

Proposal for a New Course

Department of Intelligent Systems Indian Institute of Technology, Kanpur

Course number: IS731

Course title: Verification and Testing of AI-Based Systems

Course prerequisites: An introductory course on machine learning (or equivalent); programming in Python. Familiarity with basic logic and discrete mathematics is expected; prior exposure to software testing or formal methods is helpful but not required.

Course credits: 3-0-0-0 [9]

Course duration: Full semester

Course type: Department PG Elective

Proposing instructor: Prof. Indranil Saha

Other faculty members interested in teaching the course: NA

Other departments interested in the proposed course: CSE

Course description:

a. *Objectives:* Machine-learning components are increasingly being deployed in safety-critical and high-stakes systems, creating new challenges for verification and testing. Unlike conventional software, learned models are often not derived from explicit, human-written specifications, and their behavior depends on data, training procedures, and complex model architectures. Consequently, traditional verification and testing approaches must be adapted and extended to provide meaningful assurance of AI-based systems.

This course aims to equip students with the principles and techniques needed to systematically verify and test AI-based systems and assess their readiness for deployment. Students will learn to formulate testable requirements and formal properties for machine-learning components; test datasets and deep neural networks using white-box and black-box techniques while efficiently prioritizing tests under limited labeling budgets; and evaluate properties such as robustness, uncertainty, and fairness. The course will also cover formal verification of neural networks, including quantized networks, as well as verification and testing of systems incorporating neural components. For closed-loop and autonomous systems, students will study reachability analysis, falsification, scenario-based testing, and runtime assurance to establish safety and reliability under diverse operating conditions.

b. *Logistics:* The course will serve as a Department Elective Course for the IS PG students. Undergraduate students will be able to take this course as an Open Elective course. There is no scheduled laboratory slot; hands-on work is done through take-home programming assignments and a team project using open-source tools that run on freely available GPU resources.

c. *Content*: There will be an equivalent of 39 lectures of 50 minutes each over 13 weeks. The content is organised into six major topics, most of which span several weeks: foundations of V&V for AI-based systems; testing data for ML; testing deep neural networks; robustness, uncertainty and fairness; formal verification of neural networks; and verification and testing of systems with neural components. A weekly breakup of content is given below.

d. *Evaluation*: The evaluation will be based on programming assignments, a course project (proposal, test plan, verification results, final report, presentation), a mid-semester examination and an end-semester examination.

Weekly breakup of content:

Week	Major topic	Content
1	Foundations of V&V for AI-based systems	– Classical verification and validation: test oracles, test levels, coverage criteria, formal verification – Why ML breaks classical assumptions: no complete specification, the oracle problem, faults in data, non-determinism – Aggregate accuracy vs guarantees on individual inputs – The ML lifecycle and where defects enter; case studies of AI failures – Requirements for ML components and the operational design domain – Properties: local robustness, monotonicity, fairness, safety envelopes – Overview of standards and regulation for AI systems
2	Testing data for ML	– Data validation: schemas, data unit tests, distribution checks (KS test, PSI) – Training–serving skew and data leakage – Label quality and confident learning; slice-based evaluation – Data drift and concept drift; monitoring ML systems in production
3	Testing deep neural networks	– White-box adequacy criteria: neuron, k-multisection, boundary, top-k and MC/DC-style coverage – Surprise adequacy – Test generation: gradient-guided generation, coverage-guided fuzzing, concolic testing – Does coverage correlate with fault detection?

Week	Major topic	Content
4		– The test-oracle problem and partial oracles; differential testing across models, frameworks and precisions – Metamorphic testing of DNNs; transformation-based test input generation – Input distribution coverage: learning a low-dimensional latent model of the input distribution and applying combinatorial interaction testing over latent features – Co-domain coverage: covering the output space of predicted classes and confidence levels; co-domain coverage-guided fuzzing (CoDoFuzz)
5		– Test prioritization: uncertainty-, surprise-, coverage- and mutation-based; APFD – Test selection for accuracy estimation – Mutation testing for deep learning – Testing ML code and pipelines; model tests in CI/CD
6	Robustness, uncertainty and fairness	– Adversarial examples and threat models; FGSM, PGD, black-box and transfer attacks – Pitfalls in robustness evaluation – Natural robustness and distribution shift – Adversarial training; certified robustness via randomized smoothing
7		– Uncertainty and calibration; out-of-distribution detection – Fairness metrics, bias audits and counterfactual fairness tests – Explainability as a debugging tool and its limits
8	Formal verification of neural networks	– Formal methods primer: logic, SAT/SMT, soundness and completeness – The neural-network verification problem – Complete verification: MILP encoding of ReLU networks, Reluplex, Marabou; the ACAS Xu case study
9		– Abstract interpretation and interval bound propagation – Linear relaxations: CROWN, DeepPoly – Branch and bound; VNN-COMP benchmarks – Certified vs empirical robustness

Week	Major topic	Content
10		– Quantization schemes; why guarantees on the floating-point model do not transfer – Differential testing of floating-point and quantized networks – SAT encodings of binarized networks; bit-vector SMT and ILP encodings of quantized networks – Quantization error bounds and differential verification
11	Verification and testing of systems with neural components	Theme: formal verification of closed-loop systems – Why component-level guarantees do not imply system-level safety – Closed-loop models of neural-network control systems: discrete-time and hybrid models – Reachability analysis: set representations (star sets, zonotopes, Taylor models) and the wrapping effect – Systems with learned perception: perception error models; verification of image-based controllers using generative models – Tools and benchmarks: NNV, POLAR, Verisig; ARCH-COMP
12		Theme: falsification and simulation-based testing – Specifying system requirements in signal temporal logic; quantitative (robust) semantics – Falsification as optimization: stochastic search and Bayesian optimization – Compositional falsification of systems with ML components – Scenario-based testing: scenario languages (Scenic), sampling with VerifAI; coverage of the operational design domain – The simulation-to-reality gap
13		Theme: runtime verification and runtime assurance – Runtime monitoring of signal temporal logic requirements – Runtime monitoring of neural networks: out-of-distribution and activation-pattern monitors – Runtime assurance: the Simplex architecture and verified safety controllers – Control barrier functions as safety filters for neural-network controllers

Short summary for inclusion in the Courses of Study booklet:

The course covers testing and formal verification of deep neural networks and of the AI-based systems built on them.

Textbook: None. The course will be based on the research literature in this area.

References:

The course content is based on recently published research paper

Course proposer:

Date:

Convener DPGC:

Date:

The course is approved/not approved

Chairperson, SPGC

Date: