**NAME:** Jude Dwight Oscar M. Jimenez **SUBJECT:** GEC 18 - Ethics

**YEAR/COURSE:** BSCS 3A **PROFESSOR:** Mary Antoniette S. Ariño

**Case Study #1 – Moral Agent**

**Case Study: The Ethical Dilemma of a Healthcare AI System**

**Scenario:**

A hospital implements an AI system to assist doctors in making decisions for criticalpatient care. One day, the AI is faced with a decision regarding two patients who both needthe same rare organ transplant:

**Patient A:** A 25-year-old teacher with a long-life expectancy and no other health

complications.

**Patient B:** A 60-year-old scientist working on a breakthrough in cancer research,

whose work could save thousands of lives but has a shorter life expectancy due to

age.

The AI system, programmed to prioritize "maximizing the quality of life," must decide

who will receive the organ. Doctors have agreed to follow the AI’s recommendation due to

its advanced predictive capabilities.

**Questions for Analysis:**

1. Who is the moral agent in this scenario, and why?
2. What ethical principles should the AI prioritize: saving the younger patient or the one with potential societal impact?
3. How would a utilitarian framework address this situation?
4. How would a deontological approach influence the AI's decision?
5. What role do the programmers play in shaping the moral decisions of the AI?
6. Should the AI consider societal contributions when making medical decisions? Why or why not?
7. What safeguards should be in place to ensure the AI’s decision aligns with human ethics?
8. If the AI makes a controversial decision, who should be held accountable?
9. How might biases in the AI’s training data influence its decision-making process?
10. Would you trust an AI system to make life-and-death decisions? Why or why not?

**Answer:**

1. The **doctors and the hospital administration** are the true moral agents, AI can’t be a moral agent. AI is merely a tool following programmed rules and data. The **programmers** who designed the AI also bear some moral responsibility because they set the ethical framework and algorithms that guide its decision-making.
2. I think the AI will prioritize the older scientist because his work can potentially save more lives than the young teacher's. However, the decision will ultimately be guided by how the programmers designed the AI.
3. A utilitarian framework focuses on maximizing overall benefit. In this scenario, the AI might favor the older scientist because his work has the potential to save more lives, thereby producing a greater benefit.
4. A deontological approach focuses on the inherent duty or moral rules rather than outcomes. It might argue that both patients have an equal right to life, so the decision should not be based on societal contributions but instead on fairness, such as random selection or other equitable criteria
5. The programmers play a crucial role by encoding the AI’s ethical framework and priorities. They decide how the AI evaluates "quality of life", societal impact, and fairness. Any biases or limitations in the AI's programming reflect the ethical views and decisions of its creators.
6. Whether or not the AI should consider such factors ultimately depends on the ethical guidelines provided by the programmers and approved by the hospital administration. Yes, Societal contributions could help maximize broader societal benefits and align with utilitarian goals. But the AI should remain a tool to support human decision-making rather than serving as an autonomous decision-maker.

* Regular ethical reviews by diverse committees.
* Transparency in decision-making algorithms.
* Mechanisms for human oversight or appeals against AI recommendations.
* Continuous monitoring for biases in training data and decision outcomes.

1. Since the AI is merely a tool following programmed instructions, accountability rests with the human agents involved. This includes the doctors and hospital administration who rely on the AI’s recommendation, as well as the programmers who set its ethical framework and guidelines. In effect, the decision is a reflection of human choices at every stage.
2. Biases in the training data can lead the AI to favor one outcome over another in ways that might not align with fair or ethical considerations. For example, if the data disproportionately values certain demographics or professional contributions, the AI might consistently lean toward decisions that reflect those biases. Regular data audits and updates are essential to ensure that the AI remains as balanced and unbiased as possible.
3. No, while an AI system can process vast amounts of data and offer valuable insights, trusting it entirely with life-and-death decisions is problematic. AI lacks the human qualities of empathy and moral reasoning, and there is always the risk of hidden biases. Therefore, while AI can assist in decision-making, final responsibility should lie with human professionals who can contextualize and critically assess the recommendations.