AI in Clinical and Translational Medicine

Course Administrators: Matthew Rizzo (matthew.rizzo@unmc.edu), Howard Fox (hfox@unmc.edu), Carrie Bailey (carrbailey@unmc.edu), Dorzhey Matkhanov (dmatkhanov@unmc.edu), Laura Lienemann (laulienemann@unmc.edu),

Course Format: The course will run for six weeks and consist of weekly lectures with in-person and virtual attendance options. Students are expected to complete the readings before class and come to each class prepared to ask questions.

Executive Summary

Artificial intelligence (AI) is increasingly embedded in clinical care, translational research, and health system operations. Yet many clinicians and investigators encounter AI primarily through tools rather than through a clear understanding of underlying assumptions, methodological limitations, and system-level risks. **AI in Clinical and Translational Medicine** is a six-week, clinician-facing course designed to build foundational literacy, sound judgment, and institutional readiness for responsible AI use within academic health systems.

The course emphasizes conceptual clarity, critical appraisal, and translational relevance rather than coding or algorithm development. It examines how data quality, bias, workflow, human factors, and governance shape real-world performance. Over six weeks, participants move from AI fundamentals and data literacy to interpretability, longitudinal reasoning, ethics, regulatory science, and integration within Learning Health Systems and Learning Research Systems.

The curriculum aligns with NIH priorities, FDA Good Machine Learning Practice, and Learning Health System frameworks and applies across clinical specialties. An optional, lightweight capstone supports applied reflection without technical implementation, positioning the course as both a standalone offering and an entry point to extended training. Although grounded in the UNMC environment, the structure emphasizes transferable principles and can be adopted or adapted by other academic medical centers, CTSA hubs, multisite research networks, professional societies, and continuing medical education programs.

Introduction and Overview

AI now operates across clinical care, translational research, and health system infrastructure. Clinicians, investigators, trainees, and institutional leaders often encounter AI through outputs and interfaces rather than through a working understanding of the methods, assumptions, and contextual factors that determine performance in practice.

The six-week course, **AI in Clinical and Translational Medicine**, provides a focused introduction to how artificial intelligence and machine learning are used, evaluated, and governed in academic health systems. The emphasis is on conceptual clarity, critical appraisal, and translational relevance rather than coding or model development. Participants examine what AI can and cannot do in clinical settings, how performance depends on data quality, context, and time, and why failures frequently arise from human, organizational, and sociotechnical factors rather than models alone. AI is treated explicitly as a sociotechnical system in which outcomes emerge from interactions among algorithms, data, clinicians, workflows, institutional policies, and regulatory constraints.

The curriculum integrates core clinician AI principles (Appendix C) across the translational lifecycle. Early sessions establish shared mental models around AI fundamentals, data literacy, and prompt engineering as a structured clinical reasoning interface. Mid-course sessions address interpretability, longitudinal data, disease progression, digital twins, and causal reasoning, supporting movement from static prediction toward clinically meaningful decision support. Final sessions focus on ethics, bias, human factors, governance, regulatory science, and AI's role within Learning Health Systems and Learning Research Systems.

The course applies across clinical specialties and practice settings. Core concepts generalize across disciplines, while specialty relevance is illustrated through examples spanning medicine, surgery, psychiatry, radiology, pathology, anesthesiology, primary care, and population health. Appendices A and B provide structured examples demonstrating how course topics translate across diverse domains.

Readings are tailored to each session. A concise set of required, broadly accessible readings establishes shared foundations, with additional scholarly selections aligned to that week's clinical and translational focus. These include cognitive and neurologic care, diagnostic and visual specialties, procedural and perioperative care, and longitudinal and population-based practice, while maintaining a consistent conceptual framework.

Across six weeks, participants progress from foundational literacy to safe clinical and translational application, then to sustainable governance and continuous learning. Instruction emphasizes evidence appraisal, bias recognition, transparency, and clinical accountability and does not promote specific commercial products or clinical decision replacement. The structure aligns with NIH guidance, FDA Good Machine Learning Practice, and Learning Health System frameworks, treating ethics, data stewardship, and accountability as core requirements.

The six-week offering also serves as an entry point to the extended 48-week AI in Clinical and Translational Medicine course. Participation is recommended as an orientation and rapid start but is not a prerequisite; individuals with relevant experience may enter directly. Topics introduced here are revisited in greater depth in the extended course through applied case studies and sustained project-based work.

An optional, lightweight capstone (Appendix E) may be pursued concurrently with the 6-week course. The capstone is a short, concept-driven application exercise centered on problem formulation, clinical relevance, and translational feasibility without technical implementation. Projects may inform grant concept development, pilot or feasibility studies, quality improvement initiatives, governance planning, or preparation for advanced project-based training and may be expanded within the 48-week course.

The course is accessible to participants from clinical, research, and administrative backgrounds. No prior technical expertise is required. The goal is to equip participants with shared language, disciplined judgment, and a translational perspective needed to evaluate AI tools, collaborate with technical teams, and lead responsible integration of AI into clinical and research environments.

Although grounded in the UNMC clinical, research, and governance environment, the curriculum emphasizes transferable principles rather than local tools. Core domains include data literacy, model evaluation, interpretability, ethics, regulatory science, and integration within Learning Health Systems and Learning Research Systems. Institutional examples are illustrative and can be replaced with local data sources, governance structures, IRB processes, and health system workflows at other academic medical centers, community hospitals, CTSA hubs, and health systems. The course can therefore be adapted for regional or national delivery, multisite consortia, professional societies, or continuing education programs while preserving shared learning objectives and conceptual consistency. Appendix F defines key terms used throughout the curriculum.

In total, the course prepares clinicians and translational scientists not to build AI systems, but to evaluate, govern, and integrate them responsibly within real clinical and research environments.

Week / Topic
Week 1. AI and Machine Learning at UNMC: Foundations, Resources, and Prompt Engineering
Week 2. Core Machine Learning Concepts for Clinicians and Translational Scientists
Week 3. Interpretability, Temporal Data, and Real-World Signals
Week 4. Disease Progression, Digital Twins, and Causal Reasoning
Week 5. Ethics, Governance, Human Factors, and Regulatory Science
Week 6. Translational Integration: From Models to Learning Health Systems

At least one confirmed speaker from UNMC or the broader NU system has been identified for each session. Additional speakers may rotate depending on schedules and availability as invitations and dates are finalized.

Educational Credit: Continuing Medical Education (CME) credit will be provided for eligible participants, in accordance with institutional and accreditation requirements. For those that do not need CME, “micro credentialing” certification will be available (e.g. badges) through the UNMC Office of Interactive E-Learning.

Course evaluation TBD: Qualtrics, REDCap, CES (formerly EvaluationKIT) by Watermark, or Assessment, Evaluation, Feedback, Intervention System (AEFIS). Course evaluation will emphasize participation, critical reflection, and applied reasoning. Optional capstone artifacts will be used for formative course improvement and refinement rather than summative assessment. Evaluation reports may be used to demonstrate course impact for future funding (Appendix E).

Week 1. AI and Machine Learning at UNMC: Foundations, Resources, and Prompt Engineering

Learning Objectives emphasize recognition, interpretation, and informed judgment rather than technical implementation or independent system deployment.

By the end of this session, participants will be able to:

- Recognize common uses and limitations of AI and machine learning in clinical and translational contexts.
- Describe the UNMC AI ecosystem and appropriate high-level clinical, research, and administrative use cases.
- Identify characteristics of effective and ineffective prompts for clinical and translational tasks.
- Recognize common failure modes of large language models, including hallucinations and alignment artifacts.
- Demonstrate familiarity with principles of concise and responsible prompting (green prompting).

