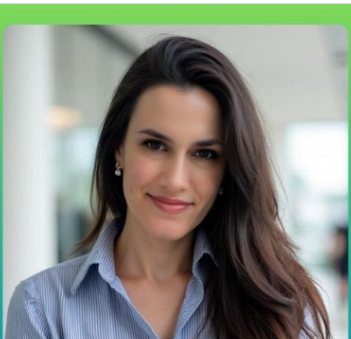
Building Foundations for AI Success: Data Infrastructure and Culture for Effective Implementation

January 29, 2026

Speakers

Jamie Fairclough, PhD, MPH, MSPharm
Director and Faculty, Engineering & Medicine
Dartmouth College

Laurah Turner, PhD
Associate Dean for Artificial Intelligence and Educational Informatics
Associate Professor of Biostatistics
Health Informatics and Data Sciences and Medical Education
University of Cincinnati College of Medicine

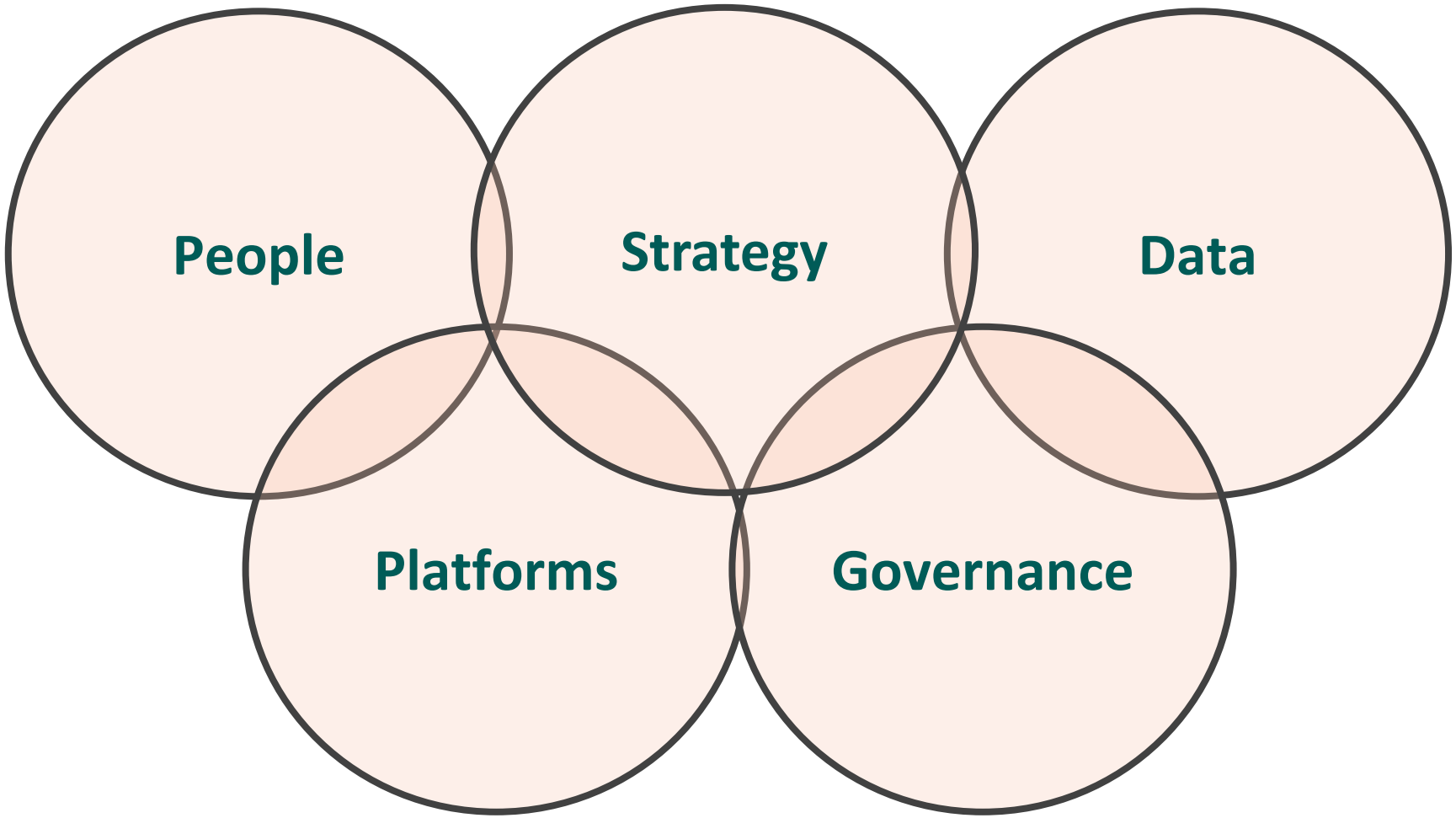


Victoria Yaneva, PhD
Director, Data Science and AI
NBME

Systems + Humans: Organizational Readiness for AI

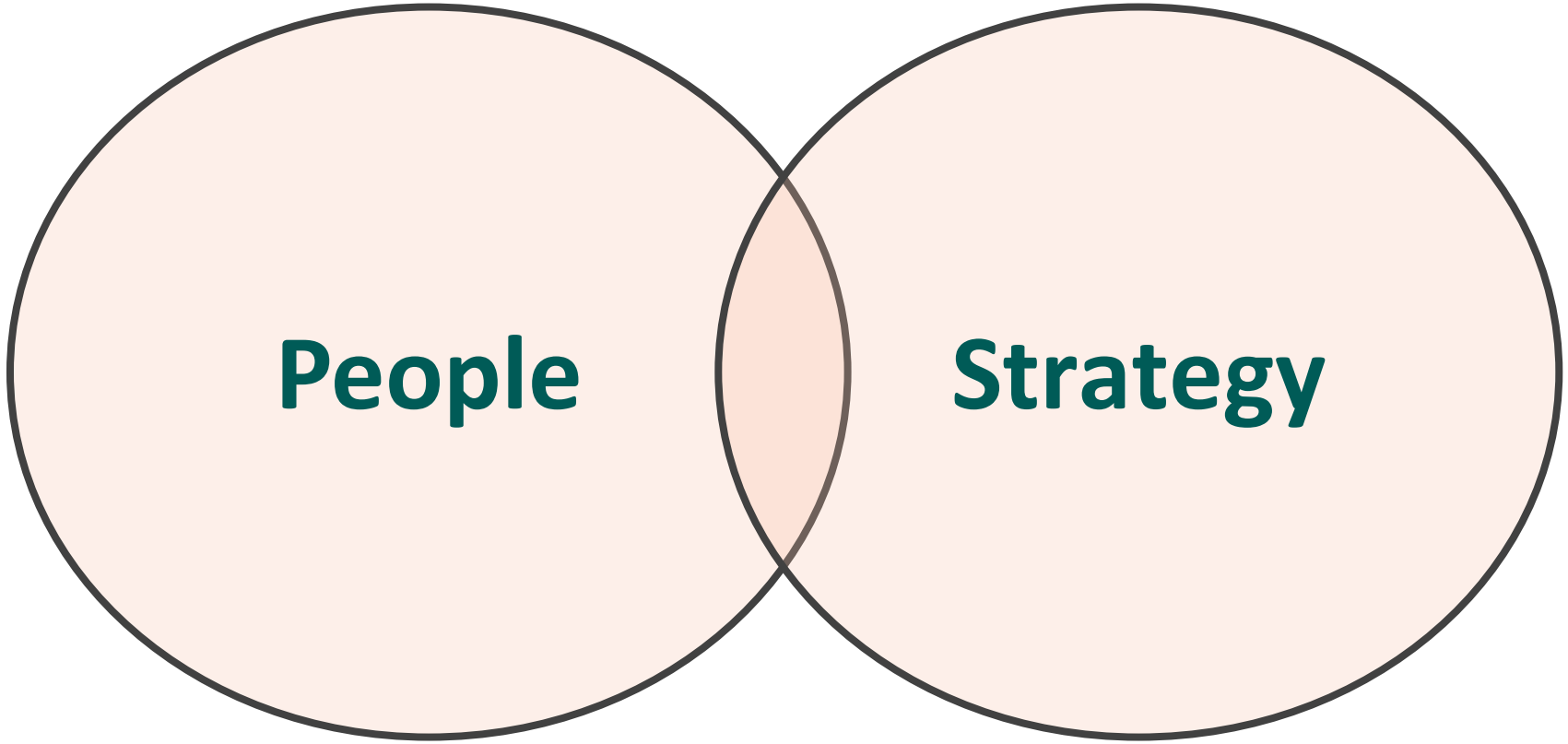
Victoria Yaneva, PhD
NBME

Organizational AI Readiness



NBME

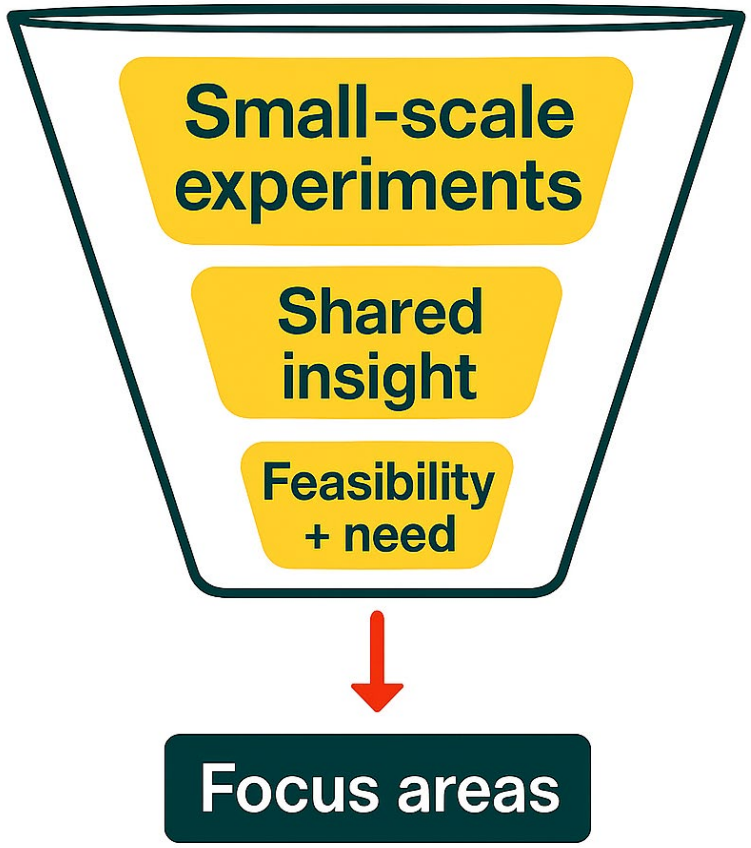
Organizational AI Readiness



AI at NBME

Find Strategic Focus

Common pitfall:
Letting technology dictate the solution

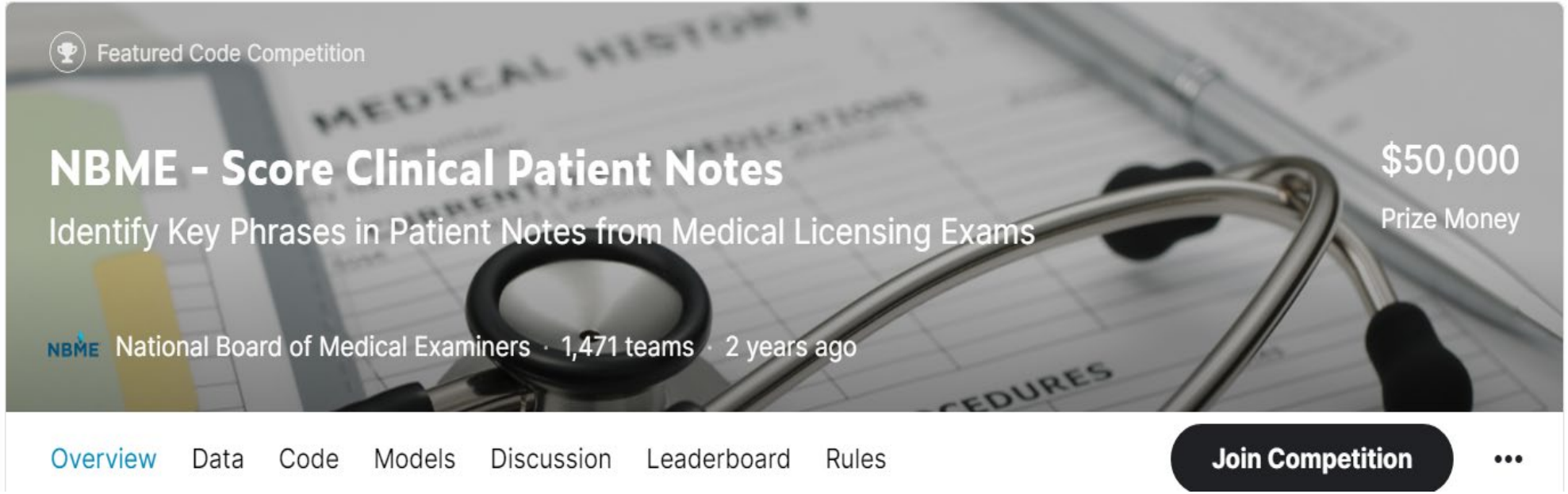


NBME

Create an Informed AI Culture

NBME

Build Trust through Transparency



