CORE TRAINING WORKSHOP

Hands-On Training on Artificial Intelligence

Understanding AI Research Oversight in Healthcare

Mark Lifson, PhD

AI/ML Director

Core Training Workshop Series

2025

AI vs Traditional Research | Clinical Intent | Risk Assessment | Oversight Framework

SPEAKER INTRODUCTION

Mark Lifson, PhD

Director, AI/ML Engineering

Mayo Clinic Center for Digital Health

IMPACT METRICS

EXPERIENCE	15+ Years
CITATIONS	1300+
PATENTS	Diagnostics, Regulatory & Mechanical Systems

TECHNICAL EXPERTISE

- Medical AI/ML
- FDA/CE Regulatory
- Diagnostic Biosensors
- Clinical Translation
- Health Equity
- SaMD Governance

KEY CONTRIBUTIONS

Regulatory & Oversight

- AI/ML Expert IRB Reviewer
- FDA/CE-Mark Approvals
- Medical & Scientific Affairs

Technical Achievements

- Hematology AI Systems
- Biosensor Platforms
- Point-of-Care Diagnostics

ENGINEERING · RIT

BIOMEDICAL · University of Rochester

RESEARCH · Stanford

HEALTHCARE · Mayo Clinic

INDUSTRY · Abbott

BIOSENSORS

CAREER IMPACT

Biosensor Development
Label-free detection platforms

AI/ML in Clinical Practice
FDA-approved diagnostic algorithms

Healthcare Governance
Frameworks for responsible AI deployment

Regulatory Innovation
Automated oversight systems & patents

Health Equity Research
Bias mitigation in diagnostic AI

WHAT MAKES AI DIFFERENT? DATA

Traditional Research

1. Lab bench
2. Animals
3. **IRB starts**
4. Humans
5. Trials

VS

AI Research

1. **Human data**
2. Training
3. Validation
4. Testing
5. *IRB?*

AI research **starts with human data** - there's no pre-human exploration phase

Phases of AI Clinical Evaluation

P1

Exploratory/Discovery

Pre-clinical

- Identify study aims & hypotheses
- Literature searches
- Secondary data analysis
- Professional society guidelines
- Initial algorithm development

P2

Pilot/Validation

Early feasibility

- Preliminary safety assessment
- Performance evaluation
- Small study cohorts
- Analytical validation
- Clinical validation metrics

P3

Intervention/Treatment

Clinical efficacy

- Confirms clinical efficacy
- Safety & risk assessment
- Impacts patient care
- Pivotal trials
- Post-market surveillance

Analytical Validation
Accuracy, Reliability, Precision

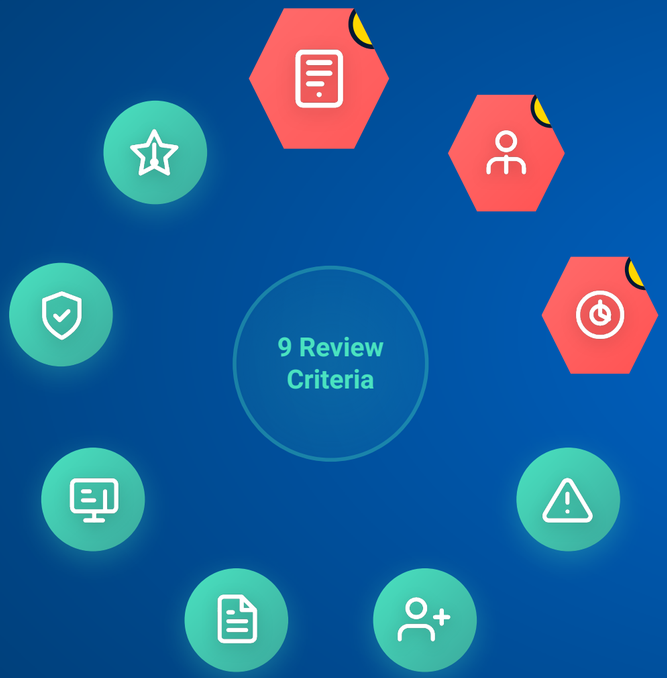
Clinical Validation
Sensitivity, Specificity, PPV/NPV

Performance Monitoring
Long-term effectiveness & safety

AI requires **three distinct phases** - each with unique validation requirements

9 CRITERIA, 1 REVIEW

All Criteria (9) | Research Definition (3) | Risk Management (2) | Subject Protection (4)



FDA-Regulated Device?

21 CFR 812, 820 | Policy | FDA Guidance | ISO Standards

Software intended for diagnosis, treatment, prevention, or mitigation of disease with quality system requirements

IRB MUST DETERMINE

Compliance with FDA Quality System Regulation (21 CFR 820) design controls, including verification and validation activities per 21 CFR 820.30(g), software validation per 21 CFR 820.70(i), and determination of significant risk vs non-significant risk device classification under 21 CFR 812.

