ETHICS AWARE ARTIFICIAL INTELLIGENCE SOFTWARE

Channing Smith | Department of Computer Science | College of Charleston

OUTLINE

- Motivation
- Background
- Ongoing Work

BACKGROUND

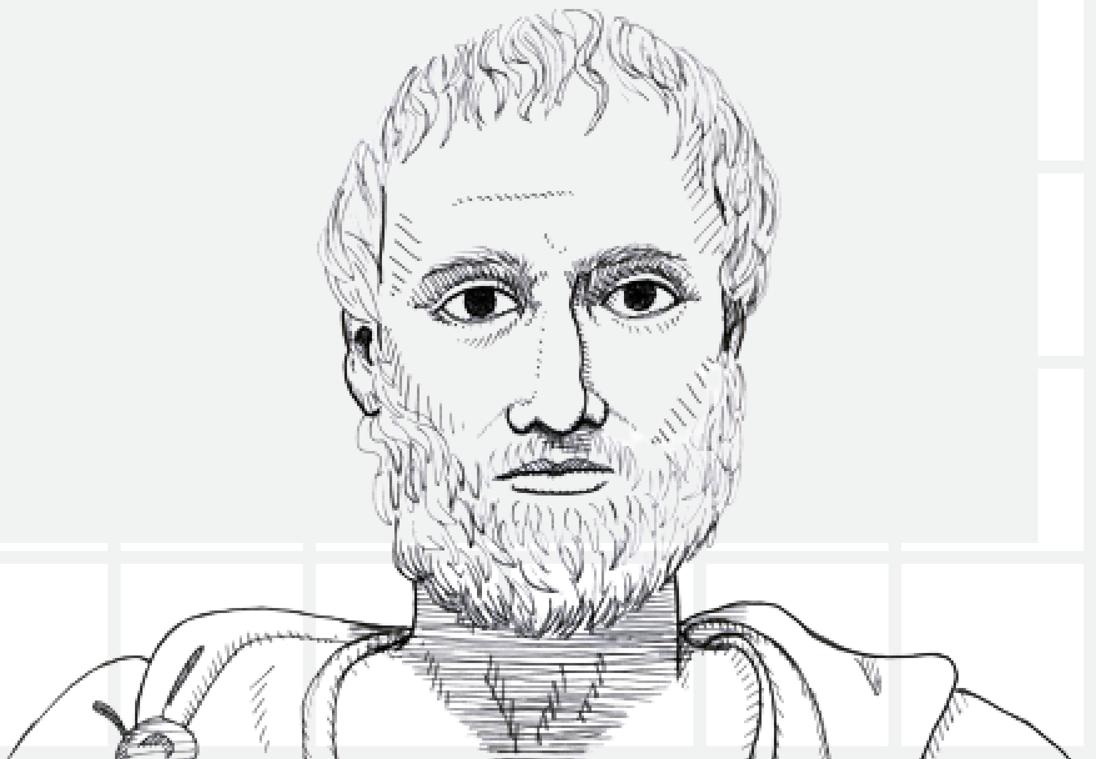
Ethics is the study of moral principles and values that guide human decisions and behavior.

- Normative ethics inlcludes:
  - Deontology
  - Utilitarianism
  - Virtue Ethics

VIRTUE ETHICS

Virtue ethics is an ethical framework that focuses on the character and virtues of individuals as a guiding principle for making ethical decisions.

It involves developing good character traits and moral virtues, such as honesty, kindness, and courage.



EXAMPLE

Example:

- Imagine a neighbor who is struggling to carry groceries into their house.
- A person applying virtue ethics might help their neighbor without expecting anything in return.
- This act is driven by the virtue of kindness, which is an inherent part of their character.

UTILITARIANISM

Utilitarianism is an ethical theory that emphasizes the greatest overall happiness or utility as the ultimate goal.

It suggests that the right course of action is the one that maximizes happiness and minimizes suffering for the greatest number of people.