Kaggle

BEA Competitions

NBME

Conclusion

Small-scale experimentation
helps find strategic focus

Education is key
for building an informed AI culture

Trust demands transparency

NBME

NBME®

Questions?

Conflicts & Disclaimers

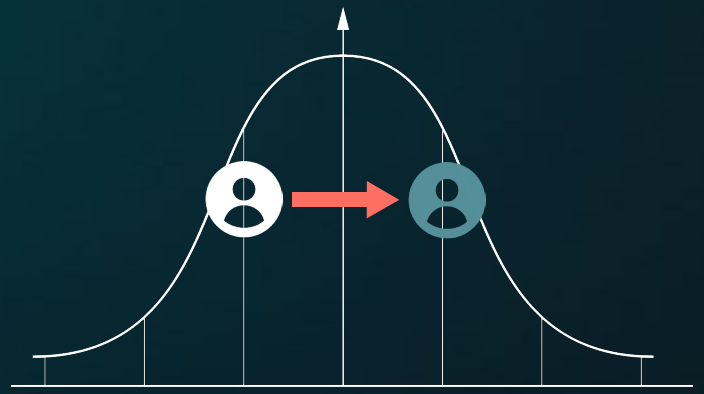
- Funding from the American Medical Association, American Board of Internal Medicine, APCCMPD, CHEST, and ATS Medical Education Research Award

- U.S. Provisional Application No. 63/620,383, Entitled: Fuzzy Logic Integrated with Ontologenic Neural Networks

Bloom's 2-Sigma Problem

In the 1980's, educational psychologist, Benjamin Bloom demonstrated one-on-one tutoring could improve performance by 2 standard deviations.

Transforming below-average learners to above-average. . .



. . . and average learners to exceptional.

Offering one-on-one experience to every learner has been historically impractical due to resource constraints, faculty availability, and financial barriers.

Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4-16.

