AI Ethics

Problem Based Learning Cases

Macy Foundation / Millennium Conference • Preclinical M1/M2 • November 2026

Two problem-based learning cases. Case 1 targets ethical principles, transparency and disclosure, patient autonomy, and algorithmic bias. Case 2 targets the environmental impact of AI.

Each case defines the technical concepts it uses — probabilistic prediction, automation bias, model drift — in plain-language concept boxes, so no prior AI coursework is assumed.

Case 1 — "The Patient Who Asked First"

Objectives 1, 2, 3, and the cross-cutting objective • Groups of 5–7 • 30 minutes

The Scenario

Mr. Henley is a 71-year-old retired schoolteacher, here for an annual physical with Dr. Okafor — his physician for fifteen years.

Before the visit begins, he says: "My granddaughter told me some doctors are using AI now to write notes and figure out what's wrong. Is anything listening to us right now? And will an AI be reading my chart?"

Dr. Okafor's clinic uses four AI tools as part of standard practice:

1. An ambient AI scribe that records visit audio and drafts the clinical note. It is off for this visit.
2. A background clinical decision support (CDS) model that runs automatically on every chart and flags readmission and sepsis risk scores.
3. **A sandboxed clinical LLM that Dr.** Okafor sometimes uses to draft patient-friendly explanations of results. She edits and signs every draft.
4. A care-coordination referral algorithm that selects patients for the clinic's chronic-disease outreach program.

About that last tool: Dr. Okafor has noticed that patients from the largely immigrant neighborhood the clinic serves are rarely flagged for outreach, despite heavy chronic-disease burdens. The vendor's documentation reveals the model was trained to predict future healthcare costs, using past utilization and spending as its main inputs.

None of these tools has been specifically disclosed to Mr. Henley. The intake form mentions "electronic systems" but does not name AI. Other patients have not asked. Mr. Henley is asking now.

Concepts Used in This Case

PROBABILISTIC PREDICTION

The CDS model does not diagnose anyone. It outputs a probability — "this patient has a 12% estimated risk of readmission within 30 days" — computed from patterns in historical patient data.

A probability can be well-calibrated and still be wrong about any individual patient. That is why CDS output is an input to clinical judgment, never a verdict.

AUTOMATION BIAS

The human tendency to over-trust automated output — accepting the machine's suggestion in place of one's own assessment.

It produces two error types: acting on a false machine output (commission), and missing something because the machine missed it (omission).

Dr. Okafor's guard against it: form her own impression first, then look at the score, then investigate any disagreement.

MODEL DRIFT

The clinic's CDS model was validated four years ago. Since then, the neighborhood's demographics shifted, coding practices changed with a new EHR, and a nearby clinic closure altered who walks in.

When the patients a model sees stop resembling the patients it learned from, performance quietly degrades. Validation is a maintenance schedule, not a one-time event.

Discussion Questions

5. Briefly define the four principles of medical ethics — beneficence, non-maleficence, autonomy, and justice — and apply each to Dr. Okafor's situation in one or two sentences.
6. For each of the four AI tools, decide whether it should be disclosed to Mr. Henley today. Justify each decision with a specific ethical principle rather than clinic policy. Where the tools differ, what feature changes the obligation?
7. **Mr.** Henley asks Dr. Okafor to turn off all AI processing of his chart. The background clinical decision support model cannot be selectively disabled. How should she respond — and what does honesty require when a real opt-out does not exist?
8. **Suppose Mr.** Henley had never asked. Would Dr. Okafor still have an obligation to proactively disclose any of these tools? Which ones, and on what principle?
9. One colleague wants a disclosure form listing every AI tool by name; another says that would alarm patients without informing them. Critique both positions and propose a defensible disclosure standard for the clinic. [optional]
10. Consider the referral algorithm trained on healthcare costs. Why might it under-select the very patients Dr. Okafor worries about? Which principle is most at stake, what should the clinic do now, and who is accountable for acting on — or ignoring — its output? [optional]

Facilitator Guide — Case 1

Purpose. Students should be able to articulate a nuanced disclosure standard in their own words, and to explain how an apparently neutral training choice (predicting cost) becomes a biased clinical instrument.

