DS 6050: Deep Learning

School of Data Science, University of Virginia

Fall 2025

Instructor: Heman Shakeri, PhD	**Email:** hs9hd@virginia.edu
TAs: Justin Lee, and Tom Lever	jgh2xh & tsl2b@virginia.edu

- **Asynchronous Lectures & Readings:** Available weekly on Canvas.
- **Zoom Sessions:** Wednesdays, 8:30 PM – 9:30 PM
- **Course Website:** Link

Other Resources:
- Assignments submission (UVA Canvas)
- Discussion Forums & Announcements (Ed Discussion)

Why Should You Care About Learning Deep Learning?

How can you design an AI that truly understands complex medical images, not just mimics patterns? How do you build a language model that can generate coherent text, or an algorithm that can innovate in scientific discovery? These are the kinds of "grand challenges" that today's data scientists and ML engineers tackle, requiring more than just off-the-shelf code. They demand a deep, quantitative understanding of how these models work, from first principles.

You've likely seen countless AI tutorials online, much like the millions of fitness videos on YouTube. Yet, professional athletes and Olympic medalists dedicate themselves to rigorous training camps with expert coaches. Why? Because true mastery—the kind that forges resilient skill and deep intuition—isn't about mimicking surface-level actions. It's built on structured guidance, disciplined hands-on practice with expert feedback, and a profound grasp of fundamentals.

This course is your intensive "training camp" for deep learning. It's where you'll move beyond superficial understanding to forge genuine expertise, from the core out, embracing the deliberate practice needed to push your boundaries and achieve lasting skill.

How Will This Course Help You Succeed (As a Data Scientist/ML Engineer)?

Grand challenges in AI require data scientists who can think critically, design robust systems, and make informed trade-offs. This course will equip you with a conceptual and practical framework to tackle such complex problems in your future research, engineering practice, or advanced studies.

By the end of this training camp, you will be better able to answer critical questions like:

1. How do I use math and core principles to truly understand why, how, and where deep learning models succeed or fail? (No more magical black boxes!)
2. When and how can I build fundamental components (like backpropagation or an attention mechanism) "from scratch" to solidify my understanding, before leveraging powerful frameworks like PyTorch?
3. I've taken calculus, linear algebra, and ML I – how do I integrate all that knowledge with programming to solve real-world AI problems and design innovative systems?
4. Beyond the buzzwords and LinkedIn cheatsheets, how do I develop the critical judgment to make sound engineering decisions, evaluate new research, and maintain my intellectual integrity?
5. How do I use the foundational skills and "learning to learn" strategies from this course to confidently tackle future AI advancements and projects long after graduation?

Course Structure & How You'll Master Deep Learning

This course follows a carefully designed 12-module progression that builds deep learning expertise through three integrated phases:

Phase 1: Foundations & From-Scratch Understanding (Modules 1-3) Master the mathematical underpinnings and implement core algorithms from scratch. You'll build linear models, neural networks, and backpropagation using only NumPy, developing deep intuition about how and why deep learning works. Critical emphasis on optimization foundations and ablation methodology that will recur throughout the course.

Phase 2: Architectural Innovations & Domain Specialization (Modules 4-9) Explore how different data modalities drive architectural choices. From CNNs for spatial data to RNNs for sequences to Transformers for universal modeling, you'll understand why specific architectures excel at specific tasks. Each module revisits optimization challenges unique to that architecture, building your ability to diagnose and fix real-world training problems.

