged, and explained to someone who has to build with them.

Six Reading Paths

The Map



Each path is a doorway, not a curriculum outline.

Path 1: Data-Centric Deep Learning

When the missing skill is documentation, provenance, or reporting.

- **Datasheets for datasets** (Gebru et al.).
- **Model cards** (Mitchell et al.).
- **Data cascades in high-stakes AI** (Sambasivan et al.): data quality problems are not isolated bugs.
- **NIST AI Risk Management Framework** and the generative-AI profile: a vocabulary for risk.

Build the discipline first. A model card is an applied artifact.

Path 2: Agents and Tool Use

When the system must reason *and* act.

- **ReAct**: connect reasoning traces with actions.
- **Toolformer**: models that learn tool use from self-supervised signals.
- Revisit **RAG**, **dense passage retrieval**, and approximate nearest-neighbor search before treating an agent as a single model call.
- **Indirect prompt injection** and **OWASP LLM Top 10**: engineering constraints, not afterthoughts.

Read failure modes alongside capability papers.

Path 3: Deployment and Inference Operations

Making models fit real hardware budgets.

- **Quantization** (Jacob et al.) and **mixed precision** (Micikevicius et al.): the entry points.
- **FlashAttention**, **grouped-query attention**, **speculative decoding**: same Transformer, very different latency and memory.
- **MLPerf Inference**: keeps accuracy, latency, throughput, and hardware context in one report.

Reports that hide hardware are not reproducible.

Path 4: Structured Data, Recommenders, Graphs, Time

Where memorization, embeddings, and feature interactions meet production data.

Recommenders Wide & Deep, DeepFM.

Graphs GraphSAGE, PinSAGE — neighborhoods become learnable signals at different scales.

Forecasting N-BEATS, Temporal Fusion Transformers — horizon, covariates, interpretability, temporal validation.

Tabular and temporal applied DL often beats flashier defaults.

Path 5: Video, 3D, Robotics, Embodied AI

When inputs become spatiotemporal or actions must reach the world.

- **Kinetics + I3D**: video action recognition as a supervised perception problem.
- **CLIP**: language supervision became central to open-vocabulary visual systems.
- **Gaussian splatting**: real-time 3D scene representation.
- **RT-2**: vision-language-action models connecting web-scale pretraining with robotic control.

Path 6: Scientific, Medical, High-Stakes Applications

When DL changes a workflow, not just a benchmark.

- **AlphaFold**: landmark for changing a scientific workflow.
- **AI in medicine** (Rajpurkar et al.) and the foundation-model framing (Moor et al.).
- **U-Net, nnU-Net, MedSAM**: segmentation quality matters.
- **Calibration** (Guo et al.): high-stakes models need uncertainty and human review — not raw novelty.

Validation protocol $\geq$ model novelty in clinical contexts.

When to Promote a Topic to a Chapter

The Filter for Future Editions

A topic deserves sustained treatment only when it has:

1. A clear applied task a non-specialist can understand.
2. A small runnable example with data and metrics.
3. Known failure cases.
4. A durable lesson — not merely a temporary interface or vendor convention.

Without all four, the topic stays in the reading map.

Same Filter After Closing the Book

When a new model, benchmark, or tool sounds important, ask:

- What problem does it change?
- What evidence would make that change visible?
- What would break first in real use?

Keep the habit that has run through every chapter:

- Make claims concrete.
- Compare them with baselines.
- Look closely at failures.
- Explain the tradeoff clearly enough that someone else could build from it.

Closing

What Should Outlast Any One Architecture

- **Names** change: Mamba, xLSTM, MoE, the next acronym.
- **Habits** stay:
  - baseline before novelty,
  - measurement before marketing,
  - failure analysis before scaling,
  - clear tradeoffs before strong claims.

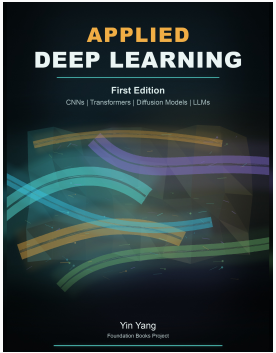
That is the part of applied deep learning that should outlast any single architecture name.

Make claims concrete.

Compare with baselines.

Look closely at failures.

Explain the tradeoff.



Applied Deep Learning

Chapter overview slides for the book by Yin Yang.

Book page	`foundation-books.github.io/applied-deep-learning`
Code	`github.com/foundation-books/applied-deep-learning-code`
Print	`amazon.com/dp/B0GZF7QRSH`
Kindle	`amazon.com/dp/B0GZF9221X`
License	CC BY-NC-ND 4.0

These free overview slides may be shared non-commercially with attribution and without modifications. Full explanations, exercises, and companion-code context are in the book and linked repository.