Building Foundations for AI Success

A Four-Layer Framework for Robust Data Culture

🎯 Educational Impact and Improvement

- ACGME competency mapping
- Validity evidence
- Equity in outcomes measurement

✓ *MAESSCR case generation accuracy; no significant errors found*

👥 People, Process & Fairness

- Multi-stakeholder governance
- Student agency & transparency
- Continuous feedback mechanisms

✓ *Example: 602 M2 students completed MAESSCR encounters in Clinical Skills, Learner involvement on all build and red teams*

⚙️ Technical Infrastructure & Quality

- Model versioning & monitoring
- Data quality validation pipelines
- Security: RBAC, encryption, audit logs

✓ *MAESSCR logs all LLM interactions (GPT-4→Claude 3.5); transcript quality checks automated*

🛡️ Governance Ethics and Risk

- IRB + FERPA compliance
- Data lifecycle: retention (4yr), deletion/opt out rights
- Bias detection & risk mitigation protocols

✓ *Consent includes secondary use, equity audits, data deletion rights; incident response plan established*

⚡ Critical for Success

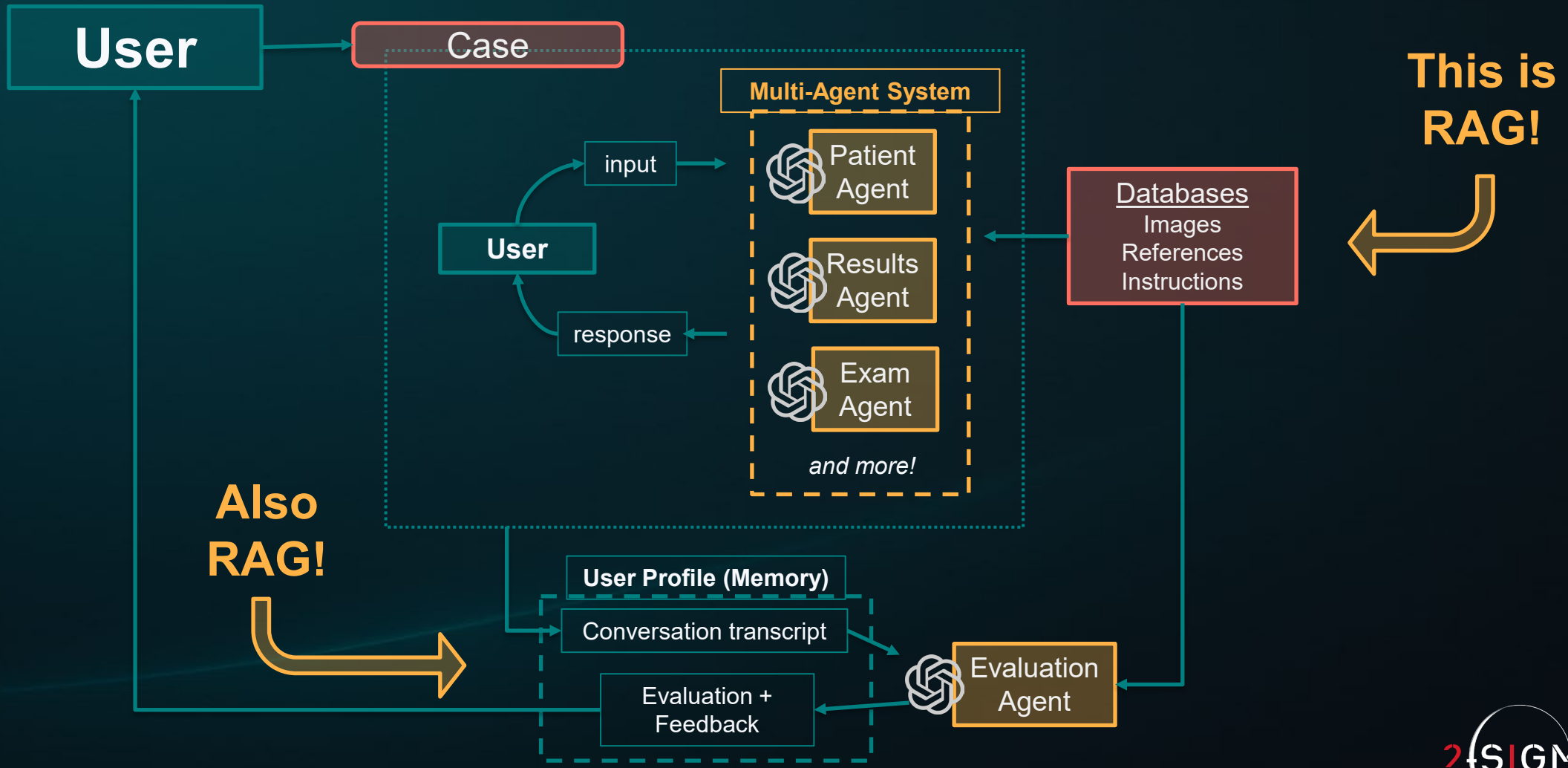
- ✦ Algorithmic fairness monitoring
- ✦ Model version control (MLOps)
- ✦ Data lifecycle management
- ✦ Transparency mechanisms for learners
- ✦ Risk assessment framework
- ✦ Continuous quality improvement cycles

Each layer depends on the one below it. Start with governance, not tools.

Framework informed by: NASEM AI Guidelines (2022) • AAMC AI in Medical Education (2024) • Kane Validity Framework • NIH Data Sharing Policy

MAESSCR

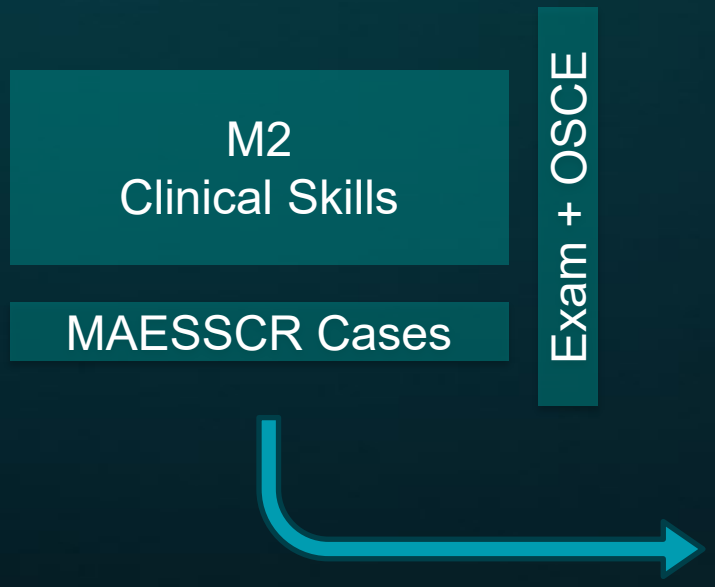
Multi-Agent Educational Scenario Simulator for Clinical Reasoning



Turner, Laurah, et al. "Harnessing the Generative Power of AI to Move Closer to Personalized Medical Education." *Academic Medicine* (2025): 10-1097.

2-SIGMA LABS

Any educational tool needs to be designed for the system it is meant to be used in.



Backend architecture allow for conversational transcripts.