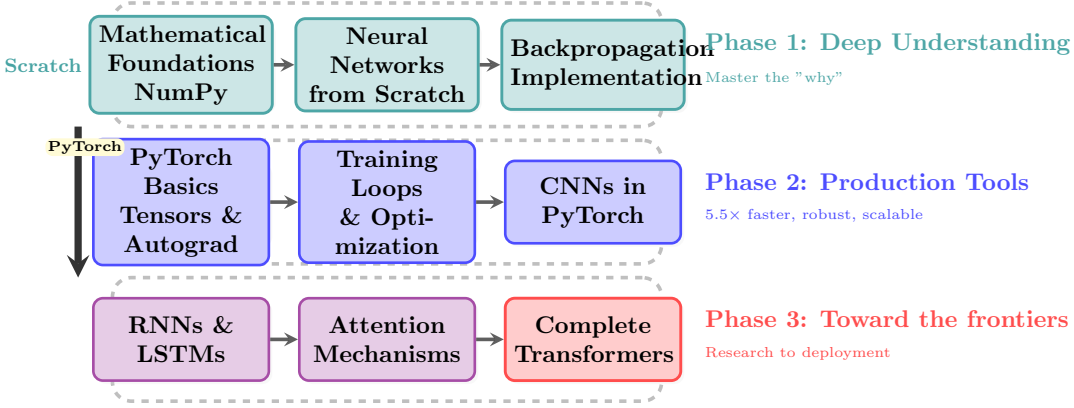

Figure 1: Learning Progression: From Mathematical Foundations to Modern AI Systems

Phase 3: Modern Practice & Research Skills (Modules 10-12) Master contemporary techniques including vision transformers, large-scale pretraining, and generative models. Learn systematic ablation studies, paper reproduction, and research methodology. The course culminates with your team tackling a "grand challenge" using all the skills you've developed.

Course Calendar

Table 1: Course Schedule

Week	Dates	Live Session Meetings
Week 1	Kick off/Logistics.	Wed 8/27, 8:30–9:30PM
Week 2	Module 1	Wed 9/3, 8:30–9:30PM
Week 3	Module 2	Wed 9/10, 8:30–9:30PM
Week 4	Module 3	Wed 9/17, 8:30–9:30PM
Week 5	Module 4	Wed 9/24, 8:30–9:30PM
Week 6	Module 5	Wed 10/1, 8:30–9:30PM
Week 7	Module 6	Wed 10/8, 8:30–9:30PM
Week 8	Module 7	Wed 10/15, 8:30–9:30PM
Week 9	Module 8	Wed 10/22, 8:30–9:30PM
Week 10	Module 9	Wed 10/29, 8:30–9:30PM
Week 11	Module 10	Wed 11/5, 8:30–9:30PM
Week 12	Module 11	Wed 11/12, 8:30–9:30PM
Week 13	Module 12	Wed 11/19, 8:30–9:30PM
Week 14	No class	Wed 11/26 is a UVA holiday
Week 15	Project presentation and Curse wrap up.	Wed 12/3, 8:30–9:30PM

Module Design Tables

Table 2: Module Design — Deep Learning (Domain-Agnostic)

	Module	Learning Goals	Learning Objectives	Instructional Materials
Chapter 1 Foundations & MLPs	**Module 1: Linear → Nonlinear**	• Linear ops & geometry • Why nonlinearity • Build an MLP • SGD overview	• Compute matrix–vector products and visualize linear transforms. • Explain XOR nonlinearity and demonstrate failure → fix with ReLU. • Implement a minimal MLP (NumPy/PyTorch) and describe SGD updates.	• **D2L:** 2.3; 3.1.4; 4.1; 5.1; 12.4 • PyTorch basics; NumPy refresher • Short geometric notes/figures
	Module 2: Backprop & Autograd	• Chain rule & comp. graphs • Backward pass • Autograd usage	• Derive gradients for a 1–2 layer network and draw the computation graph. • Implement gradient checking and validate layer derivatives. • Use `torch.autograd` to build a tiny autograd toy and inspect backward.	• **D2L:** 2.4; 2.5; 5.3 • PyTorch autograd guide
	Module 3: Training & Experiments	• Ablations & HPO • Reproducibility & logging	• Run an optimizer sweep (SGD/Momentum/Adam) with LR schedules; compare results. • Design a 2–3 factor ablation; log all runs with consistent seeds and configs.	• **D2L:** 3.6–3.7; 12; 19 • W&B (or equivalent) quickstart
Chapter 2 Convolutions & Depth	**Module 4: CNN Basics**	• Why convolution & sharing • Spatial invariance • Regularization	• Show MLP spatial limits (shifted inputs) vs a simple CNN. • Implement convolutional layers and train a small CNN. • Apply augmentation and dropout; run a brief augmentation ablation.	• **D2L:** 5.6; Ch. 7; 8.1; 14.1 • *AlexNet* (Krizhevsky et al. 2012) • Conv visualizers/notes
	Module 5: Modern CNNs & Transfer	• Deeper blocks & normalization • Residual connections • Transfer learning	• Inspect gradient flow; compare networks with/without BatchNorm. • Implement a ResNet block; reproduce a mini-ResNet training. • Fine-tune a pretrained model; summarize efficiency trade-offs (e.g., MobileNet).	• **D2L:** 5.4.1; 8.2–8.6; 14.2 • *ResNet* (He et al. 2016); BN (Ioffe–Szegedy) • Model zoo pointers

