Course Outline – STAT 946: Generative AI and Large Language Models (LLMs)

Instructor: Ali Ghodsi
Prerequisite: Machine learning, probability, calculus, linear algebra, and statistics

---

⚠️ **Important Notice on Course Overlap**

There is significant overlap between STAT 946 and STAT 940 (Deep Learning).
Since STAT 946 focuses on Generative AI and LLMs, we revisit foundational models such as Transformers, Stable Diffusion, VAE, BERT, GPT, and GAN – many of which are already covered in STAT 940.

Taking STAT 946 after completing STAT 940 or taking the two courses at the same time is strongly discouraged due to this content overlap. While these two courses are not officially antirequisites, **in practice, they are.** Students are encouraged to carefully review the syllabus and consult with the instructor if they are uncertain about enrolling in this course.

Course Description

This course offers an in-depth exploration of generative artificial intelligence and large language models (LLMs). Students **would** develop foundational knowledge of core architectures, including transformers, diffusion models, and variational autoencoders. The course covers practical aspects of large-scale model training, optimization, and techniques for improving model efficiency (e.g., knowledge distillation, efficient transformers).

Additionally, the course addresses theoretical challenges in deep learning that diverge from traditional statistical machine learning frameworks, such as phenomena like double descent and the role of inductive biases. Some topics will be delivered through formal lectures, while others will be explored through paper presentations and in-class discussions, fostering a deeper understanding of the latest advancements and open problems in the field.

Tentative Topics

Core Topics:

- **Recurrent Neural Networks (RNNs):** Architecture and applications to sequential data
- **Attention Mechanisms and Self-Attention:** Concepts and applications in transformers
- **Sequence-to-Sequence (S2S) Models:** Encoder-decoder architectures for tasks like translation
- **Transformers:** Introduction and training of transformers for NLP and other tasks
- **BERT and GPT:** Understanding bidirectional and autoregressive LLMs
- **Reinforcement Learning from Human Feedback (RLHF):** ChatGPT and alignment techniques in LLMs

- **Variational Autoencoders (VAEs) and Performers:** Dimensionality reduction and scalable attention
- **Generative Adversarial Networks (GANs) and Adversarial Autoencoders (AAEs):** Adversarial training techniques
- **Diffusion Models (DDPM):** Denoising diffusion probabilistic models for generative tasks

Advanced/Optional Topics (Subject to Available Time):

- **Efficiency in Models:** Techniques like knowledge distillation, efficient transformers, and sparsity methods to reduce computational overhead without sacrificing performance.
- **Double Descent Phenomenon:** Exploring the non-monotonic relationship between model capacity and generalization error in deep learning.
- **Inductive Biases in Deep Learning:** Understanding how implicit biases in model architecture and training affect learning dynamics and generalization.
- **Limitations of Classical Statistical Learning Theory:** Discussion on why traditional machine learning theory falls short in explaining deep learning behavior, and emerging frameworks to address this gap.
- **Multimodal Generative Models:** CLIP, DALL·E, Whisper, and audio-text models
- **Large-Scale Training and Distributed Systems:** Parallelism, memory management, and precision strategies
- **Parameter-Efficient Fine-Tuning:** LoRA, prompt tuning, and adapters
- **Model Compression and Acceleration:** Distillation, quantization, and pruning
- **Advanced RLHF and Constitutional AI:** Safety and reward modeling
- **Advanced Inference Strategies:** Sampling (Top-k, nucleus), decoding methods
- **Retrieval-Augmented Generation (RAG):** Knowledge-enhanced generation techniques
- **Curriculum and Continual Learning:** Learning efficiency and catastrophic forgetting
- **Safety and Robustness:** Adversarial defenses and alignment checks
- **System Design and Deployment:** Latency optimization and edge inference
- **Ethics and Governance:** Bias, fairness, and regulatory frameworks
- **Emerging Research:** Multimodal models, autoformalization, and future directions

---

TEntetive Assignments and Evaluation

Activity	Weight (%)
Assignment / Data Challenge (2 at 15% each)	30%
Paper Presentation	10%
Wiki Contribution (for Paper)	10%
Final Group Project (Report)	50%

Wiki Contributions for Paper Presentations

The course includes group paper presentations. Students who are not presenting must contribute by scribing papers presented by their peers. Each student is required to contribute to at least half of the presented papers. Contributions must be completed by the final class. Contributions to the paper a student presents are not counted. This ensures continuous engagement and a shared understanding of the presented materials. See details in Activities and Assessments.

Tracking and Evaluation

The wiki platform records all contributions, and this history will be used for grading. Contributions will be assessed for quality, completeness, and consistency. This process is designed to encourage active participation and reinforce collaborative learning in the asynchronous environment.

Student Resources

Student Resources

- Academic advice
- Student success
- WatCards
- Library services and more

Turnitin or other plagiarism detection tools will be used to verify the originality of your submissions. Please note that submissions to Turnitin are stored on a U.S. server. If you have concerns regarding privacy and/or security, please inform the instructor within the first week of the course. Communication should be made via email with the subject line: "Opting Out of Turnitin for STAT 940."

Resources

- **Library COVID-19**: Updates on library services and operations.

Notice of Recording

Some lectures for STAT 946 will be recorded, including audio and video, as part of the course activities. Recorded materials may be posted on YouTube or other platforms, used by the University for educational purposes, and shared with course staff and students.

By participating in activities that involve recording, you consent to the use of your image, voice, and text.

If you do not consent to being recorded, you must email the instructor within the first week of the course with the subject line: **"Opt Out Recording – STAT 946."**

University Policies

- **Academic Integrity**:
  To maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect, and responsibility. [Check the Office of Academic Integrity for more information.]

When utilizing ideas, charts, text, or any other intellectual property created by someone else, proper citation of the original source is mandatory.

If you directly copy text—be it a sentence or a paragraph—from another's work, you must not only cite the source but also enclose the copied material within quotation marks.

Evidence of plagiarism in final project reports, codes, or any other submitted materials will result in a failing grade for the course. All reports and codes will be checked by plagiarism detection software.

- **Grievance**:
  A student who believes that a decision affecting some aspect of their university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Policy 70, *Student Petitions and Grievances*, Section 4. For further assistance, contact the department's administrative assistant.

- **Discipline**:
  Students are expected to know what constitutes academic integrity to avoid committing an academic offence and to take responsibility for their actions.
  For information on categories of offences and penalties, refer to Policy 71, *Student Discipline*. For typical penalties, see *Guidelines for the Assessment of Penalties*.

- **Appeals**:
  A decision made or penalty imposed under Policy 70 (*Student Petitions and Grievances*) or Policy 71 (*Student Discipline*) may be appealed if there are grounds.

- **Note for Students with Disabilities**:
  AccessAbility Services, located in Needles Hall, Room 1401, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum