SESSION 2

AI in Clinical Skills and Skills Assessment

Clinical reasoning, documentation, and reflective practice

Preclinical M1/M2

OBJECTIVE 1

Learning objectives: clinical tools

Recognize AI-enabled tools and the tasks they support.

Distinguish tasks AI can augment from competencies learners must practice independently.

Apply institutional and clerkship policies for privacy, permitted use, and disclosure.

OBJECTIVE 2

Learning objectives: clinical reasoning

Explain how input quality, training data, and model assumptions affect suggestions.

Form an independent assessment, then integrate, question, or override AI output with justification.

Recognize bias, uncertainty, and threats to foundational skill development.

OBJECTIVE 3

Learning objectives: documentation

Describe ambient listening and AI scribe capabilities and limitations.

Review and correct draft notes for accuracy, completeness, and fidelity to the patient.

Preserve attention, empathy, and patient-centered communication.

OBJECTIVE 4

Learning objectives: reflective practice

Use AI selectively and transparently to support personal reflection and learning.

Balance AI suggestions with clinical judgment and patient values in shared decisions.

OPENING CASE

Three tools in one visit

A risk flag	The EHR displays a score generated before the visit.
A decision-support tool	A system can suggest a differential from entered information.
An ambient scribe	A microphone and language model can draft the note.

Four families of clinical tools

Clinical decision support	Differentials, risk scores, workup suggestions, and alerts
Ambient scribes	Audio capture, transcription, and draft notes
Documentation aids	Chart summaries, draft messages, and text completion
Skill-tracking platforms	Learner performance records and feedback signals

How a clinical suggestion is produced

1. Inputs	History, examination, laboratory results, and information already in the chart
2. Model	Patterns, labels, assumptions, and limitations of the system
3. Output	A suggestion, classification, or risk estimate requiring interpretation

Failure modes and matched safeguards

Biased target	Ask whether the predicted label represents clinical need.
Validation gap or drift	Look for external evaluation and ongoing monitoring.
Automation bias or alert fatigue	Compare output with an independent assessment and investigate discrepancies.

Obermeyer 2019; Wong 2021; Goddard 2012.

What an ambient scribe does

Capture	Records encounter audio under the applicable consent and privacy process.
Transcribe	Converts speech to text, with potential recognition errors.
Summarize	Generates a draft note, which may omit, transform, or add content.
Review	A clinician checks, corrects, and attests according to their role.

Tierney et al. NEJM Catalyst. 2024.

Three ways a note can be wrong

Omission	A consequential symptom or concern disappears.
Transformation	The patient's meaning changes during summarization.
Fabrication	The draft adds a finding, dose, or negative that was never established.

Koenecke et al. PNAS. 2020.

Documentation aids and skill trackers

Documentation aids

Check management-changing summaries against the original record.
Read drafted messages as the patient will.
Watch for propagated errors.

Skill trackers

Know what is measured, who sees it, and what decisions it informs.
Use the data as one input to a learning plan.

Verification depends on the task

Generated content

Check claims against sources and notes against the encounter.
Look for invented content, omissions, and contextual mismatch.

Prediction and classification

Examine local validation, calibration, subgroup performance, and change over time.
Compare with the clinical assessment.

Augmentation and independent competence

AI can support

Checking a differential
Drafting a note
Summarizing records
Finding alternatives to evaluate

Learners must practice

History-taking and examination
Clinical reasoning
Communication and judgment
Building a therapeutic relationship

Deskilling, mis-skilling, and never-skilling

Deskilling	Losing a previously developed skill through reduced practice
Mis-skilling	Learning the task inaccurately through how the tool shapes practice
Never-skilling	Failing to develop a skill because the tool performs the learning work

Cabitza et al. JAMA. 2017; Budzyń et al. Lancet Gastroenterol Hepatol. 2025.

CONSTRUCTED TEACHING CASE

The tempting shortcut

A tired M3 opens decision support before seeing a patient with a sore throat. She copies a five-diagnosis differential into the note.