The opening session establishes a shared understanding of artificial intelligence and machine learning as they are practiced at UNMC, while orienting participants to the local ecosystem of tools, data resources, and institutional supports available for clinical, translational, and health systems research. In addition to clarifying what AI is and is not in medicine, this session explicitly reviews prompt engineering as a practical, immediately usable skill for clinicians, researchers, and administrators. Discussion focuses on how prompt formulation shapes outputs, bias, transparency, and overreliance, with concrete examples relevant to literature synthesis, data exploration, documentation support, and hypothesis generation. The session will also include a high-level overview of how large language models are trained and aligned, with emphasis on limitations relevant to clinical and translational use, including optimization for human preference rather than objective truth and the origins of hallucinations and related artifacts. *Green prompting* will be introduced as the practice of using concise, task-appropriate prompts that reduce unnecessary computation, cost, and overgeneration while supporting transparent, efficient, and responsible clinical use. The session frames realistic expectations for AI performance and emphasizes the central role of problem formulation and clinical context.

Required common reading (accessible)

Liu, J., Liu, F., Wang, C., & Liu, S. (2025). Prompt engineering in clinical practice: A tutorial for clinicians. Journal of Medical Internet Research, 27, e72644. https://doi.org/10.2196/72644

Additional scholarly reading

Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. New York, NY: Basic Books. https://www.basicbooks.com/titles/eric-topol/deep-medicine/9781541644649/

Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Hoboken, NJ: Pearson. https://aima.cs.berkeley.edu/

Week 2. Core Machine Learning Concepts for Clinicians and Translational Scientists

Learning Objectives

By the end of this session, participants will be able to:

- Recognize major categories of machine learning used in clinical and translational research.
- Interpret at a high-level common performance metrics reported in clinical AI studies.
- Identify common sources of bias, overfitting, and limited generalizability.
- Articulate why apparent model accuracy may not translate into clinical usefulness.

This session grounds discussion in the core ideas underlying machine learning as it is used in clinical research and healthcare operations. The session examines supervised and unsupervised learning, model families, performance metrics, and generalizability, with particular attention to why apparent accuracy can fail to translate into clinical usefulness. Discussion emphasizes interpretation and critical appraisal rather than implementation details. The session will focus on how to interpret AI performance metrics and research claims, including why commonly reported results may fail to generalize across populations, settings, or time. Classic machine learning approaches such as random forests and gradient-boosted models are used as reference points, while extending the discussion to contemporary AI and large language model evaluation (deep learning), including robustness testing, red-teaming, safety guardrails. The session will also introduce prompt hacking, defined as attempts to manipulate model outputs through input phrasing rather than evidence quality, and jailbreaking, a related practice that seeks to bypass built-in safety constraints. These practices are discussed to support evaluation, risk assessment, and governance planning, and not to encourage circumvention or misuse of deployed systems.

Required common reading (accessible)

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

Additional scholarly reading

Hastie, T., Tibshirani, R., & Friedman, J. (2017). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). New York, NY: Springer. https://link.springer.com/book/10.1007/978-0-387-84858-7

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Week 3. Interpretability, Temporal Data, and Real-World Signals

Learning Objectives

By the end of this session, participants will be able to:

- Describe why interpretability matters in clinical AI applications.
- Recognize situations in which explanations may mislead rather than clarify.
- Identify challenges associated with longitudinal and real-world data streams.
- Discuss how temporal context influences trust and clinical interpretation.

This session focuses on trust and understanding in AI-enabled clinical systems. Discussion explores why explainability matters, when explanations help, and when they may mislead. The group then considers real-world and time-dependent data streams such as wearables, speech, sleep, and driving behavior, and how these challenge traditional clinical assumptions and evaluation frameworks.

Required common reading (accessible)

London, A. J. (2019). Artificial intelligence and black-box medical decisions: Accuracy versus explainability. Hastings Center Report, 49(1), 15–21. https://doi.org/10.1002/hast.973

Additional scholarly reading

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Week 4. Disease Progression, Digital Twins, and Causal Reasoning

Learning Objectives

By the end of this session, participants will be able to:

- Recognize the distinction between prediction and causal reasoning in clinical modeling.
- Describe the concepts of disease progression models and digital twins.
- Identify assumptions and limitations underlying counterfactual or simulated outcomes.
- Discuss when AI outputs are exploratory versus potentially actionable.

This session shifts discussion from cross-sectional prediction to longitudinal reasoning. Participants examine how disease progression models are constructed and validated, using neurodegenerative disease as a motivating example. Digital twins are discussed as evolving, integrative representations of patients and populations. Causal reasoning and counterfactual thinking are emphasized as necessary for translating AI outputs into clinical decisions and trial design.

Required common reading (accessible)
Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. New York, NY: Basic Books. https://www.basicbooks.com/titles/judea-pearl/the-book-of-why/9781541698963/

Required reading: Chapter 1. Optional deeper reading: Chapters 4, 7, and 8.

Additional scholarly reading

Young, A. L., Oxtoby, N. P., Huang, J., Marinescu, R. V., Daga, P., Cash, D. M., Fox, N. C., Ourselin, S., Schott, J. M., Alexander, D. C., & Alzheimer's Disease Neuroimaging Initiative (2015). Multiple Orderings of Events in Disease Progression. Information processing in medical imaging: proceedings of the ... conference, 24, 711–722. https://doi.org/10.1007/978-3-319-19992-4_56

Corral-Acero, J., Margara, F., Marciniak, M., et al. (2020). The "digital twin" to enable the vision of precision cardiology. European Heart Journal, 41(48), 4556–4564. https://doi.org/10.1093/eurheartj/ehaa159

Week 5. Ethics, Governance, Human Factors, and Regulatory Science

Learning Objectives

By the end of this session, participants will be able to:

- Identify ethical risks and bias across the AI lifecycle.
- Recognize how human factors and organizational context contribute to AI failure or misuse.
- Describe governance, oversight, and regulatory considerations for clinical AI deployment.
- Discuss risks and tradeoffs associated with synthetic data.

This session addresses the ethical, organizational, and regulatory realities of deploying AI in healthcare. Discussion centers on bias and fairness, human–AI interaction, automation risk, and governance across the AI lifecycle. Regulatory science is addressed broadly, encompassing FDA-regulated medical devices and non-device AI applications such as documentation support, workflow augmentation, operational analytics, and research infrastructure, all of which require institutional oversight even when formal regulatory clearance is not required. The session will also address how bias can be introduced and propagated across the AI lifecycle, including through data selection, labeling practices, model updating, and downstream use. Attention will be given to synthetic dataset construction, including when synthetic data may mitigate bias or privacy risk and when it may amplify existing distortions or create false assurances of representativeness. Emphasis is placed on the idea that many AI failures are sociotechnical failures rather than purely algorithmic ones.

Required common reading (accessible)
National Academies of Sciences, Engineering, and Medicine. (2025). Machine learning for safety-critical applications: Opportunities, challenges, and a research agenda (Executive Summary). Washington, DC: National Academies Press. https://doi.org/10.17226/27970

Additional scholarly reading
Wiens, J., Saria, S., Sendak, M., et al. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

Health, C. for D. and R. (2025). Good Machine Learning Practice for Medical Device Development: Guiding Principles. *FDA*. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

Week 6. Translational Integration: From Models to Learning Health Systems

Learning Objectives

By the end of this session, participants will be able to:

- Describe how AI fits within Learning Health and Learning Research Systems.
- Recognize the importance of monitoring, updating, and adaptation over time.
- Identify clinician roles in responsible AI integration and oversight.
- Discuss how AI supports, rather than replaces, clinical judgment.

The final session situates AI within Learning Health Systems and Learning Research Systems. Discussion focuses on how data infrastructure, interoperability standards, and real-world evidence enable continuous learning and improvement. The session concludes by considering how AI moves from models to sustained clinical, research, and policy impact within academic health systems.

Required common reading (accessible)

Friedman, C., Rubin, J., Brown, J., Buntin, M., Corn, M., Etheredge, L., Gunter, C., Musen, M., Platt, R., Stead, W., Sullivan, K., & Van Houweling, D. (2015). Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. Journal of the American Medical Informatics Association: JAMIA, 22(1), 43–50. https://doi.org/10.1136/amiajnl-2014-002977

Additional scholarly reading

Friedman, C. P., Wong, A. K., & Blumenthal, D. (2010). Achieving a nationwide learning health system. Science translational medicine, 2(57), 57cm29. https://doi.org/10.1126/scitranslmed.3001456

Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., Suchard, M. A., Park, R. W., Wong, I. C., Rijnbeek, P. R., van der Lei, J., Pratt, N., Norén, G. N., Li, Y. C., Stang, P. E., Madigan, D., & Ryan, P. B. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Studies in health technology and informatics, 216, 574–578.