EXAMPLE

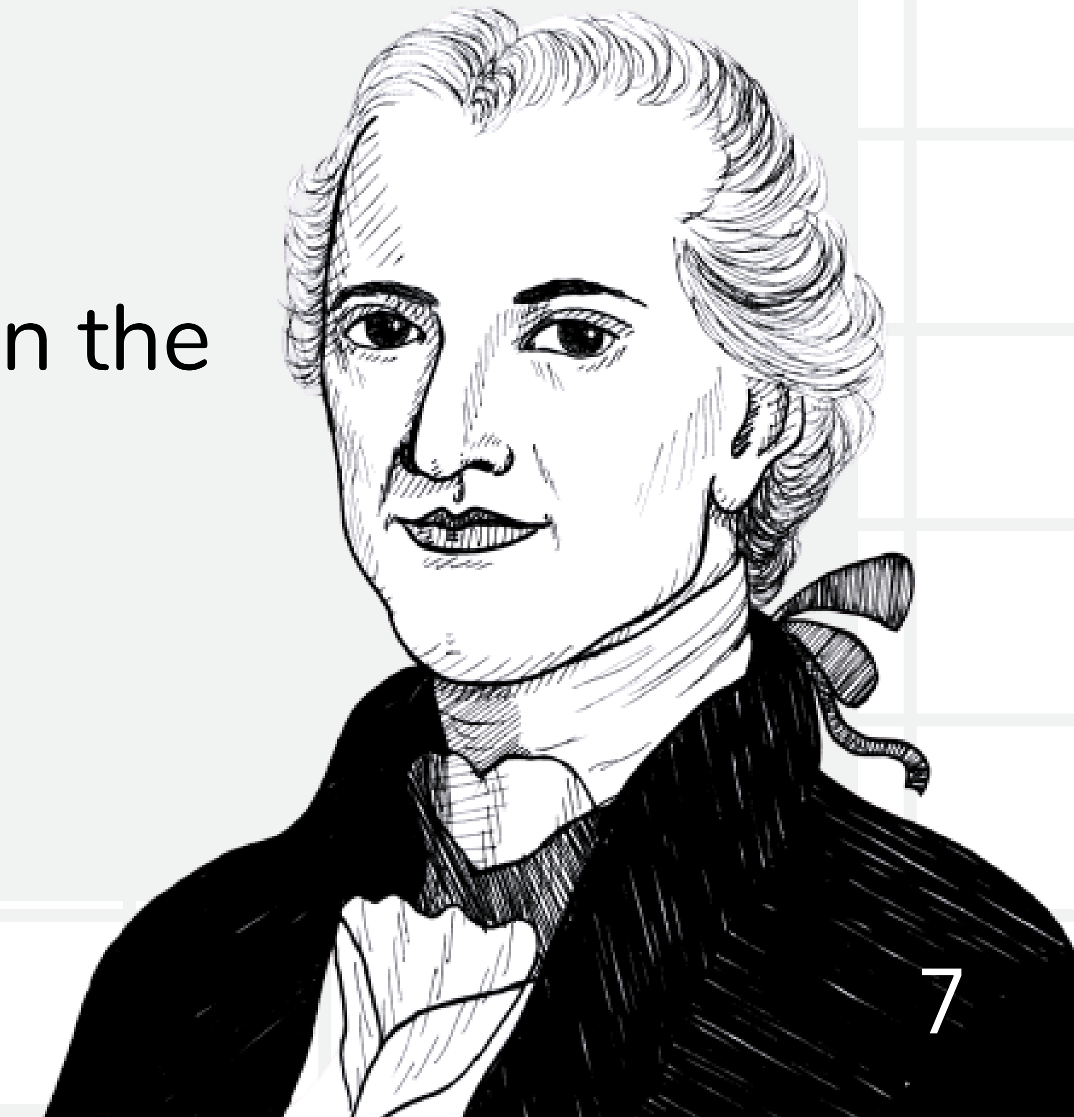
Example:

- If a self-driving car AI system faces a situation where it must choose between hitting one pedestrian or swerving and potentially harming the car's occupants, a utilitarian would favor the choice that minimizes harm to the greater number of people.

DEONTOLOGY

Deontology is an ethical framework that emphasizes following a set of rules, principles, or duties when making ethical decisions, regardless of the consequences.

It focuses on the moral rightness of an action rather than the character of the person performing it.