Question 1 — The four principles

11. Beneficence: several tools plausibly improve care — documentation attention, earlier risk recognition, clearer explanations.
12. Non-maleficence: mis-transcription, false reassurance from an unvalidated score, biased referral patterns are all foreseeable harms.
13. **Autonomy: Mr.** Henley cannot accept or decline what he does not know about. His question is autonomy in action.
14. Justice: benefits and failures may be distributed unevenly — the referral algorithm makes this concrete.

Question 2 — Which tools require disclosure

Push students past a yes/no list toward a materiality analysis: what would a reasonable patient want to know?

15. The scribe: clear disclosure and agreement when in use — it records the patient. This end of the spectrum is not controversial.

16. The background CDS: system-level transparency plus an honest, plain-language answer when asked. Per-visit consent for background decision support is not the prevailing norm — students should be able to say why (it functions like other clinical infrastructure, one input among many).
17. **The drafting LLM: the judgment zone.** It shapes patient-facing content, so disclosure costs little and supports trust. Nearly all agree an honest answer is owed when asked.
18. The referral algorithm: it allocates a clinic resource. Affected patients have a justice-based interest in transparency about how outreach is allocated — and the clinic has a duty to interrogate the tool itself (Question 6).

Question 3 — The opt-out problem

19. Honesty is the obligation that survives the constraint: say plainly what runs, what it does, and what she does with its output.
20. A defensible script: “I can keep the scribe off and write your explanations myself. The risk-score system runs on every chart here and I can't switch it off for one patient — but here is what it does.”
21. Pretending an opt-out exists when it does not would be a worse autonomy violation than the constraint itself.

Question 4 — The proactive obligation

22. Yes, for the first two zones: autonomy attaches to the care actually delivered, not to the questions a patient happens to ask.
23. The patient who never asks is precisely the one whose autonomy is most easily bypassed.
24. But the obligation is calibrated, not total — which is why the zone framework matters.

Question 5 — The disclosure standard

Both colleagues fail in opposite directions. Listing everything buries the material in the trivial; the alarm argument is paternalism. A defensible standard is anchored in materiality — the reasonable-patient test students know from informed consent:

25. Disclose and seek agreement: tools that record the patient or substantially direct decisions.
26. Judgment zone — err toward telling: tools that shape patient-facing content or materially influence a diagnosis or plan.
27. System-level transparency: routine computational infrastructure, covered by honest policies and truthful answers to any question.

Facilitator note. Disclosure norms vary by tool, setting, and local requirements. Teach this as a framework for judgment, and follow applicable institutional policy.

Question 6 — Algorithmic bias

28. Mechanism: the model predicts cost, and historical cost reflects access, insurance, and trust — not need alone. Patients who faced barriers to care generated less spending at the same illness burden, so the model under-scores them. No demographic variable required. (This is the mechanism documented by Obermeyer and colleagues in a widely used commercial algorithm.)
29. Principle: justice, primarily — the tool distributes a clinic benefit inequitably. Non-maleficence joins once the clinic knows and continues unchanged.
30. What to do: audit selections against clinical need; change the target variable (predict conditions, not cost); add human review of near-misses; schedule re-validation (connect to the drift box).
31. **Accountability: the clinic and its clinicians.** The algorithm is a tool the clinic chose — responsibility does not transfer to the vendor's code. This is the 'tool, not team member' objective in institutional form.

Common pitfalls

32. Students may collapse the case into HIPAA compliance. Redirect: the obligations here are ethical and exist independent of regulation.
33. Students may say the background CDS 'needs no thought because the patient never sees it.' Push back with Question 6 — invisible tools can carry the largest justice stakes.

Debrief prompt: “What is the one sentence you wish your own doctor would be ready to say about AI in your care?”

Case 2 — “The Cooling Tower”

Objective 4 and the cross-cutting objective • Groups of 5–7 • 25 minutes

The Scenario

A medical school sustainability committee has been asked to advise on the institution's expanding AI portfolio: administrative chatbots, ambient clinical scribes, AI-assisted research computing, and a model-training partnership with two technology vendors.