Appendices

Appendices A-F extend the core curriculum by providing specialty-specific examples, principle-to-week mappings, capstone alignment tools, evaluation instruments, and a shared glossary. They support cross-disciplinary applicability, consistent terminology, and responsible integration within our Learning Health and Learning Research Systems.

Appendix A. Cross-Disciplinary Relevance of Weekly Sessions

Appendix B. Surgery (General and Subspecialties): Relevance by Week

Appendix C. Mapping Key Clinician AI Principles to the Curriculum

Appendix D. Alignment of Capstone Projects With the Six-Week Course and Clinical Specialties

Appendix E. Example of Curriculum Evaluative Questions

Appendix F. Glossary

Appendix A

Cross-Disciplinary Relevance of Weekly Sessions

This appendix provides brief, illustrative examples of how each weekly session in the AI in Clinical and Translational Medicine curriculum is relevant across multiple clinical and translational disciplines. The intent is to highlight applicability and foster engagement across specialties, not to provide exhaustive coverage for any single field.

<u>Week 1. AI and Machine Learning at UNMC: Foundations, Resources, and Prompt Engineering</u>

Medicine / Research / Ethics
Applying prompt engineering to synthesize clinical guidelines, draft research questions, or explore hypotheses while maintaining transparency, provenance, and professional accountability.

Radiology
Using AI-assisted prompts to summarize imaging reports, reconcile longitudinal findings, or flag discrepancies across studies, with awareness of local PACS infrastructure and validation limits.

- Akinci D'Antonoli, T., Stanzione, A., Bluethgen, C., Vernuccio, F., Ugga, L., Klontzas, M. E., Cuocolo, R., Cannella, R., & Koçak, B. (2024). Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagnostic and Interventional Radiology*, 30(2), 80-90. https://doi.org/10.4274/dir.2023.232417

Anesthesiology
Leveraging prompts to review perioperative risk factors or summarize anesthesia records, while recognizing automation bias in time-sensitive clinical environments.

- Daccache, N., Zako, J., Morisson, L., & Laferrière-Langlois, P. (2025). The applications of ChatGPT and other large language models in anesthesiology and critical care: A systematic review. *Canadian Journal of Anesthesia,* 72(6), 904–922. https://doi.org/10.1007/s12630-025-02973-9
- Chung, P., Fong, C. T., Walters, A. M., Aghaeepour, N., Yetisgen, M., & O'Reilly-Shah, V. N. (2024). Large Language Model Capabilities in Perioperative Risk Prediction and Prognostication. *JAMA Surgery*, 159(8), 928–937. https://doi.org/10.1001/jamasurg.2024.1621

Neurology
Prompting AI tools to organize complex longitudinal histories, such as cognitive decline or multiple sclerosis progression, while avoiding spurious inference or overconfidence.

- Moura Junior, V., Hadar, P., Murphy, S., & Moura, L. M. V. R. (2025). AI Prompt Engineering for Neurologists and Trainees. *Seminars in Neurology*, 10.1055/a-2742-2349. Advance online publication. https://doi.org/10.1055/a-2742-2349

Ophthalmology
Using prompts to synthesize rapidly evolving evidence in retinal disease screening or glaucoma management while understanding model scope and limitations.

- Kleinig, O., Gao, C., Kovoor, J. G., Gupta, A. K., Bacchi, S., & Chan, W. O. (2024). How to use large language models in ophthalmology: From prompt engineering to protecting confidentiality. *Eye*, 38(4), 649–653. https://doi.org/10.1038/s41433-023-02772-w

Pathology
Using prompt engineering to summarize pathology reports, harmonize terminology across subspecialties, or assist with literature review on diagnostic criteria, while preserving diagnostic responsibility and verification.

- Cheng J. (2024). Applications of Large Language Models in Pathology. *Bioengineering* (Basel, Switzerland), 11(4), 342. https://doi.org/10.3390/bioengineering11040342

Psychiatry

Using prompt engineering to explore clinical documentation, psychoeducation materials, and research questions related to mental health while maintaining clear boundaries between support tools and clinical care. This includes understanding the role of mental health chatbots and social or emotional robots as adjunctive tools, not substitutes for psychiatric judgment. Examples include reviewing how tools like Woebot are positioned as supportive, skills-based interventions rather than diagnostic or therapeutic replacements.

- Grabb, D. (2023). The impact of prompt engineering in large language model performance: a psychiatric example. *Journal of Medical Artificial Intelligence*, 6(20). https://doi.org/10.21037/jmai-23-71

Week 2. Core Machine Learning Concepts for Clinicians and Translational Scientists

Medicine / Research / Ethics
Understanding why predictive models trained on historical electronic health record data may fail when applied to new populations, workflows, or care pathways.

Radiology
Interpreting performance metrics for imaging classifiers and recognizing why high accuracy does not guarantee reliability across scanners, protocols, or institutions.

- McBee, M. P., Awan, O. A., Colucci, A. T., Kansagra, A. P., Tridandapani, S., & Auffermann, W. F. (2018). Deep learning in radiology. *Academic Radiology*, 25(11), 1472–1480. https://doi.org/10.1016/j.acra.2018.03.006

Anesthesiology
Evaluating perioperative risk prediction models and assessing generalizability across case mix, acuity, and institutional context.

- Ocampo Osorio, F., Alzate-Ricaurte, S., Mejia Vallecilla, T. E., & Cruz-Suarez, G. A. (2024). The anesthesiologist's guide to critically assessing machine learning research: a narrative review. *BMC anesthesiology*, 24(1), 452. https://doi.org/10.1186/s12871-024-02840-y

Neurology
Critically appraising machine learning models for diagnosis or prognosis, such as progression from mild cognitive impairment to dementia, and recognizing risks of overfitting.

- Smith, C. M., Weathers, A. L., & Lewis, S. L. (2023). An overview of clinical machine learning applications in neurology. *Journal of the Neurological Sciences*, 455, 122799. https://doi.org/10.1016/j.jns.2023.122799

Ophthalmology
Understanding how supervised learning underpins retinal image classifiers and how dataset composition influences fairness and clinical performance.

- Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., Tan, G. S. W., Schmetterer, L., Keane, P. A., & Wong, T. Y. (2019). Artificial intelligence and deep learning in ophthalmology. *The British Journal of Ophthalmology*, 103(2), 167–175. https://doi.org/10.1136/bjophthalmol-2018-313173

Pathology
Evaluating machine learning models for histopathology or molecular classification, with attention to training data composition, slide preparation variability, and generalizability across labs.

- Serag, A., Ion-Margineanu, A., Qureshi, H., McMillan, R., Saint Martin, M. J., Diamond, J., O'Reilly, P., & Hamilton, P. (2019). Translational AI and Deep Learning in Diagnostic Pathology. *Frontiers in Medicine*, 6, 185. https://doi.org/10.3389/fmed.2019.00185

Psychiatry

Understanding how supervised learning underlies mental health chatbots and risk prediction models, and why performance can vary widely across populations, diagnoses, and cultural contexts. Psychiatrists examine how training data, labeling practices, and outcome definitions affect model behavior in depression, anxiety, substance use, and serious mental illness.

- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, 14, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Week 3. Interpretability, Temporal Data, and Real-World Signals

Medicine / Research / Ethics
Balancing the desire for explainable AI with the risk of misleading explanations in high-stakes clinical decision-making.

Radiology

Interpreting saliency maps or heatmaps in imaging AI and distinguishing clinically meaningful explanations from post hoc visual artifacts.

