Deep Learning: Modern Research Topics

Methods, systems, evaluation, and reliability in contemporary deep learning.

A research-facing course on how modern deep-learning claims are made, tested, scaled, and stress-tested, with emphasis on credible experimentation, efficient systems, long-context modeling, reliability, representation learning, and adaptation. The course is organized around reading modern papers while keeping the methodological and systems assumptions explicit.

TOPIC MAP

Research Methodology

- Baselines, ablations, uncertainty, and reproducibility.
- Fair comparison and experimental design.
- Benchmark leakage, evaluation mismatch, and reporting practice.
- Reading papers through claims, evidence, and limitations.

Systems and Efficiency

- Deep-learning frameworks and execution models.
- Parameters, FLOPs, memory bandwidth, latency, and scaling laws.
- Profiling, batching, memory movement, and serving constraints.
- Quantization, pruning, distillation, and speculative decoding.

Long-Context and Efficient Sequence Models

- Efficient Transformers and attention alternatives.
- Recurrent and state-space sequence models.
- Retrieval, memory, positional methods, and context extension.
- KV-cache behavior and cache-efficient inference.

Reliability and Robustness

- Calibration, reliability diagrams, and selective prediction.
- Uncertainty estimates, confidence failures, and abstention.
- Out-of-distribution detection and robustness.
- Evaluation mismatch and deployment assumptions.

Causality and Generalization

- Causal views of prediction and distribution shift.
- Spurious correlations and intervention-based thinking.
- Invariance, selection bias, and counterfactual reasoning.
- Stress-testing claims beyond benchmark accuracy.

Representation and Adaptation

- Contrastive and non-contrastive self-supervised learning.
- Multimodal alignment and vision-language representations.
- Continual learning, forgetting, and replay.
- Test-time adaptation under changing target distributions.