AI Ethics of Autonomy (Part 2)

By: Aman Matin, Michael Matthews, & Martin Kapiti

Introduction

Aman

Intro

- **Definition in Autonomy:**
  - Autos ‘self’ + nomos ‘law’
  - Philosophical, Physical, Mental, and Psychological Autonomy
- **Importance:** AI is being used to influence (nudge) real world decisions that humans make
- **Talking points:** 2 case studies, nudges, dark patterns, conclusion

Autonomy Case 1: AI in Insurance

Martin

Case Study #1: AI in Cardiac Health

What the company does:

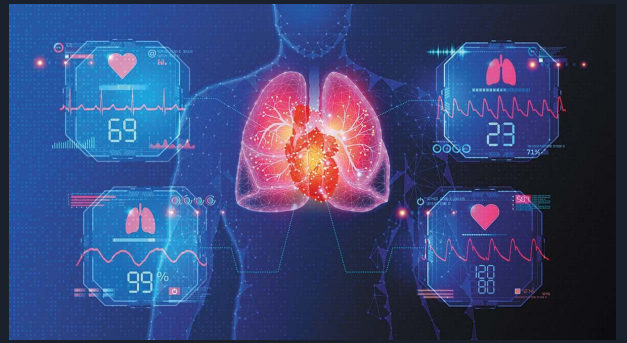
- Takes a 5-minute EKG
- AI compares millions of data points
- Matches heartbeat patterns against large datasets

Why AI?

- Humans cannot analyze massive pattern comparisons
- AI identifies statistical correlations quickly

Initial Output:

- Green = Good
- Red = Risk



The Threshold Problem

The ethical issue here is not just detection it is also between classification.

Where do we draw the line between green and red?

- 99% chance of death in 2 hours → clearly red
- 9% chance over 10 years → red? green?

There is a key ethical shift here:

- Biological health optimization does NOT EQUAL autonomy optimization

Health vs Autonomy

If the system maximizes biological safety:

- Almost everyone gets a red warning
- People become anxious
- Life becomes restricted

If it maximizes autonomy:

- Avoid overwhelming people with low-risk alerts
- Preserve lifestyle freedom
- Provide meaningful but proportionate information

So an Ethical Question here is:

What threshold best supports informed freedom rather than fear?

Extreme Case: The Right Not to Know

If AI predicts death within 24 hours, should the system tell the person?

- Yes → supports informed autonomy
- No → preserves meaningful final hours

So…. Can withholding information ever increase autonomy?

This introduces tension between:

- Autonomy
- Psychological well-being
- Human dignity

Autonomy Case 2: AI in Insurance

Martin

Case Study #2: Dynamic Insurance

Dynamic insurance collects behavioral data:

Car insurance:

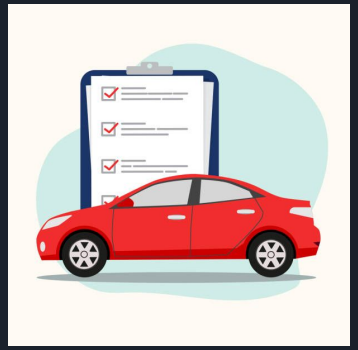
- Tracks speed, braking, acceleration
- Premium adjusts accordingly

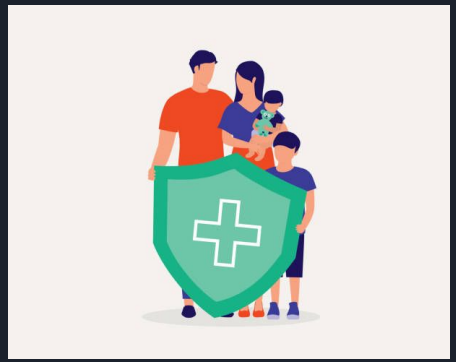
Health insurance:

- Tracks exercise, risk behavior
- Rates change based on real-time data

Initially seems autonomy-enhancing:

- You voluntarily sign up
- You control your lifestyle choices
- But there is a caveat here...





The Autonomy Trap

At what point does autonomy flip?