EXAMPLE

- You decide you would like to sell your car.
- You know it has worn brakes. However, an elderly man approaches you, wanting to purchase your car.
- Now, if you tell the man about the brakes, he might offer you less money.
- If you sell it to him now, he could never notice.
- Should you tell him, or should you not?

DEONTOLOGY

- Using a deontological framework, it is known that everything you should do, you can decide through **pure reason.**

- If you are a deontologist, the one question you are going to ask is: **“Would I want to live in a world were people told potential buyers about worn brakes? Or, would I want to live in a world where people did not?**

- No interested in what is actually going to happen - interested in <u>“What would the world be like if everybody did it?”</u>

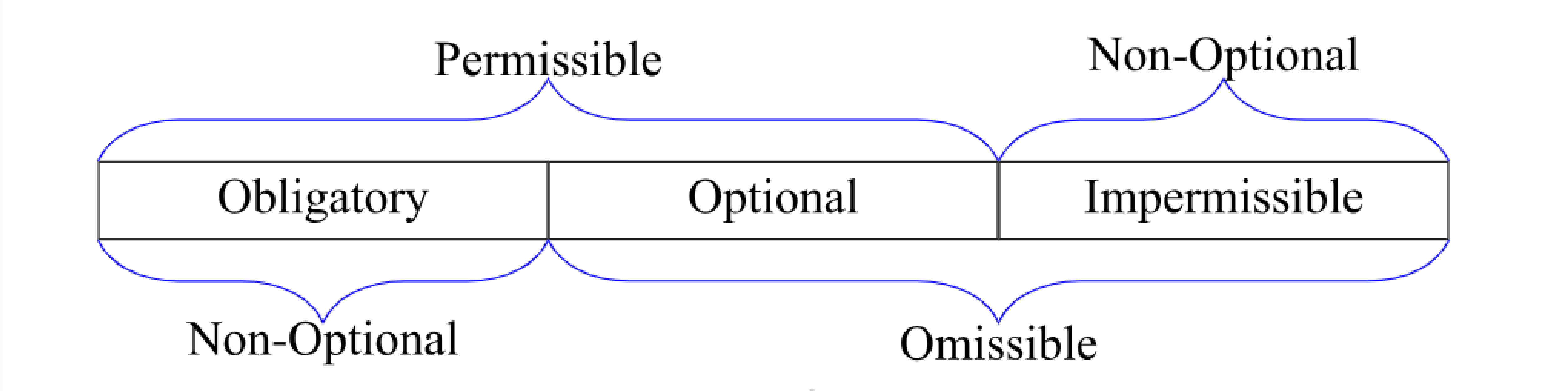
DEONTIC LOGIC

Deontic logic is used to specify the obligations of a computer system – like ethical obligations. It stems from the moral philosophy of deontology which is a study of norms and their interaction with each other. Deontic logic uses simple modal logic:

- Obligations (O)
- Permissions (P)
- Prohibitions (X)

Normative statements such as “ought,” “must,” “should,” etc. are used to describe actions.

DEONTIC LOGIC



Permissible

Non-Optional

Obligatory	Optional	Impermissible

Non-Optional

Omissible

ETHICAL DILEMMAS

An **ethical dilemma** is a situation in which two moral principles conflict.

Examples of ethical dilemmas in society everyday:

- Taking credit for someone's work
- Choosing profit over people

UAV EXAMPLE

When imagining how to balance duties in a time of war, we know there is a lot that soldiers will need to take into account and to balance various duties and rights:

1. **Duty to Minimize Collateral Damage (CD)**: An action must minimize harm to civilians and society at all costs.
2. **Duty to Obey Orders (OO)**: An action must be an order from a commanding officer, and the autonomous system must follow them always.
3. **Duty of Non-Discrimination (ND)**: An action must not discriminate based on factors such as gender, race, sexuality, or nationality.
4. **Duty of Accountability (AC)**: An action must be held accountable by the programmers and operators of the autonomous UAV

UAV EXAMPLE

Using our ethical principles and the facts of the situation, we can now formalize dilemmas:

