Ethical and Practical Oversight of AI Research

Adapting existing frameworks for emerging technology

AUTHORS

Mark Lifson

Tamiko Eto

2025

Research Presentation

AI RESEARCH 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 causes angst about AI within research communities?

Let me share what I keep hearing...

NOT UNIQUE TO AI

Traditional research challenges persist in AI — amplified by scale and speed

Outcome Uncertainty Non-deterministic results	→	RISK CALCULATORS **Proven Framework** Clinical risk scores provide probability estimates — AI extends this capability
Limited Explainability Complex decision pathways	→	COMPLEX ALGORITHMS **Proven Framework** Statistical models have inherent opacity — AI increases complexity
Data Requirements Extensive training sets needed	→	CLINICAL RESEARCH **Proven Framework** Large studies require thousands of records — AI requires millions
Potential Bias Training data limitations	→	ALL RESEARCH **Proven Framework** Selection bias exists in all research — AI propagates it at scale

WHAT MAKES AI DIFFERENT? DATA

Traditional Research

VS

AI Research

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

Phases of AI Clinical Evaluation

Exploratory/Discovery

Pre-clinical

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

Pilot/Validation

Early feasibility

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

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

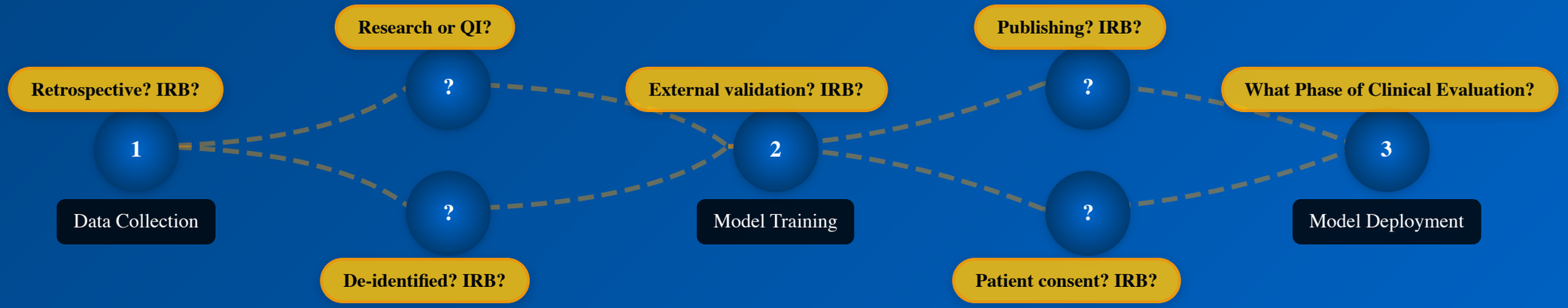
WHAT MAKES AI DIFFERENT? AMBIGUITY

Traditional Research

Clear IRB entry point

AI Development

Multiple unclear decision points



AI development has **multiple ambiguous IRB decision points** throughout the process

GOOD NEWS: SAME FRAMEWORKS APPLY

FDA Regulations

Medical Device & Software

21 CFR 820.30

Design Controls - Standard V&V requirements now applied to AI models

21 CFR 812

IDE regulations - Clinical trial requirements now covering AI devices

21 CFR 11

Electronic records - Existing compliance rules for digital data

ICH GCP E6(R3)

Good Clinical Practice - International standards for trials

FDA SaMD Guidance

Software as Medical Device - Pre-AI framework being adapted

ISO Standards

International Standards

ISO 14971

Risk management - Standard medical device risk framework

ISO 62304

Software lifecycle - Traditional software development processes

ISO 13485

Quality management - Standard QMS requirements

IEC 62366

Usability engineering - Human factors requirements

ISO/IEC 82304

Health software - General software product requirements

IRB/Ethics

Human Subjects Protection

45 CFR 46

Common Rule - Human subjects protections (revised 2018)

21 CFR 50

FDA human subjects - Standard informed consent requirements

21 CFR 56

IRB requirements - Traditional institutional review processes

32 CFR 219

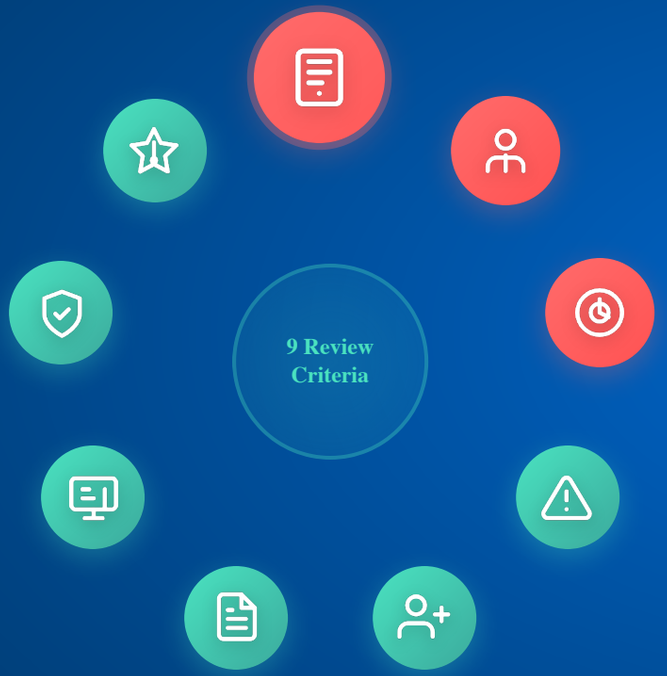
DoD protections - Military-specific human subjects rules

Belmont Report

Ethical principles - Respect, beneficence, justice (1979)

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?

