CEE 4XX Applied Machine Learning for Civil and Environmental Engineering

Wilfredo Torres Calderon | wtorresc.github.io

Proposed course design. “4XX” is a placeholder number.

Course at a glance

Format	28 lectures (two 75-minute sessions per week), 3 credit hours (undergraduate) or 4 (graduate), the same structure as CS 441.
Prerequisites	CS 101; CEE 202 Engineering Risk & Uncertainty; linear algebra (e.g., MATH 257); or consent of instructor. A linear-algebra and Python refresher notebook is released before week 1.
Texts	Forsyth, *Applied Machine Learning* (2019; primary text of CS 441, free through the Library); Brunton & Kutz, *Data-Driven Science and Engineering* (2nd ed., 2022; shared with CEE 492); Prince, *Understanding Deep Learning* (2023, free online); Barocas, Hardt & Narayanan, *Fairness and Machine Learning* (2023, free online).
Computing	Python: NumPy, scikit-learn, PyTorch, Hugging Face, WNTR, SUMO. Browser-based notebooks on Illinois Computes Jupyter (shared A100 GPUs) or Colab; PrairieLearn for weekly checks; CBTF for exams.
Assessment	Five homeworks (100-point core plus stretch goals) earning experience points (XP), three CBTF exams with the best two counting, and a custom-data final project, modeled on CS 441.

Why a civil and environmental engineering course, not a CS course with CEE examples

CS 441 teaches computer scientists to apply machine learning to any domain. This course takes CEE students to the same depth, but through the problems, data, and professional obligations of *their* domain. It keeps everything that makes CS 441 rigorous and adds what a general course leaves out. Twenty of its 28 lectures cover CS 441 material, rebuilt around CEE data and compressed where CEE 202 has already laid the probability groundwork; the other eight cover material that CEE needs and CS 441 does not.

- **P1. No reduction in depth.** Every method is derived, implemented from scratch with shape checks, and verified against a reference library at stated tolerances before students use the library or an AI assistant (*derive, implement, then delegate*).
- **P2. Real CEE data, in the forms CEE actually has.** Each method is introduced through a CEE problem and a public dataset, including data types general courses rarely reach: infrastructure networks, point clouds, geospatial rasters, and design documents.
- **P3. Validation the way infrastructure fails.** Spatial and temporal dependence, nonstationary climate, extrapolation beyond tested ranges, rare events, and leakage are central topics, not footnotes.
- **P4. Uncertainty and decisions, not only accuracy.** Calibration, conformal prediction, surrogate models, and maintenance decisions build directly on CEE 202 and on reliability-based design.
- **P5. Physics as a check on data, and data as a check on physics.** Units, conservation laws, monotonicity, and boundary conditions enter as features, constraints, and tests.
- **P6. AI impact integrated, not appended.** Following the Campus AI Curriculum Task Force, social, ethical, environmental, and economic considerations appear in every unit (the AI-impact threads below), which also addresses ABET Student Outcomes 2 and 4.
- **P7. Assessment that stays rigorous as students use AI.** Every component declares one of Grainger’s four AI-use levels (§6).

Relationship to CEE 492. CEE 492 introduces data science to CEE students. This course is designed to sit alongside it at the depth of CS 441. In consultation with CEE 492's faculty, it could run as a new elective, as a sequel, or as the basis of an expanded AI offering in the AI and data science curriculum.

Learning goals

By the end of the course, students will be able to:

	Goal	ABET SO	Task Force foundation
LG1	Formulate a CEE problem as a learning problem, with inputs, outputs, loss, and a validation scheme that respect the data's spatial, temporal, and physical structure.	1, 6	AI & ML Technologies
LG2	Derive, implement, and verify the core algorithms of machine learning, from nearest neighbors to transformers, and explain their assumptions and failure modes.	1, 7	AI & ML Technologies
LG3	Quantify and communicate predictive uncertainty, and use it in decisions under risk.	1, 2, 6	AI & ML Technologies
LG4	Embed physical knowledge in models and detect when a data-driven model violates it.	1, 6	AI & ML Technologies
LG5	Connect foundation models and language models to engineering data and tools, with verification that catches their errors.	6, 7	Agentic Systems & Workflows
LG6	Evaluate the social, ethical, environmental, and economic impacts of an AI system on the people and places it affects.	2, 4	Human-Centric AI
LG7	Recommend for or against an ML solution to a non-specialist stakeholder, and defend it orally and in writing.	3, 5	Human-Centric AI

Schedule

Columns: lecture; methods (each derived and implemented); CEE anchor problem and data; CS 441 counterpart ("new" means not covered in CS 441).