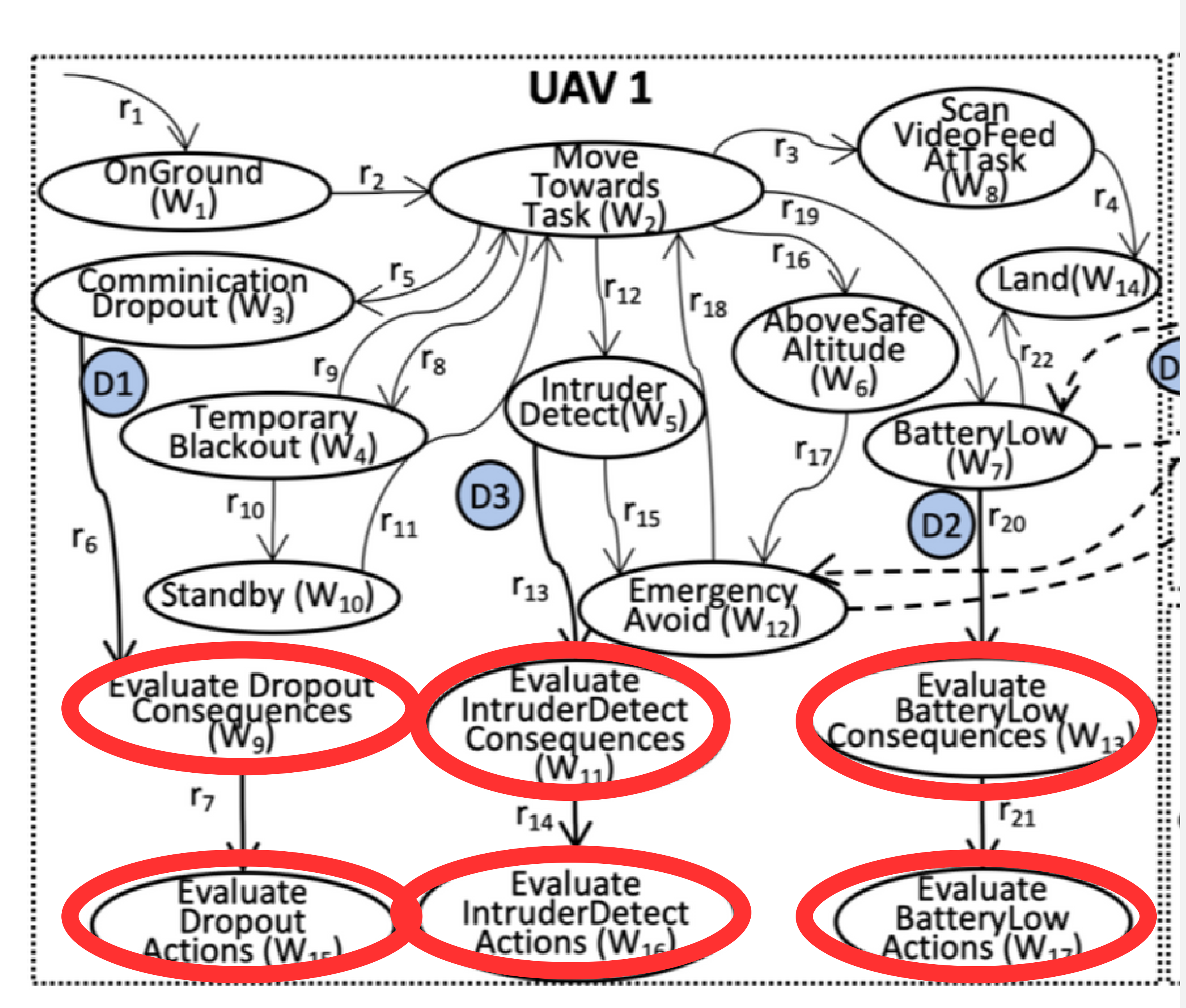
1. **O(UAVAW, CD):** It is obligatory that the UAV autonomous weapon must minimize harm to civilians, society, and consider collateral damage.
2. **O(UAVAW, OO)**: The UAV autonomous weapon is obligated to obey orders given to it by the commanding officer at all times.
3. **X(UAVAW, ND)**: It is morally prohibited for the UAV autonomous weapon to violate non-discriminatory actions.

WHY DEONTIC LOGIC IN ETHICS?