- Champendal, M., Müller, H., Prior, J. O., & Dos Reis, C. S. (2023). A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging. *European Journal of Radiology*, 169, 111159. https://doi.org/10.1016/j.ejrad.2023.111159

Anesthesiology

Working with time-series physiologic data from intraoperative monitoring and recognizing limits of explanation during rapid physiologic change.

- Ocampo Osorio, F., Alzate-Ricaurte, S., Mejia Vallecilla, T. E., & Cruz-Suarez, G. A. (2024). The anesthesiologist's guide to critically assessing machine learning research: a narrative review. *BMC Anesthesiology*, 24(1), 452. https://doi.org/10.1186/s12871-024-02840-y
- Boer, C., Touw, H. R., & Loer, S. A. (2018). Postanesthesia care by remote monitoring of vital signs in surgical wards. *Current Opinion in Anesthesiology*, 31(6), 716–722. https://doi.org/10.1097/ACO.0000000000000650

Neurology

Interpreting longitudinal signals from wearables, speech, or driving behavior and understanding how temporal context shapes inference.

- Colautti, L., Casella, M., Robba, M., Marocco, D., Ponticorvo, M., Iannello, P., Antonietti, A., Marra, C., & for the CPP Integrated Parkinson's Database. (2025). Predicting Cognitive Decline in Parkinson's Disease Using Artificial Neural Networks: An Explainable AI Approach. *Brain Sciences*, 15(8), 782. https://doi.org/10.3390/brainsci15080782
- Daniore, P., Nittas, V., Haag, C., Bernard, J., Gonzenbach, R., & von Wyl, V. (2024). From wearable sensor data to digital biomarker development: ten lessons learned and a framework proposal. *NPJ Digital Medicine*, 7(1), 161. https://doi.org/10.1038/s41746-024-01151-3

Ophthalmology

Evaluating explainability in retinal imaging AI and assessing whether highlighted features align with known anatomy and pathology.

- Singh, A., Jothi Balaji, J., Rasheed, M. A., Jayakumar, V., Raman, R., & Lakshminarayanan, V. (2021). Evaluation of Explainable Deep Learning Methods for Ophthalmic Diagnosis. *Clinical Ophthalmology* (Auckland, N.Z.), 15, 2573–2581. https://doi.org/10.2147/OPTH.S312236

Pathology

Assessing interpretability of AI-generated heatmaps or feature attributions on whole-slide images and determining whether highlighted regions correspond to known histopathologic features.

- Klauschen, F., Dippel, J., Keyl, P., Jurmeister, P., Bockmayr, M., Mock, A., Buchstab, O., Alber, M., Ruff, L., Montavon, G., & Müller, K. R. (2024). Toward Explainable Artificial Intelligence for Precision Pathology. *Annual Review of Pathology*, 19, 541–570. https://doi.org/10.1146/annurev-pathmechdis-051222-113147

Psychiatry

Interpreting models that use longitudinal behavioral and linguistic signals, such as speech patterns, text interactions, sleep, activity, or social engagement. This includes examining explainability challenges in chatbot outputs and emotional AI and understanding how apparent "empathy" or responsiveness may mask uncertainty, bias, or spurious associations.

- Joyce, D. W., Kormilitzin, A., Smith, K. A., & Cipriani, A. (2023). Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *NPJ Digital Medicine*, 6(1), 6. https://doi.org/10.1038/s41746-023-00751-9

- Welch, V., Wy, T. J., Ligezka, A., Hassett, L. C., Croarkin, P. E., Athreya, A. P., & Romanowicz, M. (2022). Use of Mobile and Wearable Artificial Intelligence in Child and Adolescent Psychiatry: Scoping Review. *Journal of Medical Internet Research*, 24(3), e33560. https://doi.org/10.2196/33560

Week 4. Disease Progression, Digital Twins, and Causal Reasoning

Medicine / Research / Ethics
Distinguishing prediction from causation when using AI to simulate treatment effects or disease trajectories.

Radiology
Understanding how longitudinal imaging contributes to disease progression modeling and emerging digital twin frameworks.

- Pesapane, F., Rotili, A., Penco, S., Nicosia, L., & Cassano, E. (2022). Digital Twins in Radiology. *Journal of Clinical Medicine*, 11(21), 6553. https://doi.org/10.3390/jcm11216553

Anesthesiology
Considering perioperative digital twins for risk simulation while recognizing causal limits inherent in observational data.

- Zhong, X., Babaie Sarijaloo, F., Prakash, A., Park, J., Huang, C., Barwise, A., … Dong, Y. (2022). A multidisciplinary approach to the development of digital twin models of critical care delivery in intensive care units. *International Journal of Production Research*, 60(13), 4197–4213. https://doi.org/10.1080/00207543.2021.2022235

Neurology
Applying progression models and digital twins to neurodegenerative disease, where timing, heterogeneity, and uncertainty are central.

- Alam, N., Basilico, J., Bertolini, D., Chetty, S. C., D'Angelo, H., Douglas, R., Fisher, C. K., Fuller, F., Gomes, M., Gupta, R., Lang, A., Loukianov, A., Mak-McCully, R., Murray, C., Pham, H., Qiao, S., Ryapolova-Webb, E., Smith, A., Theoharatos, D., … Walsh, J. R. (2024). Digital twin generators for disease modeling. *arXiv*. https://doi.org/10.48550/arXiv.2405.01488
- Dang, J., Lal, A., Flurin, L., James, A., Gajic, O., & Rabinstein, A. A. (2021). Predictive modeling in neurocritical care using causal artificial intelligence. *World Journal of Critical Care Medicine*, 10(4), 112–119. https://doi.org/10.5492/wjccm.v10.i4.112

Ophthalmology
Modeling disease progression in glaucoma or macular degeneration while separating correlation from causal inference.

- Iliuţă, M.-E., Moisescu, M.-A., Caramihai, S.-I., Cernian, A., Pop, E., Chiş, D.-I., & Mitulescu, T.-C. (2024). Digital Twin Models for Personalised and Predictive Medicine in Ophthalmology. *Technologies*, 12(4), 55. https://doi.org/10.3390/technologies12040055

Pathology
Integrating longitudinal tissue, molecular, or biomarker data into disease progression models while recognizing limits of causal inference from observational specimens.

- Pérez-Benito, Á., García-Aznar, J. M., Gómez-Benito, M. J., & Pérez, M. Á. (2024). Patient-specific prostate tumour growth simulation: a first step towards the digital twin. *Frontiers in Physiology*, 15, 1421591. https://doi.org/10.3389/fphys.2024.1421591

Psychiatry
Exploring how longitudinal symptom data from digital mental health tools could contribute to models of illness course, relapse risk, or recovery trajectories, while distinguishing prediction from causal inference. Discussion emphasizes the limits of using chatbot interactions or emotional robot engagement as proxies for underlying psychopathology or treatment response.

- Spitzer, M., Dattner, I., & Zilcha-Mano, S. (2023). Digital twins and the future of precision mental health. *Frontiers in Psychiatry*, 14, 1082598. https://doi.org/10.3389/fpsyt.2023.1082598

Week 5. Ethics, Governance, Human Factors, and Regulatory Science

Medicine / Research / Ethics
Addressing fairness, accountability, and governance when AI influences clinical judgment but does not replace it.

Radiology
Understanding regulatory expectations for imaging AI, including validation, monitoring for performance drift, and workflow integration.

- Ho, C. W. L., Soon, D., Caals, K., & Kapur, J. (2019). Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clinical Radiology*, 74(5), 329–337. https://doi.org/10.1016/j.crad.2019.02.005

Anesthesiology
Managing automation risk in safety-critical environments where overreliance can have immediate clinical consequences.

- Cascella, M., Tracey, M. C., Petrucci, E., & Bignami, E. G. (2023). Exploring Artificial Intelligence in Anesthesia: A Primer on Ethics, and Clinical Applications. *Surgeries*, 4(2), 264-274. https://doi.org/10.3390/surgeries4020027

Neurology
Navigating ethical challenges in AI-supported diagnosis and prognosis, particularly in neurodegenerative disease.

