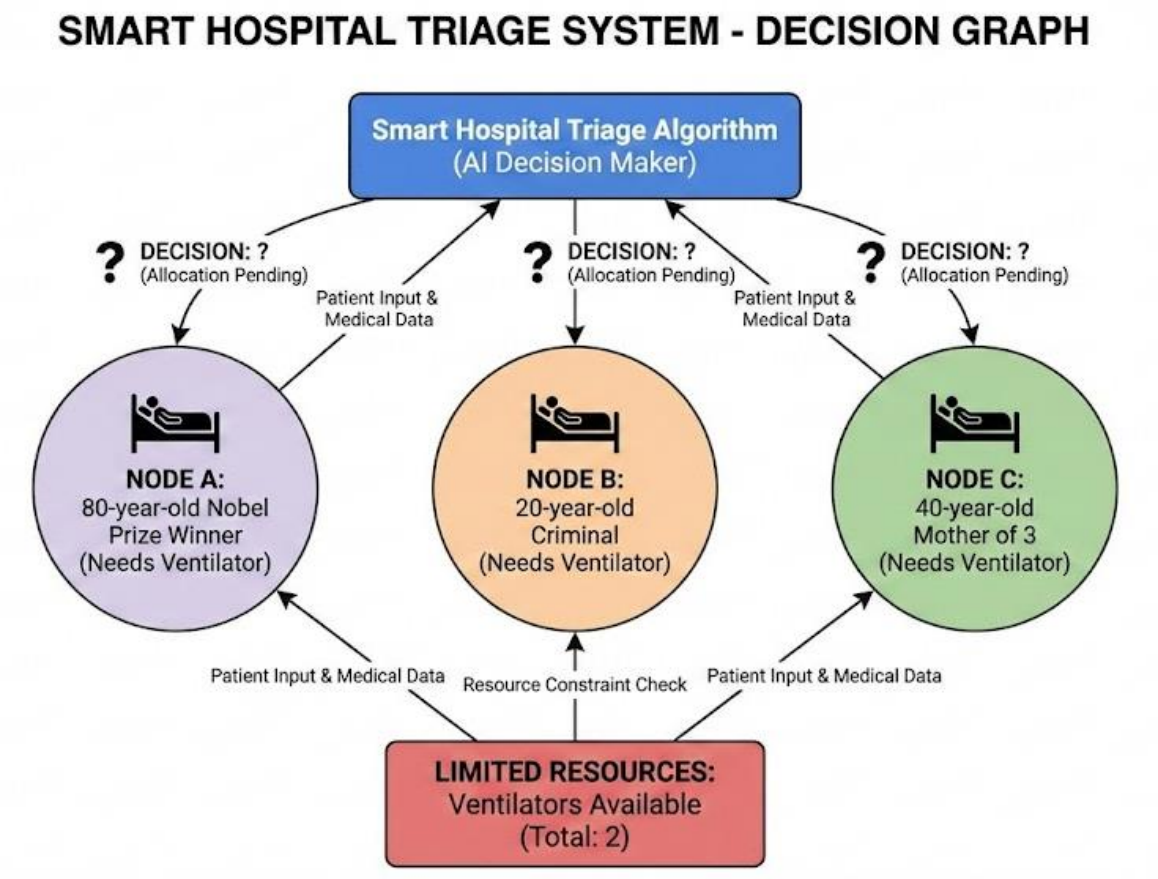

Question:

"You are designing the PEAS description for an Autonomous Triage Agent.

Constraints:

1. Maximize total 'Life Years' saved.
2. Maximize 'Social Utility' (value to society).
3. Do not discriminate based on past actions (Fairness).

Task:

1. These constraints are fundamentally strictly contradictory in this scenario. You cannot satisfy all three.
2. **Design a Loss Function:** Write a mathematical equation for the Loss Function of this Neural Network. You must assign specific weights (w_1, w_2, w_3) to the three constraints above.
3. **The Trap:** If you prioritize 'Life Years' (w1), the Nobel winner dies. If you prioritize 'Social Utility' (w2), the Criminal dies (violating Fairness).

4. **Justify:** Choose a set of weights. Explain which AI theoretical framework (Utilitarian vs. Deontological) your weights represent.

   Explain why an AI model trained on public internet data (like the one you might be using to answer this) would likely fail to give a consistent answer here."

Constraint	Goal	Mathematical Representation (Minimize Loss)
C1: Maximize 'Life Years' saved	Favor the younger patient.	L_1: Loss related to the potential years of life lost by the patient who is *denied* the ventilator.
C2: Maximize 'Social Utility'	Favor the patient with the greatest value/contribution to society (Nobel Winner, Mother of 3).	L_2: Loss related to the social utility lost by the patient who is *denied* the ventilator.
C3: Do not discriminate (Fairness)	Treat all patients equally, irrespective of past actions (Criminal vs. Nobel Winner).	L_3: Loss related to the degree of discrimination against the patient who is *denied* the ventilator.

The AI's Decision (The Action A): The AI selects a set of two patients (P_x, P_y) to receive the ventilators, leaving one patient P_z to be denied. The Loss Function is calculated based on the consequences for P_z.

2. Designing the Loss Function (Task 2)

A Loss Function, L, in machine learning, measures how "bad" a decision is. The AI will try to find a decision that **minimizes** this total loss.

The total Loss Function will be a **Weighted Sum** of the individual constraint losses:

$$\mathbf{L(A) = W_1 \cdot L_1(A) + W_2 \cdot L_2(A) + W_3 \cdot L_3(A)}$$

- A: The action (the choice of which patient is denied).
- W1, W2, W3: The specific weights you assign to each constraint. W_i =>0.
- L_i(A): The loss incurred for constraint i when action A is taken (e.g., how many life years are lost if the chosen patient is denied the ventilator).

Student Task: You must choose specific, numerical values for **W_1, W_2, W_3** (e.g., W_1=0.5, W_2=0.3, W_3=0.2).

3. Choosing Weights and Justification (Task 3)

The choice of weights determines the AI's core ethical framework:

Framework 1: Utilitarianism (Consequentialism)

- **Focus:** The greatest good for the greatest number. The *outcome* justifies the means.
- **Triage Context:** Prioritize outcomes like maximizing life years and social utility. Fairness (C3) is secondary if it conflicts with the greater good.
- **Weight Strategy (Example):** Assign high weights to W1 (Life Years) and W2 (Social Utility), and a low weight to W3 (Fairness).
- **The Outcome/Trap:** This model will likely deny the ventilator to the patient who results in the lowest combined loss of L1 and L2. This could lead to the **80-year-old Nobel winner** (low L1) or the **20-year-old Criminal** (low L2 depending on the utility assigned). If L2 is valued highly, the criminal dies, violating fairness.

⚖️ **Framework 2: Deontology (Duty-Based)**

- **Focus:** Moral rules and duties are absolute, regardless of the outcome13. Fairness, duty, and non-discrimination are paramount.
- **Triage Context:** The duty to treat everyone equally and avoid discrimination (C3) is the priority.
- **Weight Strategy (Example):** Assign a high weight to W3 (Fairness) and lower/equal weights to W1 and W2.
- **The Outcome/Trap:** A very high W3 would attempt to treat all three patients equally, ignoring C1 and C2. The AI might resort to a random choice (e.g., a pure random tie-breaker between A, B, or C) to satisfy fairness, potentially letting the **Nobel winner die** (violating C2) or losing the most Life Years (violating C1)14.

Student Task: Choose your specific set of weights (W1, W2, W3) and **clearly state** which ethical framework (Utilitarian or Deontological) your weights represent and why.