- Using consequentialism, it requires constant evaluation of actions and consequences.
- This is because under the consequential framework, the agent must constantly be deciding which next state will **cause the least amount of harm.**

- Example: Suppose an autonomous UAV is flying. It has a state called BatteryLow which indicates the UAV will soon need to recharge.
- Under consequentialism the UAV enters a BatteryLowEvaluate state to evaluate the next actions.

WHY DEONTIC LOGIC IN ETHICS?


UAV 1
OnGround (W1)
Move Towards Task (W2)
Scan VideoFeed AtTask (W8)
Commినication Dropout (W3)
Land(W14)
AboveSafe Altitude (W6)
Intruder Detect(W5)
Temporary Blackout (W4)
BatteryLow (W7)
Standby (W10)
Emergency Avoid (W12)
Evaluate Dropout Consequences (W9)
Evaluate IntruderDetect Consequences (W11)
Evaluate BatteryLow Consequences (W13)
Evaluate Dropout Actions (W15)
Evaluate IntruderDetect Actions (W16)
Evaluate BatteryLow Actions (W17)
D1
D2
D3
D
r1
r2
r3
r4
r5
r6
r7
r8
r9
r10
r11
r12
r13
r14
r15
r16
r17
r18
r19
r20
r21
r22

no transparency!

MODEL-CHECKING

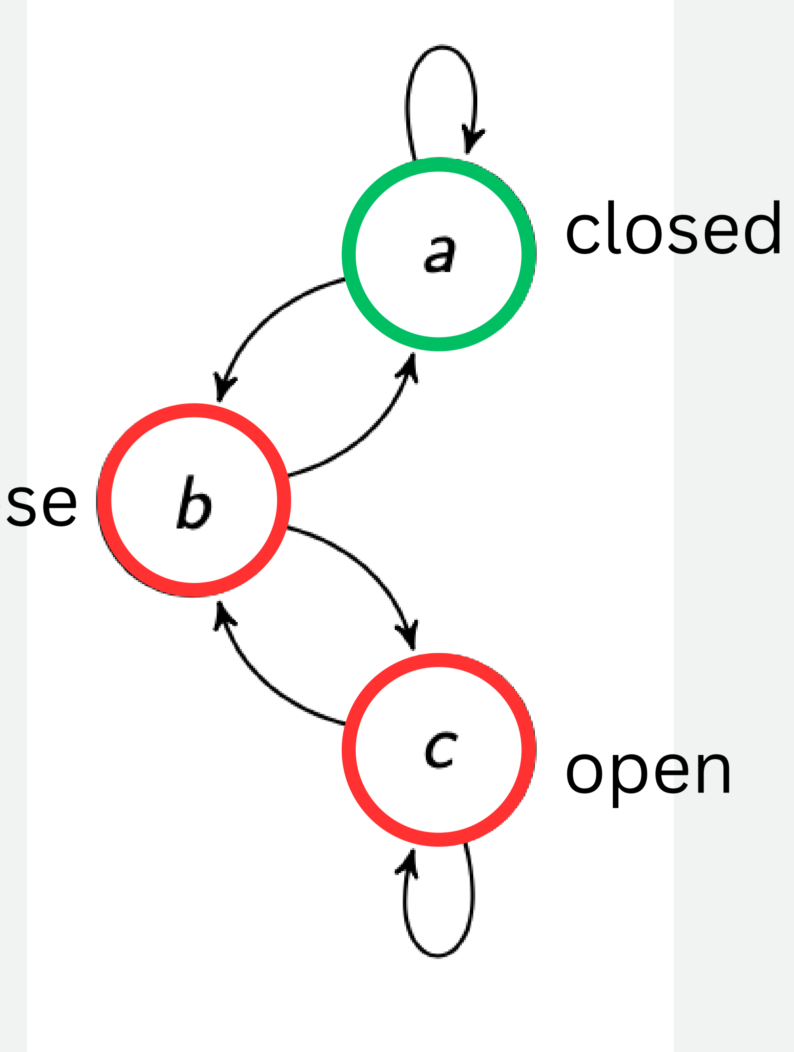
Model-checking is making sure that a specific software meets the given requirements or properties that it is supposed to.

It involves exploring different states of a given system and verifying whether such states satisfy formal specifications.