- Chiang, S., Picard, R. W., Chiong, W., Moss, R., Worrell, G. A., Rao, V. R., & Goldenholz, D. M. (2021). Guidelines for Conducting Ethical Artificial Intelligence Research in Neurology: A Systematic Approach for Clinicians and Researchers. *Neurology*, 97(13), 632–640. https://doi.org/10.1212/WNL.0000000000012570

Ophthalmology
Ensuring equitable deployment of screening AI across diverse populations and practice settings.

- Abdullah, Y. I., Schuman, J. S., Shabsigh, R., Caplan, A., & Al-Aswad, L. A. (2021). Ethics of Artificial Intelligence in Medicine and Ophthalmology. *Asia-Pacific Journal of Ophthalmology* (Philadelphia, Pa.), 10(3), 289–298. https://doi.org/10.1097/APO.0000000000000397

Pathology
Addressing governance, validation, and liability issues for AI-assisted diagnostics, including responsibility for final interpretation and documentation.

- Chauhan, C., & Gullapalli, R. R. (2021). Ethics of AI in Pathology: Current Paradigms and Emerging Issues. *The American Journal of Pathology*, 191(10), 1673–1683. https://doi.org/10.1016/j.ajpath.2021.06.011

Psychiatry

Addressing ethical concerns unique to psychiatric AI, including vulnerability, consent, privacy, emotional reliance, and the risk of anthropomorphism in chatbots and social or emotional robots. This includes critical discussion of governance, liability, and boundary-setting for tools like Woebot, particularly around crisis management, escalation, and clinician oversight.

- Putica, A., Khanna, R., Bosl, W., Saraf, S., & Edgcomb, J. (2025). Ethical decision-making for AI in mental health: the Integrated Ethical Approach for Computational Psychiatry (IEACP) framework. *Psychological Medicine*, 55, e213. https://doi.org/10.1017/S0033291725101311

Week 6. Translational Integration: From Models to Learning Health Systems

Medicine / Research / Ethics
Embedding AI within Learning Health Systems so models are continuously evaluated, updated, and improved over time.

Radiology
Using real-world performance monitoring to recalibrate imaging AI across sites, technologies, and patient populations.

- Park, W. Y., Sippel Schmidt, T., Salvador, G., O'Donnell, K., Genereaux, B., Jeon, K., You, S. C., Dewey, B. E., Nagy, P., & Alzheimer's Disease Neuroimaging Initiative (2025). Breaking data silos: incorporating the DICOM imaging standard into the OMOP CDM to enable multimodal research. *Journal of the American Medical Informatics Association.* JAMIA, 32(10), 1533–1541. https://doi.org/10.1093/jamia/ocaf091

Anesthesiology

Incorporating perioperative data into continuous learning loops to refine risk models and optimize workflows.

- Giri, R., Firdhos, S. H., & Vida, T. A. (2025). Artificial Intelligence in Anesthesia: Enhancing Precision, Safety, and Global Access Through Data-Driven Systems. *Journal of Clinical Medicine*, 14(19), 6900. https://doi.org/10.3390/jcm14196900

Neurology

Linking clinical, imaging, and digital biomarker data to support longitudinal learning and personalized care.

- Poldrack, R. A., Markiewicz, C. J., Appelhoff, S., Ashar, Y. K., Auer, T., Baillet, S., Bansal, S., Beltrachini, L., Benar, C. G., Bertazzoli, G., Bhogawar, S., Blair, R. W., Bortoletto, M., Boudreau, M., Brooks, T. L., Calhoun, V. D., Castelli, F. M., Clement, P., Cohen, A. L., Cohen-Adad, J., … Gorgolewski, K. J. (2024). The past, present, and future of the brain imaging data structure (BIDS). *Imaging Neuroscience* (Cambridge, Mass.), 2, 1–19. https://doi.org/10.1162/imag_a_00103

Ophthalmology

Scaling AI-enabled screening programs while maintaining data quality, governance, and clinical oversight.

- Colman, B. D., Zhu, Z., Qi, Z., & van der Walt, A. (2024). From real world data to real world evidence to improve outcomes in neuro-ophthalmology. *Eye* (London, England), 38(12), 2448–2456. https://doi.org/10.1038/s41433-024-03160-8

Pathology

Integrating pathology data, including digital slides and molecular results, into learning systems that support quality assurance, continuous validation, and translational research.

- Clunie D. A. (2021). DICOM Format and Protocol Standardization—A Core Requirement for Digital Pathology Success. *Toxicologic Pathology*, 49(4), 738–749. https://doi.org/10.1177/0192623320965893

Psychiatry

Integrating digital mental health tools into Learning Health Systems in ways that support monitoring, evaluation, and continuous improvement without displacing human care. Examples include using aggregated, de-identified chatbot interaction data to inform population-level insights while preserving patient trust, clinician responsibility, and ethical safeguards.

- Ranallo, P., & Tenenbaum, J. (2021). Precision Medicine and a Learning Health System for Mental Health (pp. 1–30). https://doi.org/10.1007/978-3-030-70558-9_1

Appendix B

Surgery (General & Subspecialties): Relevance by Week

Week 1. AI and Machine Learning at UNMC: Foundations, Resources, and Prompt Engineering

Using prompt engineering to summarize operative notes, synthesize perioperative guidelines, or draft research protocols, while maintaining surgeon responsibility for decisions and documentation.
Subspecialty examples

Urology

Structuring operative summaries for robot-assisted prostatectomy, including lymph node dissection details and perioperative complications.

- Shah, M., Naik, N., Somani, B. K., & Hameed, B. M. Z. (2020). Artificial intelligence (AI) in urology-Current use and future directions: An iTRUE study. *Turkish Journal of Urology*, 46(Supp. 1), S27–S39. https://doi.org/10.5152/tud.2020.20117

Gynecology

Summarizing minimally invasive hysterectomy or endometriosis cases, integrating pathology, imaging, and operative findings.

- Drukker, L., Noble, J. A., & Papageorghiou, A. T. (2020). Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound in Obstetrics & Gynecology*, 56(4), 498–505. https://doi.org/10.1002/uog.22122

Cardiothoracic Surgery

Organizing complex operative courses for CABG or valve surgery, including cardiopulmonary bypass times and postoperative trajectories.

- Mumtaz, H., Saqib, M., Ansar, F., Zargar, D., Hameed, M., Hasan, M., & Muskan, P. (2022). The future of cardiothoracic surgery in artificial intelligence. *Annals of Medicine and Surgery*, 80, 104251. https://doi.org/10.1016/j.amsu.2022.104251

Neurosurgery

Drafting structured summaries of tumor resections or spine surgeries, incorporating intraoperative monitoring and imaging findings.

- Mofatteh M. (2021). Neurosurgery and artificial intelligence. *AIMS Neuroscience*, 8(4), 477–495. https://doi.org/10.3934/Neuroscience.2021025

Week 2. Core Machine Learning Concepts for Clinicians and Translational Scientists

Understanding why surgical risk prediction models may perform well in one population but fail in others due to case mix, surgeon technique, robotic versus open approaches, or institutional workflows.
Subspecialty examples

Urology

Comparing complication and continence outcomes across open, laparoscopic, and robot-assisted prostatectomy cohorts.

- Welk, B., Zhong, T., Myers, J., Stoffel, J., Elliot, S., Lenherr, S. M., & Lizotte, D. (2024). Identifying Bladder Phenotypes After Spinal Cord Injury With Unsupervised Machine Learning: A New Way to Examine Urinary Symptoms and Quality of Life. *The Journal of Urology*, 212(1), 114–123. https://doi.org/10.1097/JU.0000000000003984

Gynecology

Evaluating prediction models for conversion to open surgery in complex minimally invasive cases.

