Lecture 0: Overview

Block 4.2 Advanced AI

Eoin O’Brien

The major

Advanced AI

- Eight weeks, 7 September to 30 October
  - Effectively 7 teaching weeks, 14 examinable studio days
  - Note that we'll lose Monday Week 8 to a bank holiday!
- Two studio days a week (Mondays and Fridays)
- Two 2-hour lecture blocks with a studio hour in between

The plan

Weeks 1 to 7

- **Week 1**: supervised learning
- **Week 2**: probability and loss functions
- **Week 3**: fitting models, gradients and initialisation
- **Week 4**: measuring performance and regularisation
- **Week 5**: network architectures
- **Week 6**: transformers, attention and reinforcement learning
- **Week 7**: generative models, autoencoders and diffusion
- No plan survives contact with the enemy!
  - ...so we'll move things around if we need to.

Week 8

- Final exam and project submission
- Project viva voce (oral defence)

Assumptions

New year, new paradigm

- This is an MSc course now!
  - The expectations are higher, and the content is more advanced
  - Work that merited an A1 in the undergraduate course won’t necessarily merit an A1 here
- Independent reading and thinking is expected
  - The lectures are a guide, not a substitute for the text or for doing the work
  - If you don’t do the reading...
- Participation is expected
  - The lectures are a two-way street, and your questions and contributions are welcome
  - If you don’t participate...

What I assume about you

- You should be **expert programmers** at this point!
  - We're not going to spend time on Python syntax
  - I'm assuming that I can throw a library at you, and you can read the docs and figure out how to use it
- You should be familiar (-ish) with **linear algebra**, **differential calculus** and **probability**
  - Notation and concepts get a refresher as we go
  - ...but you'll need to do some reading if there are gaps!
- In general, Leaving Cert HL maths is a good baseline
  - So long as you can read a formula and follow a derivation, you should be fine

How to get the most out of this major

- **Reading isn’t optional**: the slides are dense enough to revise from, deliberately
  - They’re not a substitute for the text, or for doing the work
  - Read the assigned Prince chapter *before* the lecture. Flipped classroom only works that way
  - Assigned reading is examinable
- **Come to the lectures**: mandatory, but not graded
  - …because merely showing up doesn’t merit anything!
- **Don’t be afraid to be wrong**: challenge yourself, take a stand on an open question, and be prepared to defend it
  - …because one of the best ways to learn is to make mistakes and correct them!
- **Ask questions**: if you are lost, you are not the only one
  - If you’ve got an interesting thought, someone else will want to hear it

Assessment

Structure

Component	Weight	Assessed
Continuous Assessment	20%	weekly, weeks 2-7
Final Exam	40%	09:00 Friday, week 8
Project	40%	submission and viva, week 8

What is assessed

- **Continuous Assessment**: in-class quizzes and lab submissions, from week 2
  - How it divides across them: I'll go through it in the room, and post the split once it's fixed
- **Final Exam**: drawing on material from across the major
  - Closed-book, probably a mix of MCQs and written-response questions
  - In the studio, unless I tell you otherwise

- **Project**: one system you build, and then defend in person
  - Format and requirements will be provided in the upcoming weeks
  - Assessed on both the submission and the viva voce (oral defence)
  - Superb project that you can’t defend? Expect a zero.

Academic integrity

- **Plagiarism is a serious offence**: all work submitted must be your own
- **Collaboration is encouraged**: but you must acknowledge any help received
- **Use of AI tools**: is permitted (and encouraged where stated), but you must disclose any AI assistance in your submissions
  - If you can't defend it, don't submit it.
- **Exam integrity is paramount**: any attempt to cheat will result in *at minimum* an automatic fail for that component
  - And may result in further disciplinary action

- Alt-tab during an exam? Expect a zero.
    - The same goes for notes in a sleeve, and for text parked on your clipboard.
- Better to fail honestly than permanently mar your academic record.
- If you are struggling, reach out to me or your peers for help.
    - That is not a formality. Come and find me early, while it is still cheap to fix.

AI and learning

The obvious problem

- Everyone has access to LLMs now.
  - They generate code, derivations and answers on demand.
  - Pretending that nobody's going to use them would be laughably naive of me...
- And I'm not a Luddite!
  - AI is a tool, and like any tool it can be used well or poorly
- So let's deal with evidence rather than vibes.
  - We're scientists here, right?

What the evidence says

- A carefully designed AI tutor **beat in-class active learning** in a Harvard physics randomised trial, 194 students (Kestin et al. 2025)
- Give high-school students **unrestricted** access instead and they score 48% better on the practice problems, then 17% *worse* on the exam than students who never had access (Bastani et al. 2025)
- Same technology. Opposite result. Theories?
  - The variable is **who does the thinking**.
  - The tutor *forces* the student to think first, then gives feedback.
  - Unrestricted access lets the model do the thinking.