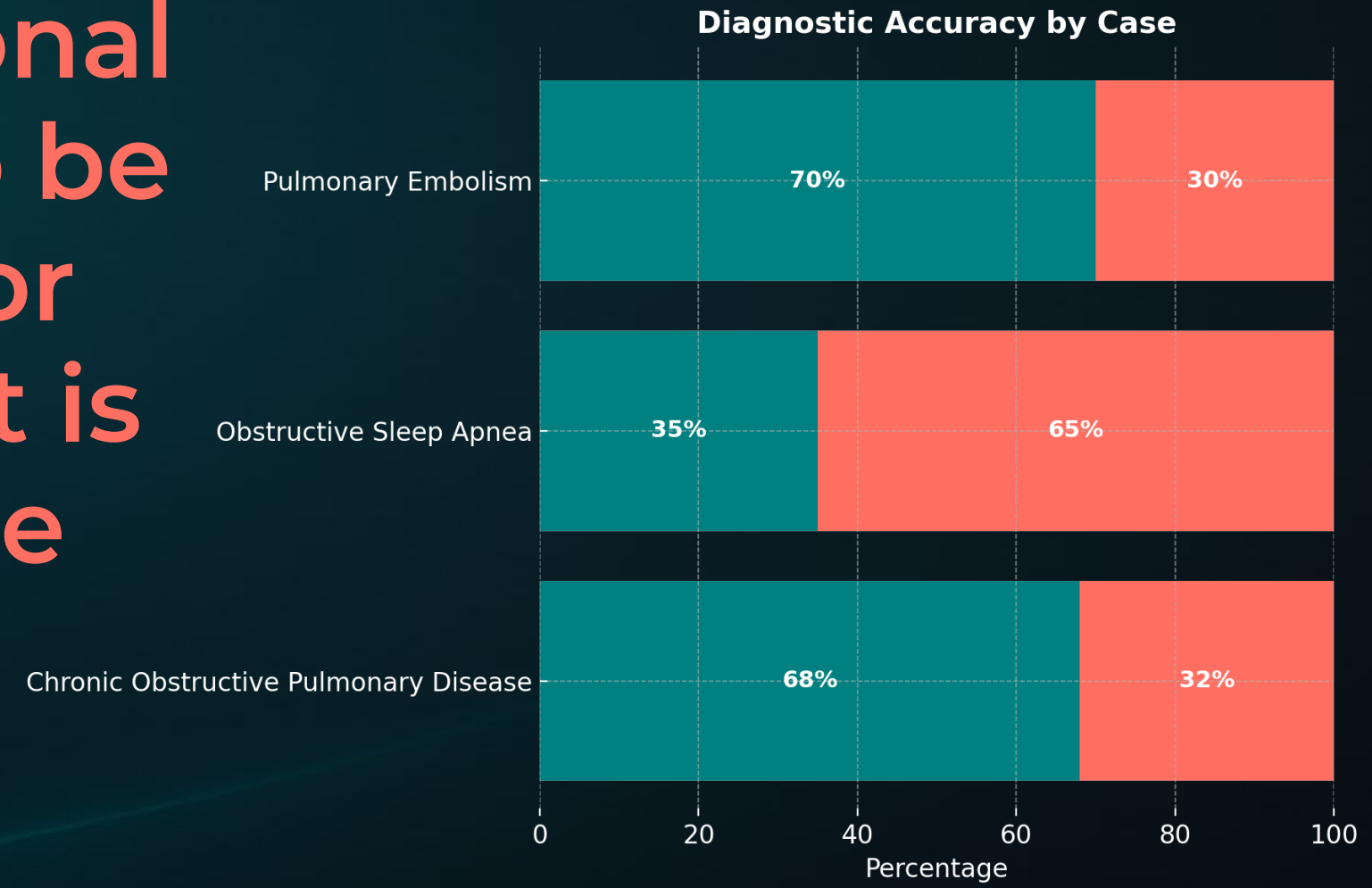
Student	What brings you in today Mr. Doe
AI	I've been having some chest pain and I'm short of breath.
Student	could you tell me more about that?
AI	The chest pain started about 12 hours ago and it's been getting worse. It hurts more when I breathe in and out.
Student	Could you describe the pain?
AI	It's sharp and gets worse when I breathe deeply. It's kind of centered on the right side of my chest.
Student	does the pain stay in that area of your chest?
AI	Yes, it stays right there in that area; it doesn't move around.
Student	any back pain?
AI	No, I don't have any back pain.

2-SIGMA LABS

Turner L, et al. Academic Medicine. August 2025. DOI: 10.1097/ACM.0000000000006185

602 transcripts from 2nd-year medical students in AI-simulated encounters

Any educational tool needs to be designed for the system it is meant to be used in.



602 Sessions

303 Correct (50%)

221 Incorrect (37%)

***Zhou C and Turner L**. Decoding Diagnostic Success: An AI-Powered Analysis of Student Behaviors in Virtual Pulmonary Encounters. Manuscript in progress.*

Well-designed tools can transform systems.

Pulm Cases	*Finished Sessions*	*Pulm Exam*	*%*	*Cardiac Exam*	*%*	*Oxygen Check*	*%*	*All 3*	*%*
Case 3	177	155	**87.6%**	138	**78.0%**	63	**35.6%**	49	**27.7%**
Case 5	172	159	**92.4%**	162	**94.2%**	27	**15.7%**	26	**15.1%**
Case 6	175	146	**83.4%**	123	**70.3%**	155	**88.6%**	103	**58.9%**
Overall	524	460	**87.8%**	423	**80.7%**	245	**46.8%**	178	**34.0%**

Methods:

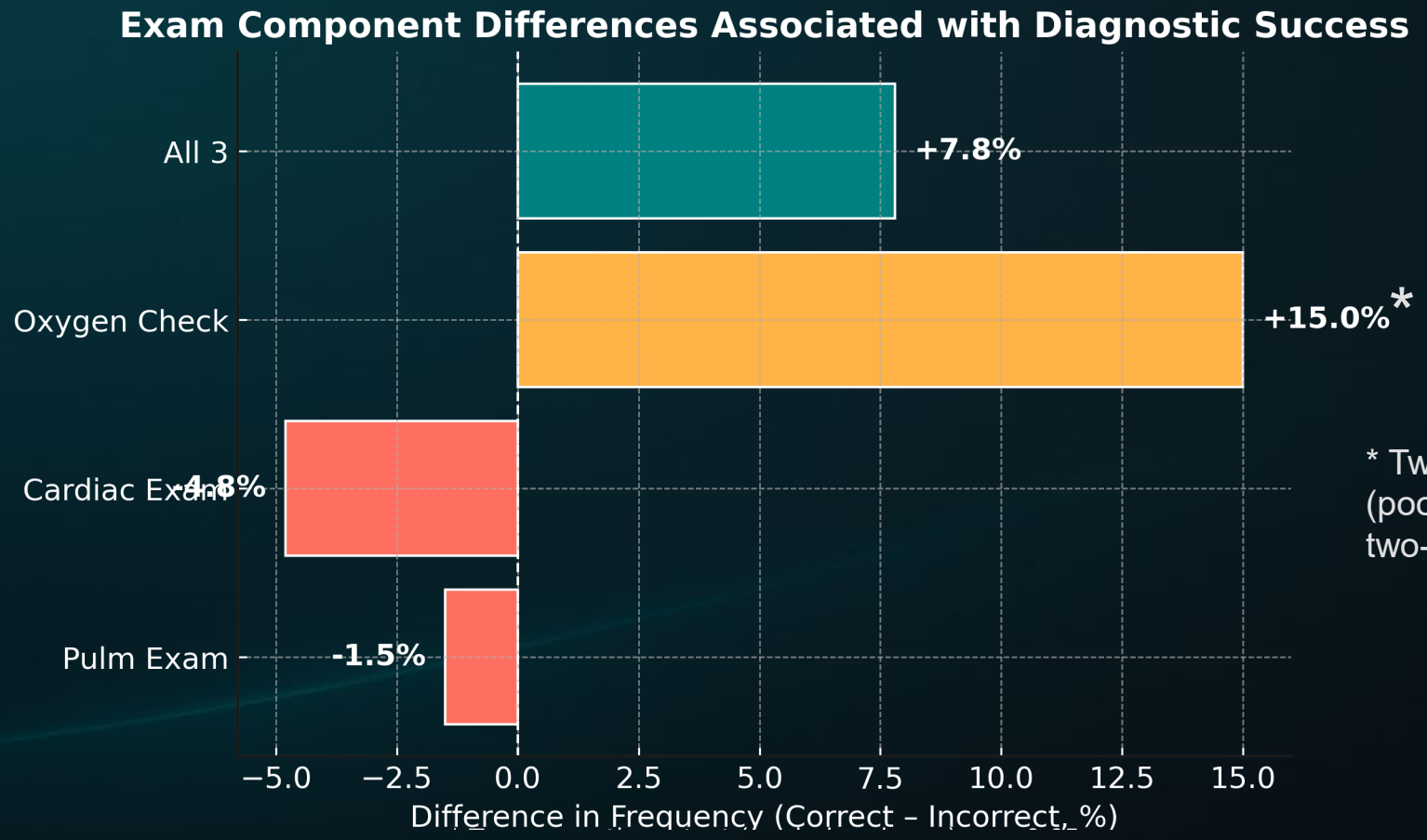
- Rule-based natural language processing (NLP) approach using a curated dictionary (23 components), each with standardized terms and comprehensive regex to capture lexical variation
- All student inputs underwent preprocessing (text normalization), then flagged component presence via regex matches