- Arezzo, F., Cormio, G., La Forgia, D., Santarsiero, C. M., Mongelli, M., Lombardi, C., Cazzato, G., Cicinelli, E., & Loizzi, V. (2022). A machine learning approach applied to gynecological ultrasound to predict progression-free survival in ovarian cancer patients. *Archives of Gynecology and Obstetrics*, 306(6), 2143–2154. https://doi.org/10.1007/s00404-022-06578-1

Cardiothoracic Surgery

Assessing why mortality or length-of-stay models differ across centers with varying operative volumes and perfusion practices.

- Wang, Z., Zhe, S., Zimmerman, J., Morrisey, C., Tonna, J. E., Sharma, V., & Metcalf, R. A. (2022). Development and validation of a machine learning method to predict intraoperative red blood cell transfusions in cardiothoracic surgery. *Scientific Reports*, 12(1), 1355. https://doi.org/10.1038/s41598-022-05445-y

Neurosurgery

Understanding heterogeneity in readmission or complication prediction across cranial versus spinal procedures.

- Schilling, A. T., Shah, P. P., Feghali, J., Jimenez, A. E., & Azad, T. D. (2022). A Brief History of Machine Learning in Neurosurgery. Acta Neurochirurgica. Supplement, 134, 245–250. https://doi.org/10.1007/978-3-030-85292-4_27

Week 3. Interpretability, Temporal Data, and Real-World Signals

Interpreting AI models that use time-series perioperative data such as vitals, labs, intraoperative events, and postoperative recovery trajectories.
Subspecialty examples

Urology

Interpreting time-series data linking console time, instrument motion, and postoperative urinary outcomes.

- Syversen, A., Dosis, A., Jayne, D., & Zhang, Z. (2024). Wearable Sensors as a Preoperative Assessment Tool: A Review. Sensors, 24(2), 482. https://doi.org/10.3390/s24020482

Gynecology

Examining intraoperative workflow phases and postoperative pain or recovery patterns following minimally invasive surgery.

- Syversen, A., Dosis, A., Jayne, D., & Zhang, Z. (2024). Wearable Sensors as a Preoperative Assessment Tool: A Review. Sensors, 24(2), 482. https://doi.org/10.3390/s24020482
- Kodipalli, A., Devi, V. S., Guruvare, S., & Ismail, T. (2025). Explainable AI-based feature importance analysis for ovarian cancer classification with ensemble methods. *Frontiers in Public Health*, 13, 1479095. https://doi.org/10.3389/fpubh.2025.1479095

Cardiothoracic Surgery

Analyzing temporal relationships between intraoperative hemodynamics, bypass duration, and postoperative organ dysfunction.

- Syversen, A., Dosis, A., Jayne, D., & Zhang, Z. (2024). Wearable Sensors as a Preoperative Assessment Tool: A Review. Sensors, 24(2), 482. https://doi.org/10.3390/s24020482

Neurosurgery

Interpreting multimodal data combining intraoperative neuromonitoring, imaging, and postoperative neurologic recovery.

- Syversen, A., Dosis, A., Jayne, D., & Zhang, Z. (2024). Wearable Sensors as a Preoperative Assessment Tool: A Review. Sensors, 24(2), 482. https://doi.org/10.3390/s24020482

Week 4. Disease Progression, Digital Twins, and Causal Reasoning

Applying disease progression models and digital twin concepts to simulate surgical outcomes while distinguishing prediction from causal inference.
Subspecialty examples

Urology

Simulating cancer progression with or without prostatectomy and evaluating functional tradeoffs such as continence or sexual function.

- Görtz, M., Brandl, C., Nitschke, A., Riediger, A., Stromer, D., Byczkowski, M., Heuveline, V., & Weidemüller, M. (2026). Digital twins for personalized treatment in uro-oncology in the era of artificial intelligence. *Nature Reviews Urology*, 23(1), 29–39. https://doi.org/10.1038/s41585-025-01096-6

Gynecology

Modeling symptom progression in endometriosis with surgical versus medical management.

- Qi, W., Wang, Y., Wang, Y., Huang, S., Li, C., Jin, H., Zuo, J., Cui, X., Wei, Z., Guo, Q., & Hu, J. (2025). Prediction of postpartum depression in women: development and validation of multiple machine learning models. *Journal of Translational Medicine*, 23(1), 291. https://doi.org/10.1186/s12967-025-06289-6

Cardiothoracic Surgery

Simulating valve disease or coronary disease progression with surgical intervention compared with medical therapy.

- Lippert, M., Dumont, K. A., Birkeland, S., Nainamalai, V., Solvin, H., Suther, K. R., Bendz, B., Elle, O. J., & Brun, H. (2024). Cardiac anatomic digital twins: findings from a single national centre. *European Heart Journal - Digital Health*, 5(6), 725–734. https://doi.org/10.1093/ehjdh/ztae070

Neurosurgery

Modeling tumor growth, spinal degeneration, or hydrocephalus trajectories with and without operative intervention.

- Chumnanvej, S., Chumnanvej, S., & Tripathi, S. (2024). Assessing the benefits of digital twins in neurosurgery: a systematic review. *Neurosurgical Review*, 47(1), 52. https://doi.org/10.1007/s10143-023-02260-5

Week 5. Ethics, Governance, Human Factors, and Regulatory Science

Addressing ethical and governance issues when AI informs surgical decision-making or robotic assistance.
Subspecialty examples

Urology

Managing automation bias when robotic metrics or AI-based guidance suggest optimal dissection planes.

- Cacciamani, G. E., Chen, A., Gill, I. S., & Hung, A. J. (2024). Artificial intelligence and urology: ethical considerations for urologists and patients. *Nature Reviews Urology*, 21(1), 50–59. https://doi.org/10.1038/s41585-023-00796-1

Gynecology

Ensuring equitable access and appropriate patient selection for minimally invasive or robot-assisted procedures.

- Sengupta, P., Dutta, S., Bagchi, S., Hegde, P., Sebastian, S., Taylor, D. C., & Henkel, R. (2025). Strategic Guidelines to Integrate Artificial Intelligence in Obstetrics and Gynecology: Best Practices and Ethical Considerations. *Reproductive sciences* (Thousand Oaks, Calif.), 32(7), 2244–2251. https://doi.org/10.1007/s43032-025-01903-w

Cardiothoracic Surgery

Clarifying accountability when AI-assisted risk stratification influences operative candidacy.

- Gross, M., Wang, H., & Scully, B. B. (2025). Ethics of Precision Medicine and Cardiothoracic Surgery. *Thoracic Surgery Clinics*, 35(4), 489–497. https://doi.org/10.1016/j.thorsurg.2025.07.002

Neurosurgery

Balancing AI-supported navigation or targeting tools with surgeon judgment in high-risk cranial procedures.

- Reyes, J. S., Lohia, V. N., Almeida, T., Niranjan, A., Lunsford, L. D., & Hadjipanayis, C. G. (2025). Artificial intelligence in neurosurgery: a systematic review of applications, model comparisons, and ethical implications. *Neurosurgical Review*, 48(1), 455. https://doi.org/10.1007/s10143-025-03597-9

<u>Week 6. Translational Integration: From Models to Learning Health Systems</u>

Embedding surgical AI tools into Learning Health Systems so outcomes, complications, and long-term recovery feed back into continuous improvement.

- Krapohl, G. L., Hemmila, M. R., Hendren, S., Bishop, K., Rogers, R., Rocker, C., Fasbinder, L., Englesbe, M. J., Vu, J. V., & Campbell, D. A., Jr (2020). Building, scaling, and sustaining a learning health system for surgical quality improvement: A toolkit. *Learning Health Systems*, 4(3), e10215. https://doi.org/10.1002/lrh2.10215

Urology

Registry-linked learning using functional outcomes and complications after robot-assisted prostatectomy.

- Lawlor, A., Beyer, K., Russell, B., Steinbeisser, C., Bjartell, A., De Meulder, B., Omar, M. I., Hulsen, T., Butler, J., N'Dow, J., Rivas, J. G., Gandaglia, G., Nicoletti, R., Sakalis, V., Smith, E. J., Maass, M., Zong, J., Fullwood, L., Abbott, T., Tafreshiha, A., … PIONEER Consortium (2025). PIONEER big data platform for prostate cancer: lessons for advancing future real-world evidence research. *Nature Reviews Urology*, 22(2), 116–124. https://doi.org/10.1038/s41585-024-00925-4

Gynecology

Continuous quality improvement for minimally invasive surgery using real-world data on recovery and readmissions.