Unit 1. Learning from engineering data (L1–L8)

L	Methods	CEE anchor and data	CS 441
1	ML when predictions carry loads; problem framing; train, validation, test	Tour of the course datasets; what an engineer would need before signing a model's output	L01
2	Nearest neighbors; representation, units, scaling; dimensionless groups	Concrete strength from mix design (UCI Concrete; Yeh, 1998)	L02
3	Generalization, bias–variance; random vs. spatial and temporal block cross-validation; leakage	Bridge condition across Illinois counties (FHWA National Bridge Inventory); optimism of random splits (Roberts et al., 2017)	L03
4	Retrieval, k-means, PCA via the SVD, embeddings	Load profiles of buildings (Building Data Genome 2); modes of a sensor array	L04–05
5	Least squares via QR and SVD; ridge and lasso; physics-based features	Abrams' water–cement law in the concrete data; in-class exercise in portfolio §5	L06
6	Logistic regression, SVM; ordinal and cost-sensitive classification	Screening for poor bridge condition when a missed bridge costs far more than a false alarm (NBI)	L07
7	Naive Bayes; mixture models and EM; kernel density estimation	Free-flow vs. congested traffic regimes from freeway loop detectors (Portland I-205, USDOT open data)	L08–10
8	Decision trees, random forests, gradient boosting; SHAP and partial dependence	Pavement deterioration (FHWA Long-Term Pavement Performance); do explanations agree with engineering knowledge?	L12–13

AI-impact thread. Whose infrastructure is in the data? A complete inventory (the NBI) differs from crowd-reported data such as Boston's Street Bump pothole app, whose coverage followed smartphone ownership (Crawford, 2013); measurement bias becomes allocation bias.

Unit 2. Uncertainty, risk, and decisions (L9–L12)

L	Methods	CEE anchor and data	CS 441
9	Outliers, robust losses, RANSAC, isolation forests; sensor faults	Leak detection in a water distribution network (BattLeDIM; Vrachimis et al., 2022)	L11
10	Calibration, reliability diagrams, proper scoring rules; quantile regression	Characteristic (5th-percentile) strength vs. mean prediction; the link to load and resistance factor design	new
11	Conformal prediction; covariate shift and nonstationarity	Streamflow intervals for the Sangamon River at Monticello (USGS, 1908–2025), trained on one period and tested on another (Milly et al., 2008)	new
12	Gaussian processes; surrogate models; Bayesian optimization; active learning	Emulating a calibrated hydraulic model (L-Town, BattLeDIM, in WNTR) to find the valve settings that leak least	new

Exam 1 (CBTF): Units 1–2

AI-impact thread. Who bears the residual risk of a model's error, and how should uncertainty be communicated to the public? Under a changing climate, historical records are a moving target.

Unit 3. Deep learning for sensing and inspection (L13–L16)

L	Methods	CEE anchor and data	CS 441
13	Stochastic gradient descent and its variants, from scratch	Fitting a deterioration model at the scale of the national inventory	L14
14	Multilayer perceptrons; backpropagation and a minimal autograd	A neural surrogate of the L12 simulator, compared with the Gaussian process	L15
15	CNNs, residual networks, transfer learning	Concrete-crack classification across decks, walls, and pavements (SDNET2018; Dorafshan et al., 2018)	L16–17
16	Segmentation and detection (U-Net); measuring from pixels	Crack width in millimeters from depth images (3DCrack); flooded buildings and roads in satellite imagery (SpaceNet 8; Hänsch et al., 2022)	L17

AI-impact thread. Automating inspection: what changes in the inspector's role, who is liable for a missed defect, and which skills grow in value for the workforce.

Unit 4. Foundation models and time series (L17–L20)

L	Methods	CEE anchor and data	CS 441
17	Word representations; attention and transformers from scratch; vision transformers	The L15 crack tiles again: a vision transformer against the CNN, and what its attention maps attend to	L18–19
18	Vision foundation models: CLIP, SAM, geospatial models (Prithvi)	Screening the L15 crack tiles with no labels, then a few; SAM on crack depth images and flooded buildings; construction-site monitoring in HW4 (portfolio §4)	L20
19	Language models; embeddings; retrieval-augmented generation; evaluation	Question answering over a public-domain standard (FHWA MUTCD); coding OSHA severe-injury reports by event type (cf. Tixier et al., 2016)	L20
20	1-D signals: spectra, 1-D CNNs, LSTMs; time-series foundation models	Day-ahead flood warning on an Illinois river (CAMELS; Kratzert et al., 2019): an LSTM against a forest and a pretrained forecaster; a railway bridge's vibration and a temperature-confounded alarm (KW51)	L23

Exam 2 (CBTF): Units 3–4

AI-impact thread. The energy, water, and carbon footprint of training and inference (Strubell et al., 2019; Li et al., 2023); fabricated code clauses; copyright in design codes and data.

