



# Evaluating the Efficacy of Artificial Intelligence Ethics Frameworks: A Case Study

Vivian Qiang, Cognitive Science Program  
Dr. AJung Moon, Department of Electrical and Computer Engineering  
Responsible Autonomous & Intelligent System Ethics (RAISE) Lab



## Introduction

### What do we mean by Artificial Intelligence (AI)?

Within the context of this project, AI technology allows a system to “learn” from data to predict events and make decisions **without explicit programming**.

### Why did we develop this project?

The increasing use of AI-powered technology raises numerous ethical concerns such as **privacy, security, accountability, fairness and bias**, etc. In response, ethical guidelines have been developed to promote responsible innovation in AI technology; however, stakeholders have found many frameworks unhelpful or difficult to implement.

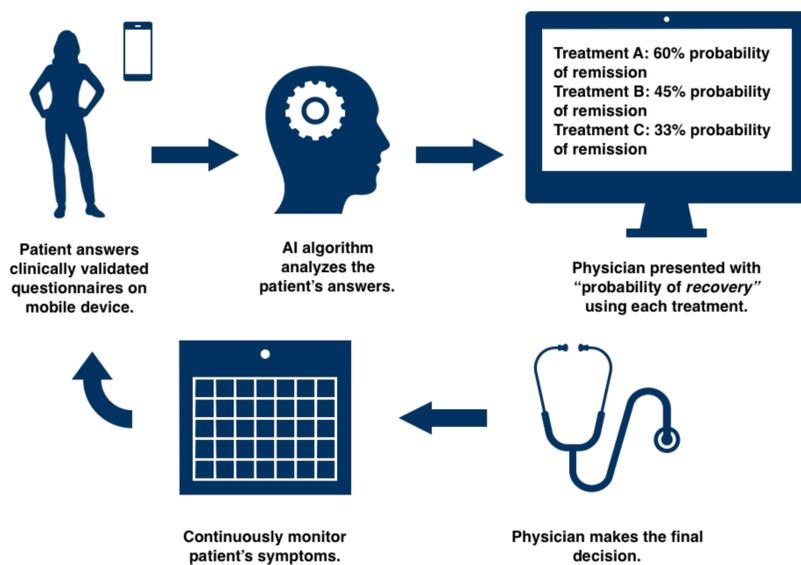
### Research Objective

**This project aims to evaluate the efficacy of existing AI ethics frameworks by applying their recommendations to a start-up’s AI-powered product.**

## Methodology

### The Technology

- Aids physicians in selecting the best treatment for patients with depression
- The product is currently in development and undergoing clinical trials



### Data Collection

- Conducted interviews virtually with company stakeholders and potential users of the product

## Results

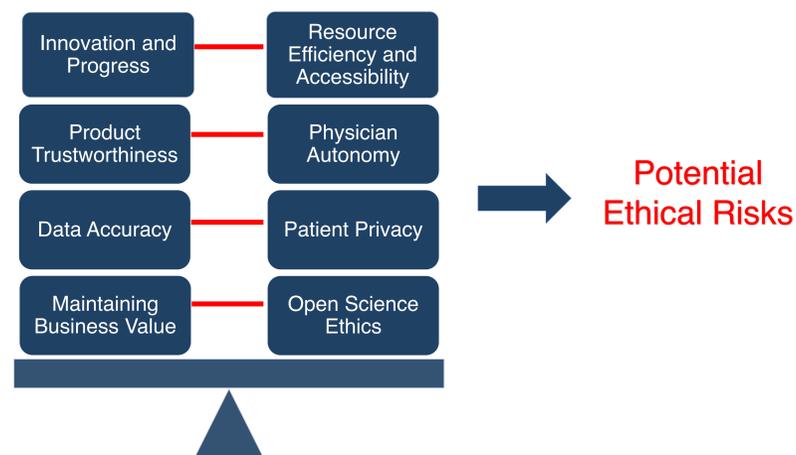
### Framework Categorization and Review

- Reviewed and categorized 38 AI ethics frameworks
- Identified gaps in current frameworks:
  - Lack in concrete, measurable outcomes
  - Minimal mention of pre-existing industry regulations (ex. HIPAA, IEEE standards)
  - Insufficient guidelines for solution-making
- Selected **four toolkits** to analyze based on:
  - Author domain diversity (NPO, international organization, academia, government)
  - Intended use (applied to product in development)
  - Appropriateness for application by third-party assessor
  - Expected timeline (3-4 months) and resources required

Principles	Toolkits	Policies and Technical Tools

### Discovered Value Tensions

- Identified several conflicts between stakeholder goals and values by applying the selected toolkits
- Produced reports for each selected framework highlighting discovered ethical risks



## Discussion and Conclusions

### What have we learned?

- Two primary forms of actionable ethical assessments emerged: checklists and process-based toolkits
- Both serve different but important purposes:

Checklists	Processes
Structure ethical principles within regulatory points	Explore value tensions and ethical issues by applying broader principles
Simpler to complete and implement	Allow for more collaboration and flexible adaptations
Goal-focused with measurable, attainable standards	Acknowledge sociocultural context and capture potential nuances
May need to be tailored to the specific context to be effective	Frameworks without checkpoints/assessments may be more vulnerable to ethics washing

- Ethics evaluations are fundamentally about **listening**, and assessors must include and encourage participation from all impacted groups

### Why are our findings important?

- **For industry actors:** recognize and manage risks associated with their products by bridging the gap between theoretical ethical principles and real-world applications
- **For policymakers:** develop AI regulations by determining which ethics assessments are most effective in discovering and mitigating risks

## Future Directions

- Will present framework reports and discovered ethical risks to company stakeholders
- Planning to hold a workshop with company stakeholders to brainstorm solutions for discovered ethical risks
- Currently studying another AI-powered decision-making tool in the health sector to develop a series of case studies

## Acknowledgements

I would like to thank my supervisor, Dr. AJung Moon, for her incredible guidance and passionate leadership. I am also grateful to our fellow RAISE lab mates, Shalaleh Rismani and Jake Chanenson, for their advice and support.