Use it to think harder

- **Explain a concept to it**, then ask it to find the gaps in your explanation
  - If you can teach it, you understand it
  - Even when the student is a glorified rubber duck
- **Do the problem first**, then ask for a critique of your attempt
  - This is an old trick (Slamecka and Graf 1978)
- **Ask it to quiz you**, then come back to the same questions days later
  - Spaced repetition is evidence-based and effective. (Dunlosky et al. 2013; Roediger and Karpicke 2006)
- **Ask why, why not, and what if**: “What happens if I build a deep neural network without an activation layer?” beats “notes on ReLU pls” every time
  - Interrogate and probe the answers. Read critically!

The traps

- **The practice illusion**: getting through more problems feels productive.
  - ...whether you're benefiting from it or not.
- **Copy-paste escalation**: usage drifts toward pasting output verbatim (Kosmyna et al. 2025)
  - LLM output is usually fluent and plausible, and not reliably correct. If you can't defend it, don't submit it.
- **Skipping the struggle**: the difficulty you feel working something out is not an obstacle to the learning.
  - It *is* the learning.
- **The illusion of competence**: novice programmers using an LLM finish believing they did better than they did (Prather et al. 2024)

The expert and the beginner

- The split isn't expert against beginner. It's **doing the task now** against **learning to do it without the tool**
  - It helps you do the task, often most where you're weakest
  - It does *not* teach you to do it unaided (Bastani et al. 2025; Bassner et al. 2026)
- You're (to an extent) both at once
- **Expert programmer:** Lean on this!
  - Generate the boilerplate, argue about the architecture, let it write the plotting code.
- **Novice ML mathematician:** This is the footgun moment.
  - Do the derivation yourself first, every time.
  - Don't abdicate the thinking to the model. You won't have it in the exam or the project viva.

Content

The text

- Prince, *Understanding Deep Learning* (Prince 2023)
  - The core text, and I'll assign reading from it most weeks.
  - And it's free! Not that any of you would pirate a textbook, right?
- Research papers and other readings arrive later in the major.
  - These will be linked from the relevant lecture slides.

The slides

- A companion to the text, not a replacement for it.
  - On the site by the morning of each lecture.
  - The revision notes sit beside each deck, and carry the reasoning the slides compress.
- Once I get access to Brightspace, I'll post the slides there as well.
  - advanced-ai.eoin.ai remains the authoritative source.
  - It's generated from my course repo and re-renders every time I push a change, so it will always be current.

References

Bassner, Patrick, Ben Lenk-Ostendorf, Ramona Beinstingel, Tobias Wasner, and Stephan Krusche. 2026. “Less Stress, Better Scores, Same Learning: The Dissociation of Performance and Learning in AI-Supported Programming Education.” *Computers and Education: Artificial Intelligence* 10: 100537. https://doi.org/10.1016/j.caeai.2025.100537.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. “Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics.” *Proceedings of the National Academy of Sciences* 122 (26). https://doi.org/10.1073/pnas.2422633122.

Dunlosky, John, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathan, and Daniel T. Willingham. 2013. “Improving Students’ Learning with Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology.” *Psychological Science in the Public Interest* 14 (1): 4–58. https://doi.org/10.1177/1529100612453266.

Kestin, Greg, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. 2025. “AI Tutoring Outperforms in-Class Active Learning: An RCT Introducing a Novel Research-Based Design in an Authentic Educational Setting.” *Scientific Reports* 15 (1). https://doi.org/10.1038/s41598-025-97652-6.

Kosmyna, Nataliya, Eugene Hauptmann, Ye Tong Yuan, et al. 2025. *Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task*. https://arxiv.org/abs/2506.08872.

Prather, James, Brent N Reeves, Juho Leinonen, et al. 2024. “The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers.” *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1*, ICER 2024, August, 469–86. https://doi.org/10.1145/3632620.3671116.

Prince, Simon J. D. 2023. *Understanding Deep Learning*. MIT Press. https://udlbook.github.io/udlbook/.

Roediger, Henry L., and Jeffrey D. Karpicke. 2006. “Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention.” *Psychological Science* 17 (3): 249–55. https://doi.org/10.1111/j.1467-9280.2006.01693.x.

Slamecka, Norman J., and Peter Graf. 1978. “The Generation Effect: Delineation of a Phenomenon.” *Journal of Experimental Psychology: Human Learning and Memory* 4 (6): 592–604. https://doi.org/10.1037/0278-7393.4.6.592.

Questions?

Eoin O’Brien · eoin@eoin.ai