2-SIGMA LABS

***Zhou C and Turner L**. Decoding Diagnostic Success: An AI-Powered Analysis of Student Behaviors in Virtual Pulmonary Encounters. Manuscript in progress.*

Well-designed tools can transform systems.

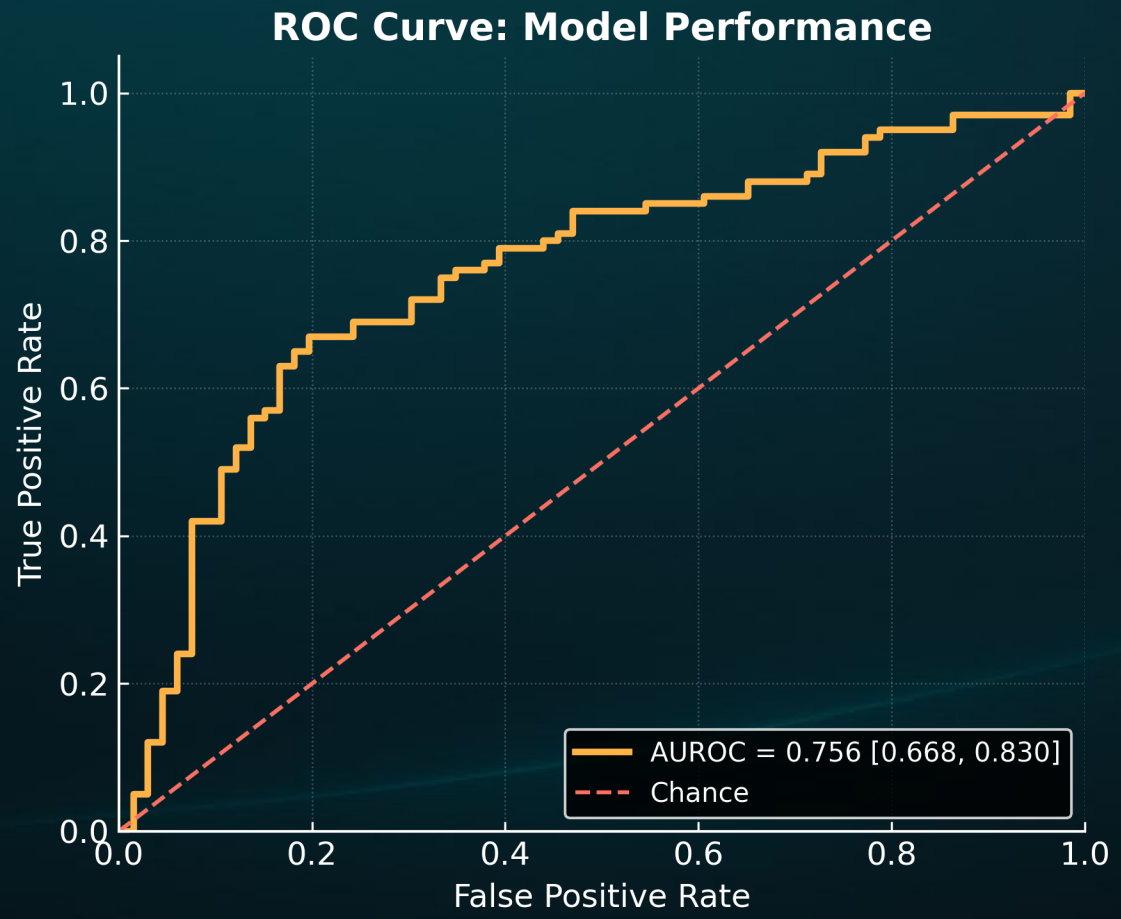
Question: Can we figure out what student behaviors are associated with diagnostic success?



* Two-proportion z-test (pooled variance), two-sided **p < 0.05**

Zhou C and Turner L. *Decoding Diagnostic Success: An AI-Powered Analysis of Student Behaviors in Virtual Pulmonary Encounters. Manuscript in progress.*

Robust Data Culture can transform assessment in medical education



Question: Can we predict if a student will get 2 out of 3 cases correct based on their input?

Methods:

- Combined all 3 case transcripts
- Trained machine learning model
  - logistic regression with L2 regularization on TF-IDF features
  - 5-fold cross validation

2-SIGMA LABS

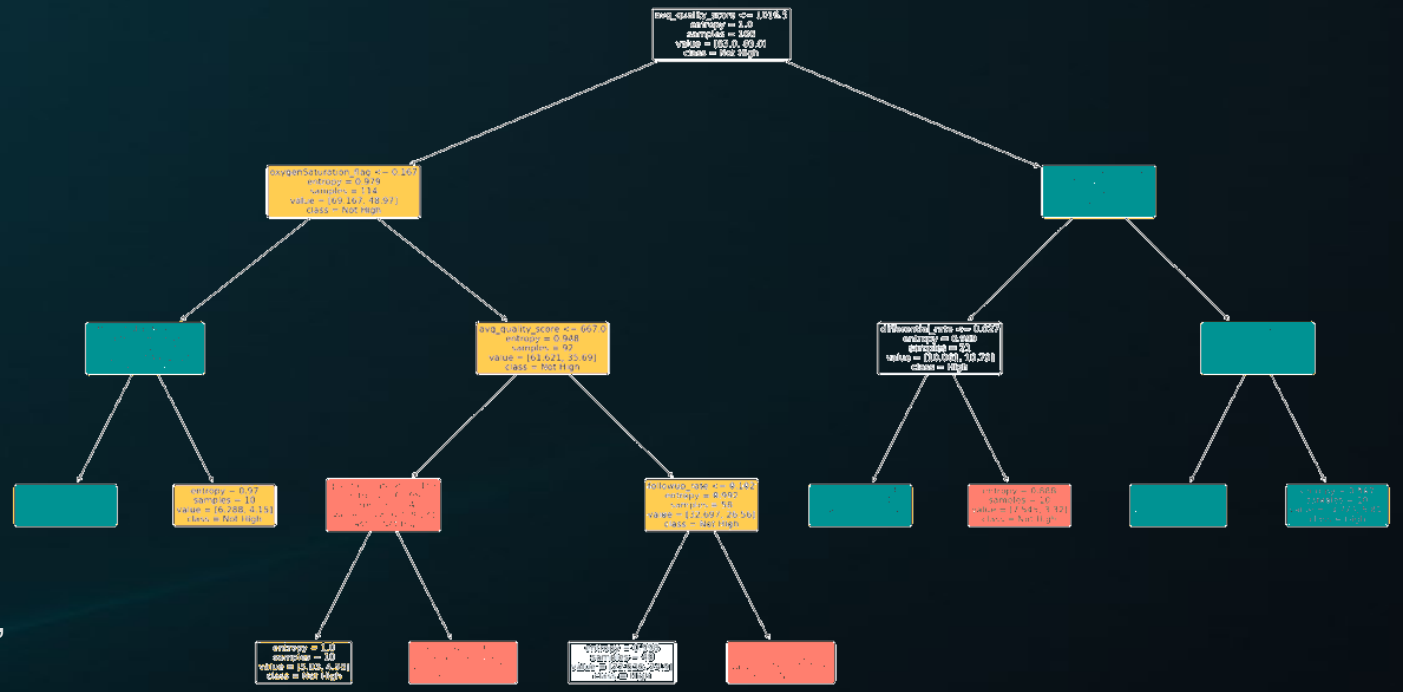
Zhou C and Turner L. *Decoding Diagnostic Success: An AI-Powered Analysis of Student Behaviors in Virtual Pulmonary Encounters. Manuscript in progress.*