- Katon, J. G., Rodriguez, A., Yano, E. M., Johnson, A. M., Frayne, S. M., Hamilton, A. B., Miller, L. J., Williams, K., Zephyrin, L., & Patton, E. W. (2023). Research Priorities to Support Women Veterans' Reproductive Health and Health Care Within a Learning Health Care System. *Women's Health Issues*, 33(3), 215–221. https://doi.org/10.1016/j.whi.2022.12.003

Cardiothoracic Surgery

Learning loops that integrate operative data, ICU outcomes, and long-term survival.

- Maddula, R., MacLeod, J., McLeish, T., Painter, S., Steward, A., Berman, G., Hamid, A., Abdelrahim, M., Whittle, J., & Brown, S. A. (2022). The role of digital health in the cardiovascular learning healthcare system. *Frontiers in Cardiovascular Medicine*, 9, 1008575. https://doi.org/10.3389/fcvm.2022.1008575

Neurosurgery

Real-world performance monitoring of image-guided and robotic-assisted systems tied to neurologic outcomes.

- Matsumoto, K., Nohara, Y., Wakata, Y., Yamashita, T., Kozuma, Y., Sugeta, R., Yamakawa, M., Yamauchi, F., Miyashita, E., Takezaki, T., Yamashiro, S., Nishi, T., Machida, J., Soejima, H., Kamouchi, M., & Nakashima, N. (2020). Impact of a learning health system on acute care and medical complications after intracerebral hemorrhage. *Learning Health Systems*, 5(2), e10223. https://doi.org/10.1002/lrh2.10223

Appendix C:

Mapping Key Clinician AI Principles to the Curriculum

Table. Mapping Key Clinician AI Principles to the AI in Clinical & Translational Medicine Curriculum

Infographic Principle	Primary Curriculum Week(s)	Where It Appears in the Course	Clinical & Translational Emphasis
1. Understand AI Basics	Week 1	AI definitions, scope, limitations, expectations	Prevents overclaiming and misuse
2. Build Data Literacy	Week 2	ML concepts, datasets, metrics, generalizability	Enables critical appraisal
3. Explore AI Tools	Week 1	UNMC AI ecosystem, vetted tools	Distinguishes clinical vs consumer AI
4. Learn Clinical Use Cases	Weeks 1–3	Specialty-specific workflow examples	Grounds AI in practice
5. Master Prompting Skills	Week 1	Structured prompting, bias control	Prompting as reasoning interface
6. Collaborate with Technologists	Week 6	LHS/LRS, team science	Clinicians as co-designers
7. Apply AI in Diagnosis	Weeks 2–3	Explainability, second-reader framing	Retains clinician judgment
8. Enhance Treatment Planning	Week 4	Progression models, digital twins	Avoids causal errors
9. Use AI for Documentation	Weeks 1, 3–5	NLP support, documentation risks	Reduces burden safely
10. Monitor Patient Outcomes	Weeks 3–6	Longitudinal signals, feedback loops	Supports continuous learning
11. Strengthen Data Security	Weeks 4–6	Governance, multimodal data	Builds translational trust
12. Address Bias & Ethics	Week 5	Fairness, automation bias	Ethics as system property
13. Continue Professional Development	Weeks 5–6	Learning cycles, adaptation	AI competency as ongoing
14. Evaluate Vendor Credibility	Week 5	Validation, regulatory science	Supports procurement decisions
15. Lead AI Transformation	Week 6	Clinician leadership in LHS/LRS	Positions clinicians as leaders

Appendix D.

Alignment of Capstone Projects With the 6-Week Course and Clinical Specialties

This table summarizes how optional capstone projects align with each week of the six-week course and accommodate a broad range of clinical specialties and practice settings. Capstones are concept-driven exercises emphasizing translational problem framing rather than technical implementation.

Week	Course Focus	Capstone Focus	Illustrative Capstone Activity	Specialty Examples (Appendices A & B)
1	Foundations, resources, prompt engineering	Problem framing	Define a clinical, research, or system problem; clarify decision context; use prompt engineering to refine questions or synthesize background	Internal medicine: longitudinal care summaries; surgery: perioperative risk framing; psychiatry: intake or note synthesis; radiology: second-reader workflows; primary care: continuity documentation
2	ML concepts, data quality, generalizability	Feasibility and data readiness	Identify data sources; assess bias and generalizability; draft a brief feasibility or data readiness note	Internal medicine: multi-site outcome prediction; surgery: cross-site complication data; pathology: dataset shift across labs; emergency medicine: data completeness and timing
3	Interpretability, temporal data, real-world signals	Trust and longitudinal reasoning	Define interpretability needs and assess how outputs are interpreted over time	Cardiology: time-dependent risk scores; psychiatry: explainability in relapse risk; anesthesiology: intraoperative trend interpretation; pulmonary or critical care: longitudinal physiologic monitoring
4	Disease progression, digital twins, causal reasoning	Actionability and causal assumptions	Distinguish prediction from intervention and articulate causal assumptions	Oncology: treatment sequencing; endocrinology: timing of diabetes interventions; rehabilitation: recovery trajectories; obstetrics: pregnancy risk trajectories
5	Ethics, governance, human factors, regulatory science	Risk and oversight	Identify ethical risks, governance needs, and human factors constraints	Primary care: equity and access; radiology: liability and workflow integration; psychiatry: automation bias; ophthalmology: screening thresholds and harms
6	Learning Health Systems and translational integration	Sustainability and learning	Describe outcome monitoring and feedback within a Learning Health System	Surgery: learning from postoperative outcomes; internal medicine: continuous outcome monitoring; population health: system-level improvement cycles; pediatrics: adaptive care pathways

Capstone outputs produced during the six-week course are intended to be reusable beyond the course and may inform grant concept development, pilot or feasibility studies, quality improvement initiatives, governance planning, or subsequent project-based training. Capstones may be expanded in the 48-week course to include formal study design, data pipelines, stakeholder engagement, and sustained evaluation without reframing the underlying problem.

Participants may optionally submit a brief structured artifact (e.g., a one-page AI readiness assessment for a clinical or research workflow, an annotated critique of a published AI study, or a short governance risk checklist). These artifacts are used solely for formative feedback and course improvement and are not graded for technical correctness or deployment readiness.

Appendix E.

Example of Curriculum Evaluative Questions

The following evaluation items are explicitly aligned with the Clinician AI Principles outlined in Appendix C. They are designed to assess foundational literacy, judgment, and responsible application rather than technical proficiency or independent deployment competence.

Participation and Critical Reasoning

Likert Scale (1 = Strongly Disagree … 5 = Strongly Agree)

"The course improved my understanding of core AI concepts, limitations, and appropriate expectations in clinical and translational contexts."
(Appendix C: Principles 1 – Understand AI Basics; 3 – Explore AI Tools)

"The course improved my ability to identify data quality, bias, and generalizability issues in AI-enabled clinical or research outputs."
(Principles 2 – Build Data Literacy; 12 – Address Bias & Ethics)

"I can explain why high model accuracy does not necessarily translate into clinical usefulness or safety."
(Principles 2 – Build Data Literacy; 8 – Enhance Treatment Planning)

"The course strengthened my ability to critique AI outputs and distinguish signal from noise in real-world clinical or translational data."
(Principles 2 – Build Data Literacy; 7 – Apply AI in Diagnosis; 10 – Monitor Patient Outcomes)

"The course increased my awareness of ethical, governance, and accountability considerations across the AI lifecycle."
(Principles 12 – Address Bias & Ethics; 14 – Evaluate Vendor Credibility; 15 – Lead AI Transformation)

Clinical Care Application

"I applied course concepts to a clinical workflow."
[Yes / No]

If yes: Briefly describe the application, including the clinical setting and patient group if relevant. Do not include any identifying information.