Unit 5. Structure and physics (L21–L24)

L	Methods	CEE anchor and data	CS 441
21	Graph neural networks; message passing from scratch	Pressure at 782 junctions of a water network from 33 sensors (L-Town, BattLeDIM); traffic forecasting (Li et al., 2018) as the same model on roads	new
22	Point clouds and 3-D learning (PointNet); registration	Indoor laser scans with their as-built BIM model (Abreu et al., 2023); construction progress monitoring by comparing scans with BIM	new
23	Physics-informed neural networks; constraints and conservation; known failure modes	Contaminant transport (1-D advection–dispersion) and Terzaghi consolidation, checked against a finite-difference solver (Raissi et al., 2019; Krishnapriyan et al., 2021)	new
24	Data-driven dynamics (SINDy); neural operators	The laws of a detention pond and of real rivers' recessions (CAMELS); a neural operator for flood routing (Muskingum–Cunge) (Brunton et al., 2016; Li et al., 2021)	new

AI-impact thread. When an emulator replaces an analysis that a design code sanctions, who is accountable for the result? Physics-informed models make that question concrete.

Unit 6. Agents, decisions, and the profession (L25–L28)

L	Methods	CEE anchor and data	CS 441
25	Agentic workflows: tool use, verification harnesses, human sign-off	An agent that runs a hydraulic model (WNTR) and must pass unit and mass-balance checks before it may report	new
26	Markov decision processes; value iteration; reinforcement learning	Bridge replacement policy from Illinois inventory transitions (NBI; Madanat et al., 1995); traffic-signal control in SUMO	L25
27	Fairness and accountability: disaggregated evaluation, model cards, datasheets	Auditing the L6 bridge model by county social vulnerability (CDC/ATSDR SVI): whose bridges are caught, flagged, and inspected	L21–22
28	AI as infrastructure; the future of the profession	Data centers' demands on power grids and water systems (EPA eGRID, EIA-930); licensure and the engineer of record; course synthesis	L21, L26

Exam 3 (CBTF, final-exam period): cumulative, with emphasis on Units 5–6. Final-project showcase.

Homework: core plus stretch goals

Each homework has a 100-point core and optional stretch goals, as in CS 441, and each includes one *critique-the-machine* problem: a plausible AI-generated analysis with a seeded physical or statistical error that students must find, explain, and fix.

Homework	Core (100 points)	Stretch goals
HW1 Tabular infrastructure data (after L8)	Nearest neighbors and ridge regression from scratch, verified against scikit-learn; random vs. spatial-block cross-validation on the NBI to measure the optimism of random splits; boosted trees with SHAP, checked against engineering expectations. Critique: a concrete-strength analysis that extrapolates beyond the tested mixes.	Monotone constraints from Abrams' law; an ordinal model of condition ratings.
HW2 Uncertainty and risk (after L12)	Leak detection on BattLeDIM with a threshold set by the cost of each error; conformal streamflow intervals at Illinois gauges with a coverage audit across decades. Critique: a forecast that leaks future rainfall.	Gaussian-process surrogate with active learning on a WNTR network; quantile regression for design values.
HW3 Inspection and sensing (after L16)	MLP and backpropagation from scratch with a gradient check; CNN transfer learning on SDNET2018 with results reported separately for decks, walls, and pavements. Critique: a crack classifier validated on photos of the same slab it was trained on.	U-Net segmentation with crack width in millimeters; flood mapping on SpaceNet 8.
HW4 Foundation models on the job (after L20; teams of three)	Can a foundation model monitor a construction site? (portfolio §4: CLIP on openly licensed site photos and the team's own, with disaggregated evaluation and a memo). Critique: an AI answer on work-zone signage that cites a fabricated clause of MUTCD Part 6.	Prompt ensembling; a fine-tuned CNN against its compute cost; temporal smoothing; a data-collection plan for the worst subgroup; an estimate of the energy and water used by the team's own computing.
HW5 Structure, physics, and decisions (after L26)	PINN for 1-D advection–dispersion compared with a finite-difference solver, with the Péclet numbers at which it fails; graph message passing from scratch for pressure estimation in a water network against a no-graph baseline. Critique: a surrogate that violates mass balance.	Inverse Terzaghi problem (identify c_v); SINDy on a structural response; PointNet on the L22 indoor scans; a verified WNTR agent; a bridge-maintenance MDP; LSTM rainfall–runoff against a linear-reservoir baseline and a pretrained forecaster.