We can transform assessment in medical education

Methods:

- Within-subject rule-based phenotyping
- Terminal leaf on the tree = phenotype

- *Features include*:
  - Engagement metrics
    - Average quality score, response count, words used
  - Rule-based component coverage
  - Process signals
    - Rate of questions, follow up, differentials, plans
    - Timing of specific actions
  - Outcomes (correct/incorrect)



Zhou C and Turner L. *Decoding Diagnostic Success: An AI-Powered Analysis of Student Behaviors in Virtual Pulmonary Encounters. Manuscript in progress.*

Can identifying student phenotypes help direct teaching/coaching to specific areas?

Phenotype	*Rules*	*n*	*Accuracy Rate*	*[95% CI]*
A	avg_quality_score ≤ 1016.500 AND oxygenSaturation_flag > 0.167 AND avg_quality_score > 667.000 AND followup_rate ≤ 0.102	48	**62.5**	[48.4–74.8]
B	avg_quality_score ≤ 667.000 AND oxygenSaturation_flag > 0.167 AND question_rate > 0.103	24	**20.8**	[9.2–40.5]
C	avg_quality_score > 1016.500 AND plan_rate > 0.174 AND followup_rate ≤ 0.108	21	**100.0**	[84.5–100.0]
D	avg_quality_score ≤ 1016.500 AND oxygenSaturation_flag ≤ 0.167 AND differential_rate ≤ 0.070	12	**91.7**	[64.6–98.5]
E	avg_quality_score > 1016.500 AND plan_rate ≤ 0.174 AND differential_rate ≤ 0.027	11	**81.8**	[52.3–94.9]
F	avg_quality_score > 1016.500 AND plan_rate > 0.174 AND followup_rate > 0.108	10	**70.0**	[39.7–89.2]

***Zhou C and Turner L**. Decoding Diagnostic Success: An AI-Powered Analysis of Student Behaviors in Virtual Pulmonary Encounters. Manuscript in progress.*

Lessons Learned from Building MAESSCR

💡 Key Principle:
Build the data culture first — MAESSCR emerged from that foundation, not vice versa. Start with governance, not tools.

Contact Us

🎯 Five Critical Success Factors

1 🛡️ Start with Governance, Not Tools

Establish data protocols, consent processes, and risk frameworks before deployment. We formed our governance committee 4 months before launch.

2 ⚙️ Design for Data from Day 1

Build transcript capture, model versioning, and quality logging into your architecture—not as afterthoughts. This enabled our research.

3 ⚖️ Be Intentional but Adaptive

Disaggregate outcomes, include diverse stakeholders in governance, and monitor for bias continuously. We found no disparities because we designed for equity.

4 💡 Build Trust Through Transparency

Be explicit about data use, give students agency, and report findings openly. Our 100% consent rate came from radical transparency.

5 🔄 Close the Feedback Loop

Use insights to improve teaching. When we found O_2 checks predict success, we updated curriculum. Research should inform practice.

2SIGMA LABS

Questions?

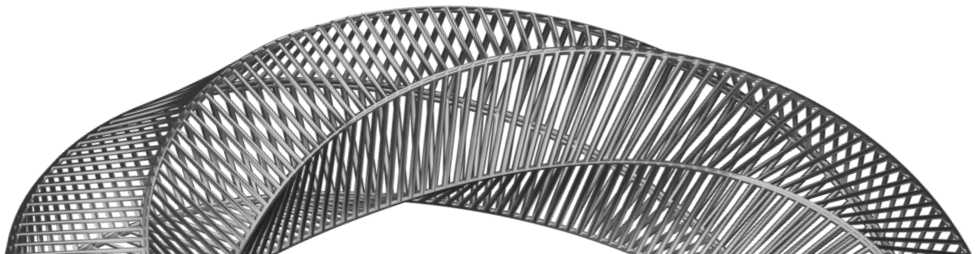
Contact Us

Tools and Infrastructure for AI Capacity Building

in Academic Medicine

Jamie Fairclough, PhD, MPH, MSPharm
Director and Faculty, Engineering & Medicine

DARTMOUTH ENGINEERING
THAYER SCHOOL



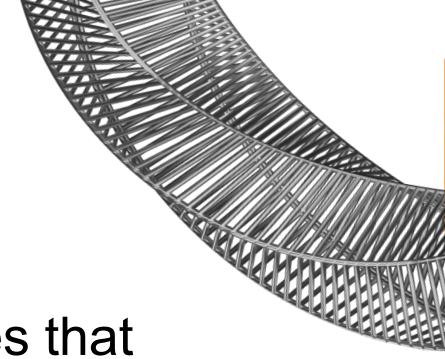
Dartmouth
GEISEL SCHOOL OF
MEDICINE

Artificial Intelligence Coined at Dartmouth

In the summer of 1956, Dartmouth College made history by hosting the world's first academic conclave on the emerging field that would come to be known as artificial intelligence.

Dartmouth and the Dawn of AI

Defining AI Capacity



AI capacity refers to the creation of systems, tools, infrastructure, and talent pipelines that enable sustained, responsible AI adoption and deployment across an institution.

Core Dimensions:

- **Data Infrastructure:** Secure and robust systems for collecting, storing, and accessing high-quality data
- **AI Workforce:** Technical teams collaborating with medical educators, clinicians, domain experts, and research scientists
- **Governance & Ethics:** Frameworks, policies, and oversight for responsible AI development and deployment
- **Technical Systems:** Scalable computing infrastructure and operational tools, platforms, and technologies

AI capacity is multi-dimensional and requires sustainable integration across departments/units, systems, and often multiple institutions to achieve meaningful impact!

Infrastructure and Culture Matter in AI Implementation

- Medical schools and academic health systems face a complex landscape of fragmented data systems, unclear data ownership and governance, and uneven institutional readiness for AI adoption for teaching, research, and care delivery.
- Challenges extend beyond technical barriers to include organizational culture, faculty and clinician engagement, workflow integration, workforce capacity, resource allocation, and alignment across education, clinical operations, IT, and research.
- The success of AI initiatives depends on institutional data culture and infrastructure maturity, just as much as algorithms and compute resources – particularly when AI is expected to support and optimize medical education, clinical decision-making, patient safety, and health system performance.

KEY QUESTION

How can institutions build sustainable, interoperable systems that support AI-driven research, education, and practice?