Based on: Belmont Report • Institutional Policy | 21 CFR 50, 56, 812, 820 • 45 CFR 46 • FDA SaMD Guidance • ISO 13485, 14971 • 21st Century Cures Act

What Both Sides Want

IRB SPECIALISTS

Speaking Regulatory

> "Per 45 CFR 46.111(a)(2), risks must be reasonable in relation to anticipated benefits..."

Need: Complete documentation demonstrating compliance with Common Rule, FDA regulations, and institutional policies

TRANSLATION GAP

AI RESEARCHERS

Speaking Technical

> "Our CNN achieves 94% AUC with cross-validation on the hold-out test set..."

Want: Clear guidance on what's needed without having to learn regulatory frameworks

The Challenge: How do we bridge this gap without requiring either side to become experts in the other's domain?

DETERMINING CLINICAL INTENT

The same technology described differently triggers different regulatory pathways

⚠ Clinical Trigger Words

⚠ diagnose | ⚠ treat | ⚠ screen | ⚠ detect | ⚠ clinical decision | ⚠ deployed | ⚠ patient care

Research Indicators

✓ investigate | ✓ analyze | ✓ explore | ✓ retrospective | ✓ patterns | ✓ correlations | ✓ hypotheses

STUDY OBJECTIVES

The primary objective is to develop a deep learning model that will ⚠ diagnose pneumonia from chest X-rays with accuracy comparable to board-certified radiologists. The system will be ⚠ deployed in emergency departments to ⚠ expedite treatment decisions for patients with respiratory symptoms.

⚠ CLINICAL INTENT - FDA PATHWAY LIKELY

RESEARCH AIMS

This study will ✓ investigate novel biomarkers in retinal images by ✓ analyzing patterns across large retrospective datasets. We aim to ✓ explore correlations between retinal features and cardiovascular risk factors to ✓ generate hypotheses for future prospective studies.

✓ EXPLORATORY RESEARCH - IRB FOCUS ON DATA PRIVACY

METHODOLOGY

Our AI platform will ✓ analyze patient data to ⚠ predict 30-day readmission risk. Results will be ? provided to clinicians for ✓ evaluation of potential ⚠ treatment modifications. The system will ? support clinical teams in ? decision-making processes.

? MIXED SIGNALS - CONTAINS BOTH CLINICAL AND RESEARCH LANGUAGE

DATA ANALYSIS PLAN

The AI algorithm will ⚠ screen mammography images to ⚠ identify suspicious lesions requiring immediate radiologist review. Cases flagged as high-risk will ⚠ trigger automatic scheduling for follow-up imaging, with the AI assessment ⚠ documented in patient records.

⚠ CLINICAL INTENT - DIRECT PATIENT IMPACT

IMPACT SPECTRUM & OVERSIGHT REQUIREMENTS

Patient impact level determines IRB oversight requirements

✓ Minimal ⚠ Moderate ✗ Significant

Silent/Shadow Mode

✓ LOW

AI runs in background without clinical use

Example: *Algorithm validation study*

IRB Oversight: *Standard data review + annual reports*

Advisory Mode

⚠ MOD

AI provides recommendations for review

Example: *Risk scores shown in EHR*

IRB Oversight: *Performance monitoring + quarterly safety reviews*

Active Intervention

✗ HIGH

AI triggers actions or alerts directly

Example: *Auto-schedules follow-ups*

IRB Oversight: *Real-time monitoring + DSMB + FDA consultation*

Oversight Intensity: Proportional to the degree of patient care impact

RISK CONTROL PRIORITIZATION MATRIX

Q2: Patient Impact

	Exploratory	Translational	Clinical
Direct Care	Enhanced Review **Medium-High Risk** Exploratory with direct impact	Full IRB **High Risk** Translational affecting patients	Full IRB + Monitoring **Critical Risk** Clinical deployment to patients
Indirect	Standard Review **Low-Medium Risk** Exploratory with indirect effects	Enhanced Review **Medium Risk** Translational validation phase	Full IRB **Medium-High Risk** Clinical decision support
No Impact	Minimal **Low Risk** Pure research, no patient involvement	Standard Review **Low-Medium Risk** Algorithm development phase	Enhanced Review **Medium Risk** Clinical readiness testing

Q1: Clinical Intent

Q3 Technology Characteristics

Select technology features to determine appropriate controls

ML/Predictive Models

- ☐ **Autonomous operation** — Operates without human input
- ☐ **Clinical decision influence** — Outputs used in care decisions
- ☐ **Continuous learning** — Updates from new data
- ☐ **Age extremes included** — Users <10 or >80 years
- ☐ **Patient data usage** — Processes medical records

Complexity Score 0/ 17

Based on ISO 14971 Risk Management