Case Studies

Moral Agency and Moral Patiency in AI

The following cases are designed to test whether you can **separate moral agency from moral patiency**, even when an AI system behaves in complex or autonomous ways.

Each group will be assigned to one case study. See if you can answer the following questions:

1. Who (if anyone) is a **moral agent**?
2. Who (if anyone) is a **moral patient**?
3. Where does **moral responsibility** ultimately lie?

And follow that up with write a short paragraph answering:

> How does separating moral agency from moral patiency clarify where moral responsibility lies — and what goes wrong if we fail to keep them distinct?

Case Study 1: Autonomous Vehicle Collision

An autonomous vehicle is driving in self-driving mode when it strikes a pedestrian.
The system detected the pedestrian late due to unusual lighting conditions and chose an evasive maneuver that minimized damage to the vehicle but resulted in injury to the pedestrian.

Questions

- Is the AI system a **moral agent**?
- Is the AI system a **moral patient**?
- Who, if anyone, should be morally blamed?
- Does the fact that the system "made a choice" imply moral responsibility?

Case Study 2: Medical Triage Algorithm

A hospital uses an AI system to help prioritize patients for limited ICU beds.
The system consistently ranks certain demographic groups lower due to patterns in historical training data. A patient receives delayed care and suffers serious harm.

Questions

- Is the AI system morally responsible for the harm?
- Who are the moral patients in this scenario?
- Can a system cause moral harm without being a moral agent?
- How does responsibility differ between *design*, *deployment*, and *use*?

Case Study 3: AI Chatbot and Emotional Attachment

A conversational AI chatbot provides companionship to users.
A user forms a deep emotional attachment and reports distress when the system is shut down or modified.

Questions

- Is the AI a moral patient?
- Is the user a moral patient?
- Does causing emotional distress automatically imply moral patiency?
- Are there moral duties *about* AI even if AI has no moral status?

Case Study 4: AI Used in Hiring Decisions

An AI screening tool ranks job applicants and filters out candidates before any human review.
A qualified applicant is rejected due to features correlated with protected characteristics.

Questions

- Can an AI be unfair without being morally responsible?
- Who bears moral responsibility for discriminatory outcomes?
- Is fairness a property of agents, systems, or institutions?