SESSION 1

Introduction to AI Concepts in Healthcare

Core concepts for medical learning and clinical care

Preclinical M1/M2

Learning objectives

1. Taxonomy

Distinguish major AI methods and match them to clinical tasks.

2. Generative AI

Explain how language models produce text and why plausible errors occur.

Learning objectives

3. Prompt engineering	Construct and refine prompts with role, task, constraints, and verification anchors.
4. Performance metrics	Interpret accuracy, sensitivity, specificity, positive predictive value, and calibration.
5. Model lifecycle	Explain training, validation, deployment, generalization, and drift.

TEACHING POINT 1

AI is a family of methods

Ask which method, which task, and which failure mode.

Machine learning

Learning from data

Models learn patterns from examples rather than relying only on explicitly programmed rules.
Deep learning uses multilayer neural networks.

Clinical uses

Risk scores, deterioration alerts, readmission predictions, and imaging triage. Inputs may include laboratory results, vital signs, codes, or images.

Rajkomar, Dean & Kohane. N Engl J Med. 2019.

Computer vision and clinical AI history

Rule-based systems	Expert systems and conventional decision support
Statistical prediction	Models trained on structured clinical data
Deep learning	Image classification in radiology, ophthalmology, and pathology
Generative AI	Language, images, and systems that combine multiple modalities

Liu et al. Lancet Digit Health. 2019;1:e271–e297.

Language models, retrieval, and agents

Large language models	Generate text from patterns learned during training and the supplied context.
Retrieval-augmented generation	Retrieve external documents and use them to inform an answer.
AI agents	Use tools and take multiple steps, such as searching a chart or drafting an order.

Different outputs call for different checks

Prediction and classification

Output: a score, probability, or label. Check validation, calibration, subgroup performance, and changes over time.

Generative systems

Output: text, images, or proposed actions. Check factual claims, source fidelity, omissions, and fit to the patient.

Activity: identify one tool, its output, and its most important check.

TEACHING POINT 2

Text generation and retrieval are different processes

A language model generates an answer.
A retrieval tool may supply the evidence it uses.

How a language model produces text

Tokens and context	Predicts a distribution of possible next tokens from the preceding context.
Learned patterns	Uses patterns learned from large training corpora and subsequent training.
Variable output	Different settings or repeated runs can produce different responses.
Useful applications	Drafting, summarizing, explaining, and reasoning tasks require task-specific evaluation.

What fluent output does not establish

Accuracy	A well-written answer may contain factual errors.
Retrieval	A citation-looking string does not prove that a source was consulted.
Applicability	A true general statement may not fit this patient or setting.
Verification	Self-confidence and self-consistency do not replace an external check.

Hallucination and persuasive fabrication

Hallucination	False or unsupported content presented as an answer.
Persuasive fabrication	A fluent, coherent explanation that is nevertheless wrong.
Checks that help	Use appropriate sources, inspect retrieved passages, and verify consequential claims.

Bender et al. FAccT. 2021. See notes for context.

CONSTRUCTED TEACHING EXAMPLE

Spot the problem

"In cardiogenic shock, the failing left ventricle leads to decreased preload, which is the primary driver of decreased cardiac output."

With a partner: is the explanation sound?
What would you check, and against which source?

Facilitator answer and clinical nuance are in the notes.

TEACHING POINT 3

Prompt design

Role, task, constraints, and verification anchors

Anatomy of a useful prompt

Role	Specify the audience and the kind of assistance needed.
Task and format	Define the question, scope, and desired output.
Constraints	State relevant boundaries, terminology, and exclusions.
Verification anchors	Ask for traceable sources and explicit uncertainty; then check them.

The same learning task, scoped more clearly

Broad request

“Tell me about acute kidney injury.”
Scope, level, and verification method are unspecified.

Structured request

Name the learner level, etiologic categories, output format, and constraints.
Request source anchors that you will check.

Next activity: compare Prompt A, Prompt B, and Prompt C.

Choosing an appropriate system

Institutionally approved use

Use a system authorized for the data and purpose.
Follow local access, privacy, documentation, and disclosure requirements.

Public or consumer tools

Use only for permitted tasks without patient information.
Removing a name alone does not make a clinical narrative safe to share.

TEACHING POINT 4

What does “95% accurate” mean?

A performance headline needs a denominator and a clinical context.

VOTE BEFORE THE EXPLANATION

Would you deploy this screening tool?

A disease has 1% prevalence.
A screening model reports 99% accuracy.

A. Deploy
B. Do not deploy
C. Cannot tell from this information

WORKED EXAMPLE

A model can be 99% accurate and miss every case

99%

Accuracy if every patient is labeled negative

100 people screened
1 person has disease
99 people do not

The model identifies none of the cases. Sensitivity is 0%.

Four performance questions

Sensitivity	Among patients with disease, how many are detected?
Specificity	Among patients without disease, how many test negative?
Positive predictive value	Among those flagged, how many actually have disease?
Calibration	Among predictions near 30% risk, does the event occur about 30% of the time?

Generalization to new patients

Overfitting

A model learns patterns specific to its training data, including noise.
Performance falls on new data.

Underfitting

A model fails to capture important relationships.
Performance is poor even where it was developed.

Ask: how does it perform on data it was never trained on?

TEACHING POINT 5

Performance can change after deployment

A model needs continuing evaluation in the setting where it is used.

The model lifecycle

1. Train	Learn patterns from historical data.
2. Validate	Test on held-out data and relevant external populations.
3. Deploy	Evaluate fit with actual patients and workflows.
4. Monitor	Track performance and decide when to recalibrate, update, or retire.

How drift can arise

Patient mix	The populations and case mix change.
Clinical practice	Assays, devices, coding, or documentation workflows change.
Disease patterns	Prevalence and relationships between predictors and outcomes change.

Five concepts to carry forward

Identify the method and task.

Distinguish generated claims from retrieved evidence.

Design prompts with a verification plan.

Interpret performance in its clinical context.

Validate locally and monitor after deployment.