Likert Scale (1 = Strongly Disagree … 5 = Strongly Agree)

"The application supported clinical reasoning or decision-making without replacing clinician judgment."
(Principles 7 – Apply AI in Diagnosis; 8 – Enhance Treatment Planning)

"The application improved efficiency or workflow while maintaining patient safety and accountability."
(Principles 9 – Use AI for Documentation; 10 – Monitor Patient Outcomes)

"I was able to recognize and mitigate risks of automation bias or inappropriate reliance on AI outputs."
(Principle 12 – Address Bias & Ethics)

"The application improved communication or coordination within the care team."
(Principles 6 – Collaborate with Technologists; 10 – Monitor Patient Outcomes)

Research and Translational Application

"I applied course concepts to a research or translational activity (e.g., grant development, study design, data collection or analysis, governance planning, or research workflow)."
[Yes / No]

If yes: Briefly describe the application and its anticipated or observed effect. Do not include any identifying information.

Likert Scale (1 = Strongly Disagree … 5 = Strongly Agree)

“The application improved the rigor, feasibility, or interpretability of the research activity.”
(Principles 2 – Build Data Literacy; 8 – Enhance Treatment Planning)

“The application clarified assumptions, limitations, or generalizability relevant to translational use.”
(Principles 2 – Build Data Literacy; 10 – Monitor Patient Outcomes)

“The application improved collaboration with data science, informatics, or technical partners.”
(Principle 6 – Collaborate with Technologists)

“The application strengthened readiness for responsible AI use within a Learning Health System or Learning Research System.”
(Principles 13 – Continue Professional Development; 15 – Lead AI Transformation)

Appendix F. Glossary

This glossary defines key terms as they are used in *AI in Clinical and Translational Medicine*. Definitions are clinician-facing and conceptual, emphasizing judgment, interpretation, and system integration rather than technical implementation.

AI Lifecycle Management: Oversight of AI systems from development and validation through deployment, monitoring, updating, and retirement.

Alignment Artifacts: Model behaviors shaped by training and reinforcement processes that optimize for human preference or fluency rather than objective accuracy.

Artificial Intelligence (AI): Computational approaches that enable machines to perform tasks such as pattern recognition, prediction, language processing, and decision support. In this course, AI is treated as a sociotechnical system rather than a standalone technology.

Automation Bias: The tendency of humans to over-rely on automated system outputs, even when those outputs are incorrect or inappropriate.

Automation Risk: Potential harm arising from inappropriate reliance on automated systems, particularly in safety-critical contexts.

Bias (AI): Systematic distortion arising from data selection, labeling, model design, deployment context, or downstream use.

Capstone Project (Course): A short, concept-driven application exercise focused on problem framing, translational relevance, and feasibility without technical implementation.

Causal Reasoning: Analytical approaches that distinguish correlation from cause-and-effect relationships.

Clinical Accountability: The principle that clinicians and institutions retain responsibility for decisions and outcomes regardless of AI involvement.

Clinical Trial Management System (CTMS): Systems supporting operational management and oversight of clinical trials.

Counterfactuals: Hypothetical scenarios representing what might have occurred under different conditions or interventions.

Data Provenance: Documentation of data origin, collection methods, transformations, and contextual meaning.

DICOM: A standard for handling, storing, and transmitting medical imaging data and related information.

Digital Biomarkers: Quantifiable behavioral or physiologic measures derived from digital devices or sensors.

Digital Twin: An evolving computational representation of an individual patient or population integrating multimodal data to simulate trajectories or scenarios. Digital twins are treated as exploratory unless causally validated.

Disease Progression Models: Models that characterize changes in disease state over time rather than static prediction.

Electronic Data Capture (EDC): Systems used to collect and manage clinical research data.

Electronic Health Record (EHR): Digital systems used to store and manage patient health information from routine care.

External Validation: Evaluation of model performance using data from different populations, institutions, or settings.

FAIR Principles: Guidelines stating that data should be Findable, Accessible, Interoperable, and Reusable.

Generalizability: The extent to which a model performs reliably across populations, settings, workflows, and time periods beyond its training data.

Good Machine Learning Practice (GMLP): Regulatory principles supporting safe, effective, and trustworthy AI development and deployment.

Governance (AI): Institutional policies, structures, and oversight mechanisms guiding AI development, deployment, monitoring, and accountability.

Green Prompting: A concise, task-appropriate prompting approach that minimizes unnecessary computation, cost, and overgeneration while supporting responsible AI use.

Hallucinations (AI): Outputs that appear coherent but are factually incorrect, fabricated, or unsupported by evidence.

Human Factors: The study of how people interact with systems, technologies, and workflows.

Institutional Review Board (IRB): A committee that reviews and oversees research involving human participants.

Interpretability / Explainability: The degree to which a human can understand how an AI system produces an output. Explanations may mislead and require critical appraisal.

Internal Validation: Assessment of model performance using data from the same source or population on which it was developed.

Jailbreaking: Efforts to bypass built-in safety constraints or usage policies of AI systems.

Large Language Model (LLM): A machine learning model trained on large text corpora to generate, summarize, or transform language. LLMs are treated as probabilistic text generators, not sources of ground truth.

Learning Health System (LHS): A health system in which data from routine care are continuously analyzed and fed back to improve quality, safety, and outcomes.

Learning Research System (LRS): An extension of the LHS concept that integrates clinical research, real-world evidence, and translational science into continuous learning cycles.

Machine Learning (ML): A subset of AI in which models learn patterns from data to make predictions or classifications without explicit programming.

Model Drift (Performance Drift): Changes in model behavior over time due to shifts in populations, data distributions, clinical practice, or system context.

Model Monitoring: Ongoing surveillance of AI system performance, safety, and unintended consequences after deployment.

Model Performance Metrics: Quantitative measures such as accuracy, sensitivity, specificity, and AUROC used to evaluate models. Metrics do not equate directly to clinical usefulness.

OMOP Common Data Model (OMOP CDM): A standardized data structure enabling harmonized analysis of observational health data across institutions.

Overfitting: A failure mode in which a model performs well on training data but poorly on new or external data.

Prompt Engineering: The practice of structuring inputs to AI systems to improve clarity, relevance, transparency, and safety of outputs.

Prompt Hacking: Attempts to manipulate AI outputs through strategic phrasing of inputs rather than evidence quality or valid reasoning.

Real-World Data (RWD): Data collected outside controlled research environments, often characterized by noise, missingness, and heterogeneity.

Real-World Evidence (RWE): Clinical or population-level insights derived from analysis of real-world data.

Red-Teaming: Structured testing of AI systems to identify vulnerabilities, misuse scenarios, or failure modes.

Regulatory Science: Application of scientific methods to evaluate safety, effectiveness, reliability, and performance of medical products and health technologies, including AI systems. In this course, regulatory science includes both formally regulated medical devices and non-device AI applications that require institutional governance, validation, monitoring, and accountability.

Safety Guardrails: Technical, procedural, and organizational controls designed to limit harmful, unsafe, or inappropriate AI behavior.

Second-Reader Model: A deployment pattern in which AI outputs support clinician judgment without replacing it.

Second-Reader Workflow: Operational use of AI as an additional review layer within clinical workflows.

Sociotechnical Failure: Failure arising from interactions among technology, people, workflows, and organizational context rather than from algorithms alone.

Sociotechnical System: A system in which outcomes emerge from interactions among technology, people, workflows, organizational structures, and policy.

Supervised Learning: Machine learning in which models are trained on labeled data to predict predefined outcomes.

Synthetic Data: Artificially generated data designed to approximate real data distributions. Synthetic data can mitigate or amplify bias.

Temporal Data: Data collected over time, including longitudinal clinical records, physiologic monitoring, wearables, speech, sleep, or behavioral signals.

Unsupervised Learning: Machine learning approaches that identify structure or patterns in unlabeled data.

Workflow Integration: Incorporation of AI tools into existing clinical or research processes without increasing burden or risk.