A staff scientist presents the committee's first environmental accounting. Her four slides:

34. **Energy.** The IEA estimated that all data centers used about 415 TWh in 2024 (1.5% of global electricity), with a base-case projection of about 945 TWh in 2030. These figures include non-AI computing; they are not an estimate of AI alone. Source: IEA, Energy and AI, 2025.
35. **Water.** Data-center cooling and electricity generation can consume water. The amount depends on location, cooling technology, electricity supply, model, and workload; a single “bottle per chat” figure is not universal. Water stress makes the location of consumption important. Source: Li et al., Making AI Less Thirsty, Commun ACM. 2025;68(7):54–61. doi:10.1145/3724499.
36. **Hardware.** Specialized chips carry impacts from mineral extraction, manufacturing, replacement, and disposal. Assess the actual equipment lifecycle rather than assuming a fixed replacement interval.
37. **Downstream public health.** U.S. healthcare already accounts for ≈8.5% of national greenhouse-gas emissions, and climate change acts on health through heat illness, air quality, vector-borne disease, food insecurity, and displacement.

The committee chair turns to the dean: “When you decide whether to adopt a new AI tool for clinical care — is environmental impact part of the decision? Or only the clinical benefit?”

Concepts Used in This Case

TRAINING VS. INFERENCE — WHERE THE ENERGY GOES

Training and subsequent retraining can require substantial computation. Inference is the computation used for each request; its cost varies and can accumulate substantially at scale.

A hospital rarely trains models — but a hospital that embeds an LLM in every clinical note generates millions of inferences a year. Deployment decisions are environmental decisions, and “we didn't train it” does not end the institution's accounting.

DRIFT AND DECOMMISSIONING — THE MAINTENANCE ANGLE

Clinical safety and environmental stewardship converge on the same practice: post-deployment review.

A model left running without re-validation risks drift; a tool left running without utilization review burns energy for benefit no one is measuring. Scheduled review — and retirement of unused tools — serves non-maleficence and justice at once.

Discussion Questions

38. Using the four categories — energy, water, hardware, downstream public health — summarize the environmental impact of AI in healthcare in your own words. Which category was least familiar, and why does it matter clinically?
39. Apply justice: for a clinical AI tool at this institution, who benefits and who bears the environmental cost? Be specific about populations and geography.
40. Apply beneficence: name one deployment in the portfolio you would defend despite a meaningful footprint, and one you would question. What distinguishes them?
41. A member argues: “Environmental review is a distraction — medicine has bigger problems, and the AI runs on the vendor's servers anyway.” Critique both halves. Whose tool is it, ethically, once the institution deploys it? [optional]
42. Propose one procurement policy that integrates environmental impact without stalling genuinely beneficial tools. Specify what it requires, from whom, and when it triggers review or decommissioning. [optional]

Facilitator Guide — Case 2

Question 1 — The footprint

43. Hold the numbers loosely and the structure firmly: energy (inference dominates over a model's life), water (regional inequities), hardware (mining to e-waste), downstream health (burdens land first on populations least likely to benefit).

Question 2 — Justice

44. Beneficiaries: the institution, its clinicians, its patients — largely in a resourced setting.
45. Cost-bearers: communities hosting data centers, water-stressed regions, climate-vulnerable populations, future patients.
46. The point: they are usually different people, and the burdened rarely consented. The asymmetry — not the footprint itself — is the justice problem.

Question 3 — Beneficence calibration

47. The reasoning is the assessment target, not the picks. Strong answers weigh magnitude and distribution of clinical benefit against footprint, per deployment.
48. Correct categorical answers in either direction: 'all AI is wasteful' and 'clinical benefit always wins' both fail.

Question 4 — Critiquing the dismissal

49. Climate change is a health problem; 'bigger problems' treats downstream morbidity as someone else's ledger — the reasoning that once excused antibiotic overuse and low-value imaging.
50. Outsourcing the computation does not outsource the accountability. The institution chose the deployment — the 'tool, not team member' objective at institutional scale.

Question 5 — Policy proposals

51. Acceptable: vendor impact disclosure in procurement; tiered review by clinical stakes; preference for renewable-powered vendors; scheduled utilization and re-validation review with decommissioning; an environmental line item in any AI business case.
52. Push for specificity: who reports, on what schedule, with what threshold for action.

Debrief prompt: “You now chair procurement. What is the first question you ask every vendor — and would you defend your portfolio to the community hosting the data center?”