Continued on next page

Table 2 – continued

Chapter	Module	Learning Goals	Learning Objectives	Instructional Materials
Chapter 3 Sequences & RNNs	**Module 6: RNNs & Optimization**	• Sequential modeling • BPTT & truncation • Stabilizing training	• Build a simple RNN/char-RNN; explain temporal dependencies. • Demonstrate gradient explosion and apply clipping effectively. • Evaluate truncation lengths and initialization choices for stability.	• **D2L:** Ch. 9 • Pascanu et al. (RNN training difficulty)
	Module 7: LSTM/GRU	• Gates & memory • LSTM vs GRU • Bidirectionality	• Implement an LSTM cell; discuss gradient "highways" via gating. • Compare LSTM vs GRU on a toy sequence task; perform gate ablations. • Build a small BiRNN and summarize when bidirectionality helps.	• **D2L:** 10.1–10.4 • Hochreiter–Schmidhuber (LSTM); Cho et al. (GRU)
Chapter 4 Attention & Transformers	**Module 8: Attention & Seq2Seq**	• Dot-product attention • Encoder–decoder • Alignment/weights	• Implement QKV attention; visualize attention weights. • Build a toy seq2seq (copy/reverse) with attention; interpret alignments. • Run a small ablation on attention variants and report findings.	• **D2L:** 10.5–10.7; 11.1–11.4 • Bahdanau et al. (additive attention)
	Module 9: Transformer Blocks	• Multi-head self-attn • Positional encodings • Norm/warmup/dropout	• Implement a transformer block and train a tiny Transformer. • Compare positional encodings; pre- vs post-norm; analyze patterns. • Use warmup and dropout; document stability/perf. impacts.	• **D2L:** 11.5–11.7 • Vaswani et al. *Attention Is All You Need*
	Module 10: ViT Essentials (Domain-Agnostic)	• Patches as tokens • Patch embeddings • CNN–ViT trade-offs	• Implement patchify + linear embedding; train a tiny ViT. • Run a patch-size ablation; inspect attention maps qualitatively. • Compare a small ViT vs small ResNet for efficiency vs accuracy.	• **D2L:** 11.8 • Dosovitskiy et al. (ViT); Touvron et al. (DeiT)