The Solution: Three Key Questions

*A framework that **translates** technical complexity into **reviewable criteria***

1

Clinical Intended Use?

Determines regulatory pathway and efficacy evidence requirements

2

Patient Care Impact?

Clarifies risk profile and human subjects protections needed

3

Technology Risks?

Identifies AI-specific considerations often overlooked

These questions **focus oversight where it matters** – no AI expertise required

DETERMINING CLINICAL INTENT

Question 1: Clinical Intended Use

Is there a clinical intended use?

Clinical Use

- Diagnostic algorithms for patient care
- Treatment recommendation systems
- Risk stratification tools
- Clinical decision support
- Patient monitoring systems

Exploratory Research

- Pattern discovery in datasets
- Hypothesis generation
- Research-only algorithms
- Retrospective analyses
- Method development studies

Regulatory Pathway Implications

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

ASSESSING PATIENT IMPACT

Question 2: Patient Care Impact

Does it impact patient care?

Direct Impact

- AI outputs influence clinical decisions
- Alerts trigger immediate actions
- Treatment recommendations provided
- Real-time decision support active
- Results documented in patient records

Indirect/No Impact

- Retrospective analysis only
- Research insights for future care
- No real-time clinical influence*
- Results inform protocol development
- Population-level studies only

Risk Profile Determines Oversight Level

Direct patient impact requires enhanced safety monitoring, real-time performance tracking, and fail-safe mechanisms. Indirect impact allows more flexibility in study design and timing.

IMPACT SPECTRUM & OVERSIGHT REQUIREMENTS

Patient impact level determines IRB oversight requirements

Minimal **Moderate** **Significant**

Silent/Shadow Mode

AI runs in background without clinical use

Example: *Algorithm validation study*

IRB Oversight & Expectations: *Standard data review + annual reports*

Advisory Mode

AI provides recommendations for review

Example: *Risk scores shown in EHR*

IRB Oversight & Expectations: *Performance monitoring + quarterly safety reviews*

Active Intervention

AI triggers actions or alerts directly

Example: *Auto-schedules follow-ups*

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

Oversight Intensity: Proportional to the degree of patient care impact

Question 3: Technology-Specific Characteristics

What technology characteristics require oversight?

Algorithmic Characteristics

- Training data demographics
- Algorithm transparency level
- Decision pathway complexity
- Update frequency requirements

Feedback Loops

- Self-fulfilling predictions
- Behavior modification
- Data contamination
- Perpetual learning risks

System Risks

- Cascading failures
- Automation dependency
- Silent failures
- Integration complexity

Why Implementation Matters

While the risk categories aren't new, AI's speed, scale, and autonomous nature create unique implementation challenges that traditional oversight must adapt to address.

TECHNOLOGY CHARACTERISTICS & CONTROLS

While risks aren't unique to AI, their manifestation requires technology-specific mitigations

Traditional ML

KEY RISKS

- Overfitting to training data
- Feature selection bias
- Class imbalance effects

POTENTIAL MITIGATIONS

- ✓ Validation strategies
- ✓ Performance monitoring
- ✓ Documentation of limitations

Deep Learning

KEY RISKS

- Black box decision-making
- Hidden spurious correlations
- Catastrophic forgetting
- Model drift over time

POTENTIAL MITIGATIONS

- ✓ Explainability approaches
- ✓ Human oversight mechanisms
- ✓ Continuous evaluation

Generative AI

KEY RISKS

- Hallucination of medical facts
- Training data memorization
- Adversarial manipulation

POTENTIAL MITIGATIONS

- ✓ Output verification processes
- ✓ Privacy safeguards
- ✓ Usage boundaries and guardrails

TECHNOLOGY COMPARISON FOR IRB REVIEW

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
- [ ] **Limited training demographics** — Trained on <3 ethnic groups

Complexity Score 0/ 29

Based on ISO 14971 Risk Management

BUILDING YOUR TECHNOLOGY CHECKLIST

Institution Risk Profile: Conservative | **Standard** | Streamlined — Balanced approach

Answer questions to build **personalized requirements** based on your technology and use case

1 Define Your Context

What type of AI technology?

Machine Learning (ML)

Q1: What phase is your research?

- Discovery/Feasibility
- Translational/Pilot
- Clinical Validation

2 Identify Characteristics & Risk Approach

3 Your Protocol Requirements

IRB Oversight Level: MINIMAL

Expedited review, minimal reporting

Review & Approval 1

- Expedited review pathway

Monitoring & Reporting 1

- Annual progress report

WORK IN PROGRESS

Final Version: EOY 2025

Seeking Collaborators & Feedback

Questions?

A practical three-question framework for AI research oversight

- Is this human subjects research requiring IRB review? (Including FDA-regulated device studies)
- What is the potential for impact on human subjects?
- Is the technical risk acceptable relative to benefits?

lifson.mark@mayo.edu
linkedin.com/in/marklifson