Building a Data Foundation for AI

Good Data is the Backbone of AI…

Data governance
- ✓ Establishing clear policies, standards, and accountability structures across educational, clinical, and operational data

Robust systems and access control
- ✓ Implementing secure systems with role-based permissions and audit trails for learners, faculty, clinicians, and staff

Quality oversight
- ✓ Ensuring accuracy, completeness, and consistency for clinical, educational, and administrative decision-making

Privacy and compliance
- ✓ Meeting HIPAA, FERPA, and institutional requirements while enabling appropriate data access

Metadata and documentation practices
- ✓ Creating comprehensive data dictionaries and lineage tracking to support reproducibility, transparency, and responsible AI use

Interoperability
- ✓ Seamless integration across educational, clinical, and operational systems (e.g., LMS, SIS, EHR, registries, legacy systems)

Without high-quality, well-governed data, even the most sophisticated AI/ML algorithms cannot deliver reliable or actionable insights in education or patient care.

AI Institutional Readiness

Assessing readiness across data infrastructure, governance, and organizational context is a critical first step that enables the identification of key gaps and opportunities for AI adoption in education, research, and clinical practice.

Common Challenges

Siloed Systems

Data trapped in disconnected platforms limits comprehensive analysis and reduces effectiveness of AI

Inconsistent Metadata

Lack of standardized documentation and shared data definitions make integration difficult across programs and units

Unclear Access Protocols

Ambiguous permissions and approval processes create bottlenecks, slow or stall innovation, and increase compliance risk

Data Infrastructure Readiness Assessment for AI in Academic Medicine

Centralized Data Catalog

Streamlined, Role-Appropriate Access

Standardized Environment

Core Tools to Support AI Development and Deployment

Building effective AI systems requires a robust technical foundation. Three core tools form the infrastructure backbone for sustainable AI development and deployment across medical education, research, and clinical environments.

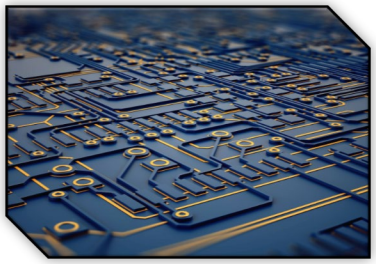
1. **DataOps Pipeline:** Automates data ingestion, validation, and transformation across clinical, educational, and operational data sources to ensure consistent data quality and availability.
   - Technical teams define data quality thresholds, schema validation rules, and preprocessing logic to support reliable, reproducible, and transparent modeling.

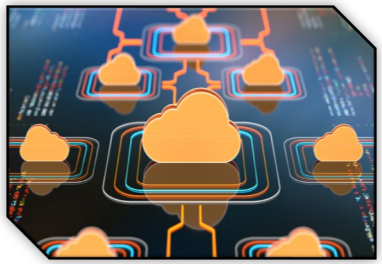
2. **ML Workbench:** Containerized environments provide standardized platforms for secure, reproducible analytics and experimentation, supporting collaboration among learners, faculty, clinicians, and technical teams.
   - Technical teams define required libraries, packages, compute resources, and container configurations to ensure reproducibility and scalability.

3. **Data Catalog:** Comprehensive metadata tagging, lineage tracking, and version control for clinical, educational, and operational datasets, supporting transparency, reuse, and responsible AI development.
   - Technical teams ensure that metadata structures and standards reflect analytic and modeling requirements, supporting interoperability and traceability.

Together, these tools enable AI to move from isolated pilots to scalable, governed infrastructure across education and clinical settings.

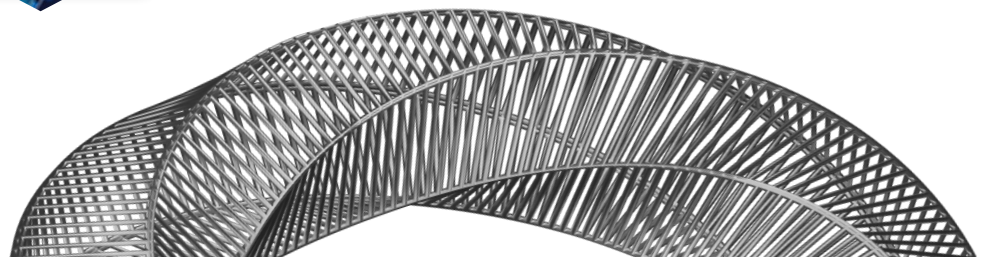
Scalable Infrastructure for AI in Academic Medicine





There are several mission-critical **technical components** required to build and deploy reliable, trustworthy, and sustainable AI systems…

- **Hybrid Cloud Model:** Secure servers or clusters combined with cloud platforms to support workloads with flexibility and scalability.
- **Containerization:** Reproducible environments that enable consistent machine learning analyses and experimentation across environments.
- **APIs and Interoperability:** Seamless integration with existing systems and databases to enable efficient data flow and cross-system functionality.
- **Security and Compliance:** Strict adherence to HIPAA and other regulatory standards ensuring patient privacy, data protection, and institutional risk management.
- **MLOps Pipelines:** Continuous integration and deployment systems that automate model updates, testing, and version control for AI applications.
- **Microservices:** Modular AI components embedded into clinical workflows, educational platforms, and operational systems for real-world use and horizontal scalability.
- **Monitoring & Evaluation:** Comprehensive tracking of model performance, safety signals, risk indicators, fairness metrics, and data drift detection in real-world deployment.



Workforce Development: Empowering Internal Teams

AI success is a team sport. Building AI capacity requires investing in people as much as in technology. A comprehensive workforce development strategy ensures that faculty, staff, providers, researchers, and learners have the knowledge and skills to leverage AI responsibly and effectively across education, research, and clinical practice.

Tiered Training Pathways

AI / Machine Learning Workshops

AI Engineering Laboratories

Professional Certificates

Micro-Credentials

Infrastructure + AI Workforce → Sustainable Impact

Cross-Functional AI Teams

Data Scientists & Informaticists

Engineers & Developers

IT & Systems Specialists

Design

Develop

Implement

Evaluate

Monitor

Medical Educators

Clinicians & Domain Experts

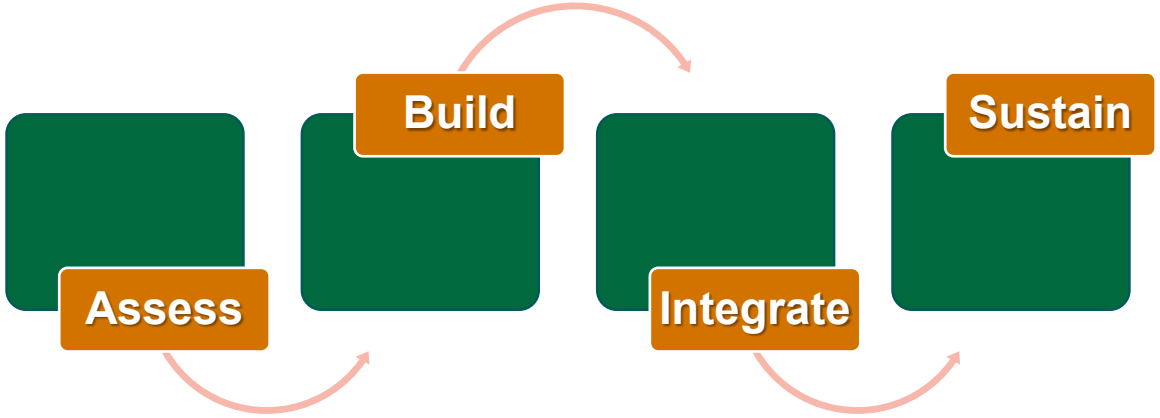
Research Scientists

From Concept to Practice

AI Systems Engineering Laboratory

DARTMOUTH ENGINEERING THAYER SCHOOL 1867

Successful AI implementation follows a structured, iterative approach that enables sustainable adoption. This framework guides AI engineering labs from initial assessment through long-term sustainability.