Continued on next page

Table 2 – continued

Chapter	Module	Learning Goals	Learning Objectives	Instructional Materials
	Module 11: Pretraining, PEFT & Scale	• Self-supervised pretraining • Masked vs autoregressive • Parameter-efficient FT	• Pretrain on synthetic sequences (masked and AR); evaluate representations. • Fine-tune on a small downstream task; compare PEFT (LoRA) vs full FT. • Summarize scaling-law takeaways for model/data/compute.	• **D2L:** 11.9; 15.8–15.10; 16.6–16.7 • BERT; GPT; Hu et al. (LoRA); scaling-law summaries
Chapter 5 Generative Models	**Module 12: VAEs, GANs & Diffusion**	• Latent-variable models • Adversarial training • Diffusion basics & ethics	• Implement a tiny VAE; explore the latent space qualitatively. • Train a small GAN; apply simple stability tweaks; assess samples. • Explain forward/reverse diffusion and demonstrate sampling. • Briefly discuss evaluation metrics and ethical considerations.	• **D2L:** 4.7.5; 20.1–20.2 • Kingma–Welling (VAE); Goodfellow et al. (GAN); Ho et al. (DDPM)

Assessment as Learning: How You'll Build and Demonstrate Mastery

Assessment in this course is designed primarily as a learning tool, with evaluation as a by-product. Each assessment provides opportunities to deepen understanding, receive feedback, and refine your skills.

Programming Assignments: Deliberate Practice with Feedback

Five comprehensive assignments that build your implementation skills progressively:

- **Learning Focus:** Each assignment includes extensive unit tests that guide your implementation and provide immediate feedback
- **Iterative Development:** You can discuss together and explore ideas in the forum
- **Reflection Component:** Brief write-ups explaining your design choices and debugging process
- **Peer Code Review:** Optional exchanges with classmates to learn from different approaches

Assignment Progression:

1. Linear Models from Scratch → Understanding gradients and optimization
2. MLP & Backpropagation → Mastering automatic differentiation
3. CNN Implementation → Spatial reasoning and feature hierarchies
4. RNN/LSTM for Sequences → Temporal dependencies and memory
5. Attention & Transformers → Modern architecture principles

Module Quizzes: Knowledge Reinforcement

- **Purpose:** Reinforce key concepts from asynchronous content and readings
- **Format:** Short (10-15 questions), open-book, focused on understanding not memorization
- **Completion-Based:** Full credit for thoughtful completion, encouraging exploration over perfection
- **Immediate Feedback:** Explanations provided for all answers to support learning

Group Project: Solving a Grand Challenge

Multi-phase project applying course concepts to real-world problems:

Formative Milestones:

1. **Proposal & Literature Review:** Receive feedback on problem framing and approach
2. **Mid-Project Check-in:** Share challenges and get guidance from peers and instructors
3. **Final Deliverables:** Complete solution with documentation and reflection

Peer Learning Through Discussion Boards:

- Teams post project summaries, key findings, and demos on Canvas
- Structured peer feedback using provided rubric
- Asynchronous Q&A threads for each project
- "Best Insight" awards chosen by class vote

Participation: Building a Learning Community

Active engagement that enhances everyone's learning:

- **Discussion Leadership:** Each student leads one module discussion thread
- **Peer Support:** Help classmates with conceptual questions and debugging
- **Reflection Posts:** Weekly "aha moments" or challenging concepts
- **Module Quiz Completion:** Demonstrating engagement with all content

Grading: How Your Learning Translates to Grades

While assessment focuses on learning, grades provide a summary of your demonstrated mastery. The grading schema and Late Policy and Extensions is the standard MSDS practice. Here is the breakdown for the total grade:

Component	Weight	Details
Programming Assignments	40%	5 assignments × 8% each
Group Project	40%	Proposal (5%), Mid-checkpoint (10%), Final deliverables (25%)
Participation	20%	See detailed breakdown below

Participation Breakdown (20%)

- **Module Quiz Completion (10%):** 12 quizzes, full credit for thoughtful completion
- **Discussion Contributions (10%):** Quality posts, peer responses, Project Peer Feedback, and discussion leadership

Professional and Academic Integrity

As future leaders in data science and AI, uphold the highest standards of ethics and integrity. All work must comply with the UVA Honor System. Individual assignments must be your own original work, though conceptual discussions are encouraged. Any external code, ideas, or resources must be appropriately cited. The pledge should be included with relevant submissions.