Foresight:
How Might Theory Let Us Anticipate PR Crises Around Using AI in Healthcare?

UNDERSTANDING MORAL OUTRAGE AND AI ETHICS

BY

BRINDA JOSHI

MASTER'S STUDENT AT UNIVERSITY OF WASHINGTON

2025 March | Persuading Ethical UX Design

Introduction

- **AI in healthcare is advancing** rapidly, offering new possibilities for Administrative Automation, treatment and patient care.

- However, ethical concerns can trigger **public backlash** if companies fail to anticipate risks.

- This analysis applies to two key AI healthcare challenges:
  - **AI generating healthcare content** (e.g., AI-written diagnoses, automated health advice).
  - **AI being trained on patient data without clear consent.**

- Moral outrage is predictable if we analyze the ethical dimensions beforehand

.

- This presentation applies **Haidt (2012)'s moral foundations** to anticipate risks in AI-driven healthcare decisions.

The Six Moral Foundations
(Haidt's Model)

Jonathan Haidt's research identifies **six core moral foundations** that shape human emotions and ethical judgments.

These principles explain why people react strongly to AI decisions in healthcare.

By understanding these triggers, companies can **avoid backlash and create responsible AI.**

CARE/HARM

FAIRNESS/CHEATING

LOYALTY/BETRAYAL

AUTHORITY/SUBVERSION

SANCTITY/DEGRADATION

LIBERTY/OPPRESSION

Moral Foundations and AI Scenarios

1. CARE/HARM

People react strongly when AI **compromises patient safety or dignity**. AI in healthcare must **prioritize patient well-being and safety** to maintain trust.

Haidt (2012): ""We respond emotionally to signs of violence or suffering, particularly when a child is involved, particularly our own child."

◆ **Guidance 1 :**
AI-generated diagnoses must undergo **human review** to prevent harmful misdiagnoses.

◆ **Guidance 2 :**
AI systems handling patient data must have **strong security measures** to prevent breaches.

2. FAIRNESS/CHEATING

People expect AI to be **fair and transparent** in decision-making. When AI systems **discriminate** or **favor certain groups,** it triggers strong public backlash.

Haidt (2012): "The Fairness/cheating foundation is about proportionality and the law of karma. It is about making sure that people get what they deserve, and do not get things they do not deserve."

◆ **Guidance 1 :**
AI chatbots should provide **consistent healthcare advice** to all users, regardless of economic status

◆ **Guidance 2 :**
AI training data must be **diverse and unbiased** to prevent discriminatory healthcare decisions.

3. LOYALTY/BETRAYAL

Trust is essential in healthcare. Patients expect **AI-driven systems** to be **loyal to their best interests**, not driven by corporate greed.

Haidt (2012): "Similarly, in The Inferno, Dante reserves the innermost circle of hell—and the most excruciating suffering—for the crime of treachery. Far worse than lust, gluttony, violence, or even heresy is the betrayal of one's family, team, or nation."

◆ **Guidance 1 :**
Hospitals must balance **AI recommendations with human expertise** to maintain trust in patient care.

.

◆ **Guidance 2 :**
Companies must **not sell patient data** without explicit consent, as it damages trust.

4. AUTHORITY/SUBVERSION

Societies function through **rules and structure**. When AI **ignores regulations** or **bypasses oversight**, it creates concerns about accountability.

Haidt (2012): "The urge to respect hierarchical relationships is so deep that many languages encode it directly."

◆ **Guidance 1 :**
AI-generated health content must be **approved by medical professionals** before use.

◆ **Guidance 2 :**
AI companies must **follow strict healthcare regulations** to avoid ethical violations.

5. SANCTITY/DEGRADATION

People value **dignity and respect** in healthcare. When AI **reduces patients to mere data points**, it disregards the **emotional and personal aspects** of medical care.

Haidt (2012): "If we had no sense of disgust, I believe we would also have no sense of the sacred."

◆ **Guidance 1 :**
 AI-powered health advice should consider **human emotions and sensitivities.**

◆ **Guidance 2 :**
 AI should **never automate life-and-death decisions** without human involvement.

6. LIBERTY/OPPRESSION

Users want **control over their own healthcare data.** If AI systems **force consent** or **manipulate choices**, it creates feelings of oppression.

Haidt (2012): "Evolved in response to the adaptive challenge of living in small groups with individuals who would, if given the chance, dominate, bully, and constrain others."

◆ **Guidance 1 :**
AI health platforms must provide **clear options for human review and second opinions.**

◆ **Guidance 2 :**
AI must offer **opt-in data-sharing choices** rather than forcing users into it.

References : "Haidt, J. (2012). The righteous mind: Why good people are divided by politics and religion. Vintage."

Conclusion

Predicting moral outrage allows us to **design AI that is ethical and human-centered.**

Understanding these moral foundations helps companies **anticipate ethical dilemmas before they happen.**

However, theory alone isn't enough, AI teams must still conduct real-world studies to confirm these insights.

AI in healthcare has **huge potential**, but ethical concerns must be addressed early.

What do you think?
How can AI companies design systems that **respect user values** and **avoid public backlash?**

Thank you

BRINDA JOSHI

MASTER'S STUDENT AT UNIVERSITY OF WASHINGTON

2025 March | Persuading Ethical UX Design