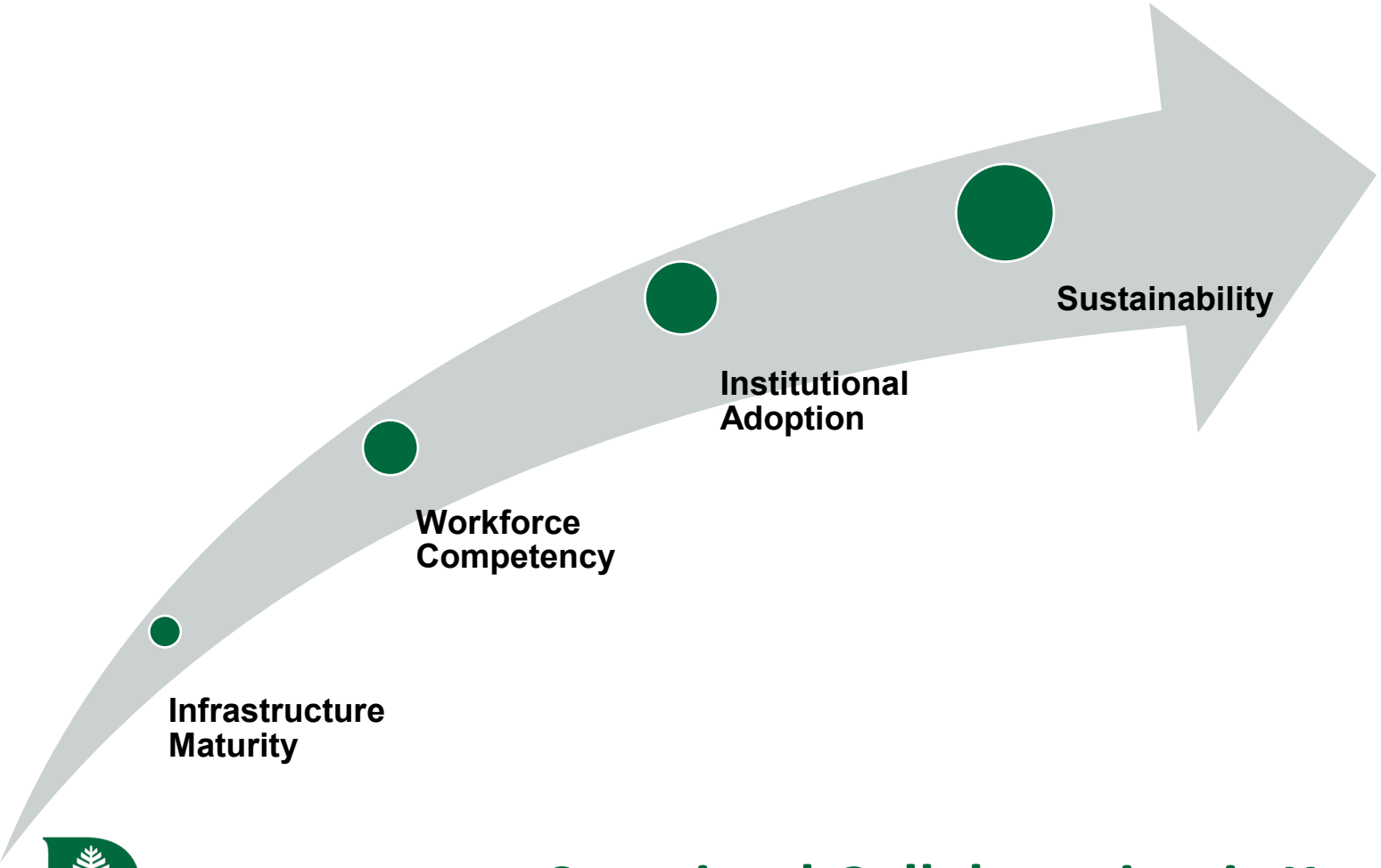


- **Expand strategic partnerships** with other peer health systems, professional societies, and industry partners to accelerate AI adoption and drive innovation.
- **Scale through shared infrastructure and resources** by joining multi-institutional consortia.

1. **Assess readiness**, including governance structures, data quality, clinical and educational workflows, infrastructure maturity, and stakeholder alignment; identify gaps and prioritize investment.
2. **Build modular infrastructure, interdisciplinary teams, and governed tools** that support AI workflows; establish data pipelines, computing environments, and development platforms with scalability, auditability, and deployment in mind.
3. **Integrate systems and engineering principles** into care delivery, clinical decision support, population health, and operational workflows; foster internal collaboration across departments, clinical units, service lines, sites of care, and health-system partners, ensuring system interoperability.
4. **Sustain systems** through ongoing governance, stakeholder training, data monitoring, and regular evaluation using clinical and educational outcomes, safety, cost, and operational efficiency metrics, incorporating user feedback and real-world performance.

Measuring Success: AI Infrastructure, Workforce, and Capacity

Input from multidisciplinary stakeholders is essential to accurately measure success across the AI lifecycle...

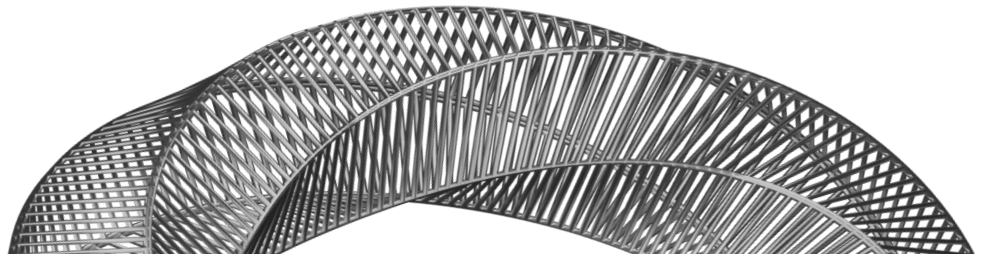


- **Infrastructure Maturity**
  - Data availability and accessibility
  - Interoperability with clinical and educational systems
  - Integration depth with existing platforms
  - Compute scalability and resource utilization
  - System uptime and reliability metrics
- **Workforce Competency**
  - AI literacy levels across teams
  - Number and quality of applied AI projects
  - Certification and training completion rates
  - Cross-functional collaboration effectiveness
- **Institutional Adoption**
  - Cross-unit / institution engagement and participation
  - Research reproducibility and transparency
  - Number of AI-enabled systems, tools, and programs
  - Clinical / operational decision support integration
- **Sustainability**
  - Budget allocation and financial planning
  - Strategic partnerships and external funding
  - Compute, cloud credits, and resource optimization
  - Long-term governance and maintenance plans

...Sustained Collaboration is Key!

Key Takeaways for AI System Development and Deployment

Prioritize governance	• **Establish clear data governance, risk and ethics frameworks, and accountability structures** before building and deploying AI models.
Consider evaluation early	• **Embed continuous evaluation and monitoring** across the AI system lifecycle to inform design, deployment, and scaling decisions.
Build cross-functional teams	• **Assemble diverse teams**, including technical specialists, clinicians, domain experts, leaders, and end-users. External partners can help.
Start small but design for scalability	• **Begin with focused pilot projects** that demonstrate value, while architecting systems with future growth and reuse in mind.
Use pilot projects to demonstrate value	• **Quick wins** from well-designed pilot projects build institutional confidence, secure buy-in, and establish trust in AI systems.
Engage stakeholders continuously	• **Continuous engagement** with stakeholders ensures systems meet real needs and builds a culture of shared ownership.
Document, document, document!	• **Comprehensive documentation** of data sources, methods, decisions, and evaluations for reproducibility and knowledge transfer.



Thank You!

Jamie Fairclough, PhD, MPH, MSPharm

Jamie.L.Fairclough@Dartmouth.edu

Connect on LinkedIn