Final project

As in CS 441, teams of two to four define a custom ML problem, collect or assemble the data, compare models, and report results. Here the problem must come from a CEE setting: the team's own measurements, or data and a question from a candidate Illinois partner such as a municipality, Illinois Extension, the Prairie Research Institute, campus Facilities & Services, or IDOT's open data. Deliverables: a datasheet for the dataset (Gebru et al., 2021), a model card with disaggregated results (Mitchell et al., 2019), a two-page engineer's memo with a recommendation, an impact statement that addresses the four dimensions the Task Force names (social, ethical, environmental, and economic), and a five-minute oral defense.

Grading and AI use

Grading is modeled on CS 441, so that the course carries the same weight on a transcript:

$$\text{Grade} = \frac{\text{XP} + \text{best exam} + \text{second-best exam} + \text{final project}}{\max(\text{XP},\, \text{XP}_{\text{target}}) + 300},$$

with $\text{XP}_{\text{target}} = 400$ for 3 credits and 525 for 4 credits. Experience points come from homework cores and stretch goals and from in-class activities; points beyond the target raise the grade. Ten free late days apply to homework; the final project cannot be late.

Component	Grainger AI-use level	Why
CBTF exams	Restricted	Measures unaided mastery; this is where the course certifies fluency.
Homework cores (from-scratch code and analysis)	Limited Use	AI may explain and help debug; the implementation and the analysis are the student's, because producing them is how the learning happens.
Critique-the-machine problems	Use When Instructed	Students work with AI output by design.
Stretch goals and final project	All Use, with a disclosure log	The grade rests on the team's own data, its verification, and its oral defense.

Delivery, materials, and accessibility

Every lecture comes with a notebook that follows the template of my applied machine learning notebooks: derivation, from-scratch implementation with shape assertions, comparison with a reference library at explicit tolerances, a worked example by hand, concept checks, and common pitfalls. Notebooks run in the browser with no installation. Lectures are recorded and captioned, and figures use colorblind-safe palettes with alt text.

Evaluating the course

The course is also a study site for the scholarship plan in portfolio §8: proctored CBTF scores measure unaided mastery; a pre/post task in which students audit AI-generated analyses with seeded errors measures error detection; project rubrics measure engineering judgment (ABET SO 6); and surveys record how students actually use AI. Results go to the department, ASEE, and the ASCE *Journal of Civil Engineering Education*.

Key readings beyond the texts

Angelopoulos & Bates (2023), conformal prediction, *Found. Trends ML*. Abreu et al. (2023), indoor point clouds with an as-built BIM model, *Data*. Brunton, Proctor & Kutz (2016), SINDy, *PNAS*. Crawford (2013), hidden biases in big data, *HBR*. Dorafshan, Thomas & Maguire (2018), SDNET2018, *Data in Brief*. Gebru et al. (2021), datasheets for datasets, *CACM*. Hänsch et al. (2022), SpaceNet 8, flooded roads and buildings, CVPR Workshops. Jakubik et al. (2023), Prithvi geospatial foundation model, arXiv. Kratzert et al. (2019), LSTM rainfall–runoff in large samples, *HESS*. Krishnapriyan et al. (2021), failure modes of PINNs, NeurIPS. Li, Yu, Shahabi & Liu (2018), diffusion convolutional recurrent networks for traffic, ICLR. Li et al. (2021), Fourier neural operator, ICLR. Li, Yang, Islam & Ren (2023), the water footprint of AI, arXiv. Maes & Lombaert (2021), KW51 railway bridge monitoring, *J. Bridge Eng*. Madanat, Mishalani & Ibrahim (1995), infrastructure transition probabilities, *J. Infrastructure Systems*. Milly et al. (2008), "Stationarity is dead," *Science*. Mitchell et al. (2019), model cards, FAT*. CDC/ATSDR (2024), Social Vulnerability Index 2022 database. Raissi, Perdikaris & Karniadakis (2019), physics-informed neural networks, *J. Comput. Phys.* Roberts et al. (2017), cross-validation for structured data, *Ecography*. Siler-Evans, Azevedo & Morgan (2012), marginal emissions factors, *Environ. Sci. Technol.* Strubell, Ganesh & McCallum (2019), energy costs of deep learning in NLP, ACL. Tixier et al. (2016), NLP for construction injury reports, *Automation in Construction*. Vrachimis et al. (2022), BattLeDIM, *J. Water Resour. Plann. Manage.* Yeh (1998), concrete strength modeling, *Cem. Concr. Res.*