- You choose to join
- You gain control over premiums
- Self-determination increases

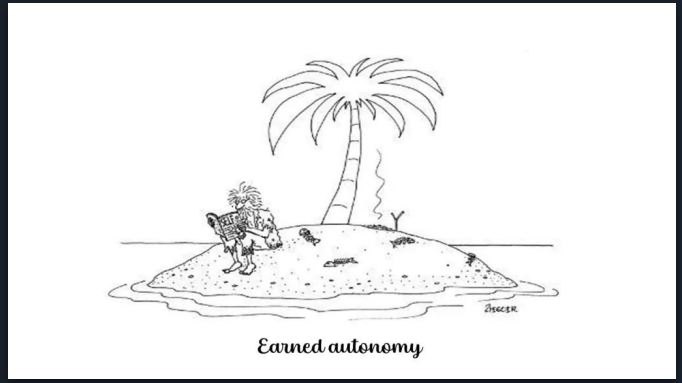
Later stage:

- Constant warnings
- Behavioral nudging
- Feeling monitored
- Fear of financial penalty

Does this become:

- A system of self-imposed surveillance?
- A soft coercion structure?

Can voluntarily limiting freedom eventually reduce freedom?



Ethics, Profit & Freedom

Can freedom destroy freedom? If I freely choose:

- To be constantly monitored
- To be nudged
- To restrict my behavior

Am I more free or less free? AI forces us to rethink:

- Self-determination
- Information
- Risk
- Control
- The limits of autonomy itself

Challenges within Autonomy

Aman

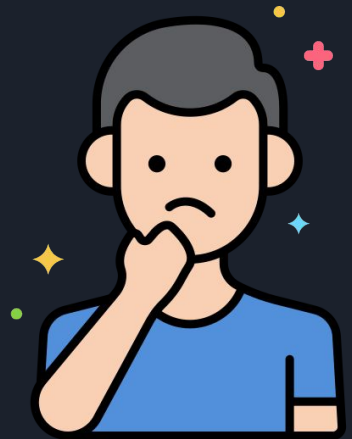
Nudges

- **What is a nudge?**
  - Nothing can be explicitly forbidden
  - No monetary, or tangible appeal, like money
  - Signage on stairs example
  - A small push in a certain direction

- **Nudges seen with AI:**
  - Appeals to our specific and unique psychological profiles
  - Optimizes which nudges work for different groups

Concerns with Nudges

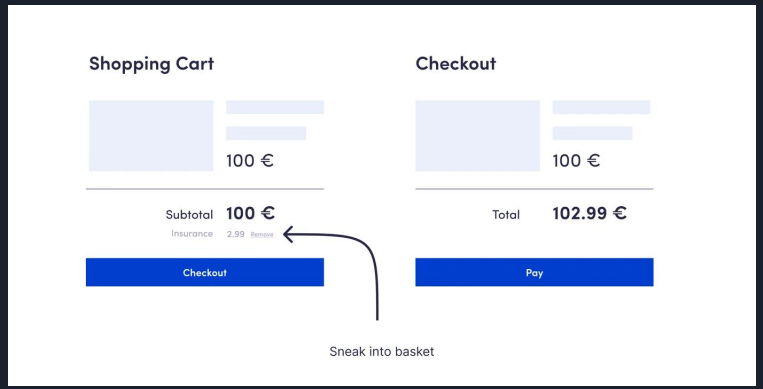
- Potential interference with a users autonomy
  - American Heart Association example
  - Appeal based on vulnerability
  - Is there a point in which the nudge can transfer into manipulation?
  - Could this be used maliciously?
- Limitations/attacks or opportunities to exercise autonomy?
  - Up to the individual
  - Unclear whether or it is a limitation or opportunity



Dark Patterns

Michael

Dark Patterns Against Autonomy



A dark pattern is a user interface design that is intentionally crafted to trick users into actions that may not be in their best interests, such as making unintended purchases or signing up for unwanted services

Examples: Misleading links, pre-checked boxes, confusing visual layout

Dark Patterns at the Limit

“The thought process that went into building these applications. Facebook being the first of them. Was all about: ‘How do we consume as much of your time and conscious attention as possible?”

-Sean Parker, Facebook’s founding president

Dopamine is a brain chemical released when experiencing pleasure (similar biological mechanism to highly addictive drugs).

Facebook discovered that receiving a "Like" releases a small hit of dopamine.

Theory

Michael

Control With Nudges

Postscript: control vs disciplinary

Key Figures: Jean-Paul Sartre, Michel Foucault, and primarily Gilles Deleuze.

Deleuze’s "Society of Control" (1990s): this is where nudge came from

Deleuze defined "control" long before AI existed. Argued that yes, it is entirely possible to take away human freedom *without* ever using prohibitions.

Conclusion

Aman

Conclusion

- **Takeaway:** AI could be used as a good catalyst to guide an individual's autonomy
- **Big Question:** Is there a point at which a line can be crossed where AI reduces autonomy
- **Synthesis:** For there to be an ethical use of AI, it must respects an individual's freedom while still being able to provide value for them