Her preceptor asks: “Walk me through how you got there.”

She cannot explain why gonococcal pharyngitis is included; she has not taken the relevant history.

Reason first, then consult

Commit	Write a problem representation, differential, and initial plan.
Consult	Use an approved tool with accurate, relevant inputs.
Compare	Investigate disagreements and missing possibilities.
Decide	Accept, revise, or override with a defensible explanation.

Goh et al. JAMA Netw Open. 2024;7:e2440969.

Five questions when reviewing output

Coverage	What did each assessment include or miss?
Verification	How will I check a consequential claim?
Concordance	Does the recommendation match applicable guidance?
Omissions	Which red flags or alternatives remain unaddressed?
Input sensitivity	Which missing or uncertain detail could change the answer?

A source check has four parts

Existence	Find the source in a database or journal.
Content	Confirm that it supports the claim.
Currency and quality	Check the publication type, date, and relevant updates.
Applicability	Assess population, setting, and relevance to the patient.

Turning disagreement into learning

Identify a verified gap	Which relevant possibility did I miss, and why?
Set a learning goal	Turn the discrepancy into a focused question or reading task.
Revisit the reasoning	Ask the tool to critique an assessment you have already developed.

DOCUMENTATION CASE

What the patient said and what the draft recorded

Patient account

“When I lie down flat… I cannot get the air.” Mrs. Alvarez sleeps in a chair on some nights.

Draft note

“Sleeps in a recliner by preference.”
“Negative for shortness of breath.”

A practical note-review workflow

Anchor	Record your own brief account of the encounter.
Compare	Check the draft against the patient account and source data.
Inspect	Look for omissions, changed meaning, and invented details.
Correct and attest	Resolve uncertainty and sign only within your knowledge and role.

Presence during an AI-assisted encounter

Potential benefits

Less keyboard attention
More time for conversation
Reduced documentation burden

Habits that preserve presence

Keep a brief independent summary.
Restate important symptoms accurately.
Check your understanding with the patient.

Privacy, permission, and disclosure

Patient information	Use only systems authorized for the data and task.
General learning	Use public tools only where policy permits and without patient information.
Disclosure	Explain use when required or material, and answer questions honestly.

Equity requires more than careful users

Risks to assess

Sparse or fragmented records
Unequal speech-recognition performance
Biased labels and historical patterns

Responsibilities

Gather relevant information.
Question unsupported assumptions.
Audit subgroup performance and workflow effects.
Report recurring problems.

Obermeyer et al. Science. 2019; Koenecke et al. PNAS. 2020.

AI in reflective practice

Begin with your own account	What happened, what did I notice, and what remains uncertain?
Use comparison selectively	Compare summaries, reasoning, or feedback to identify a learning need.
Be transparent	Describe AI assistance accurately when work is shared or assessed.

When a patient brings AI advice

Ask	What prompted the search, and what matters most to the patient?
Appraise	Evaluate the recommendations on their clinical merits.
Decide together	Explain what fits, what does not, and how patient preferences affect the plan.

Clinical habits to practice

Form your own assessment.

Check consequential claims and sources.

Use disagreements to identify learning needs.

Review notes against the encounter.

Keep patient values and professional responsibility in view.

References

Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517–518.

Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Network Open. 2024;7(10):e2440969.

Tierney AA, Gayre G, Hoberman B, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catalyst. 2024.

Koenecke A, Nam A, Lake E, et al. Racial disparities in automated speech recognition. PNAS. 2020;117(14):7684–7689.

References

Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–453.

Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Internal Medicine. 2021;181(8):1065–1070.

Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. JAMIA. 2012;19(1):121–127.

Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digital Medicine. 2020;3:17.

References

Russell RG, Lovett Novak L, Patel M, et al. Competencies for the use of artificial intelligence–based tools by health care professionals. Academic Medicine. 2023;98(3):348–356.

Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol. 2025;10:896–903.

American Medical Association. Professionalism in Health Care Systems. Code of Medical Ethics, Opinion 11.2.1.