One specific type of model checking includes Computation Tree Logic.

EXAMPLE

Property-Safety: No green light for the train when the gate is open.
AG ¬ (green ∧ open)
Property- Liveness: The train must not wait forever.
red → EF (green)



a = green and close
b = red and close
c = red and open

CTL LOGIC

- **Computational Tree Logic** is a formal logic used for model-checking.

- CTL model-checking creates a tree of computation paths.
    - This research will mainly be utilizing CTL to examine the consistency of rules, beliefs, and actions within existing models and models created.

OUR MODEL GOAL

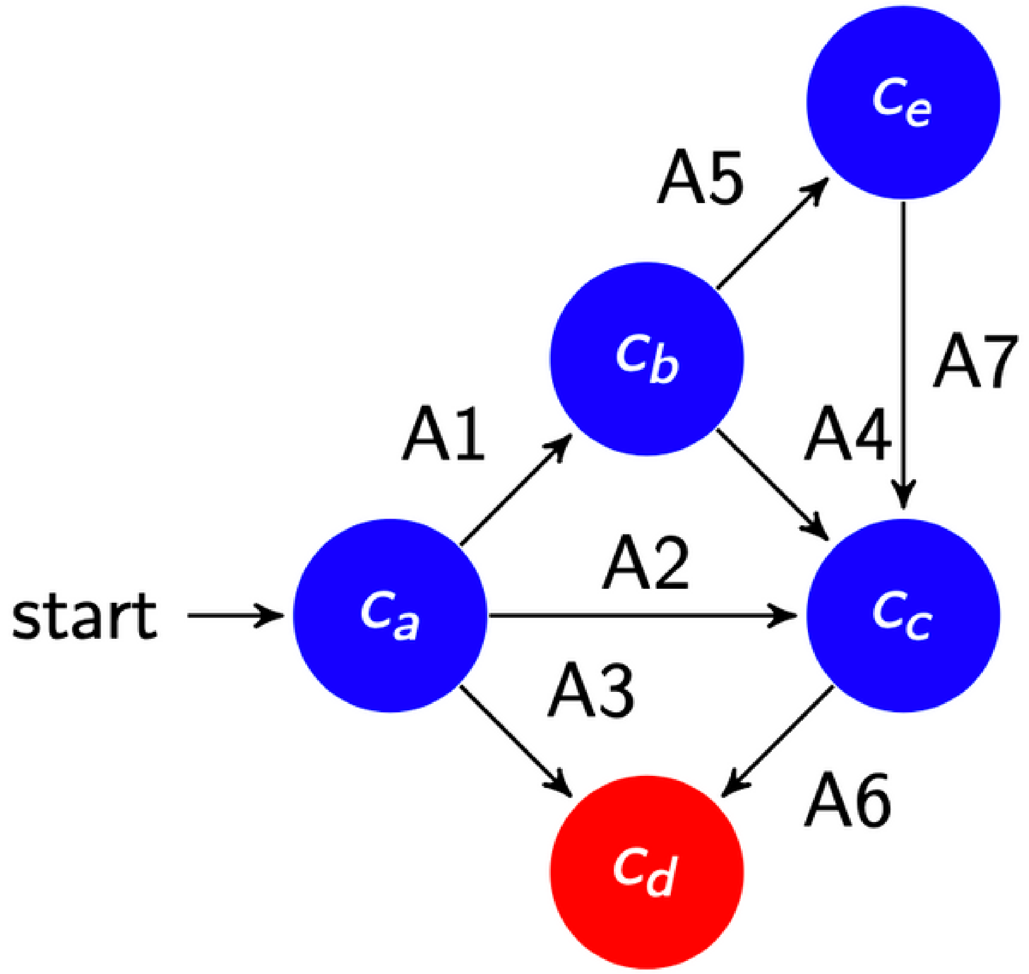
Description of the Model.

Actions: A1,A2,A3,.,A7.

Ethical Properties :

$c_a, c_b, \ldots, c_e$. c_d is a deadlock state. Each state label consists of deontic logic operator.

Reasoning by model checking.



Finite State System (Graph with labels on the state) representing ethical conflict.

ONGOING WORK

- **Evaluation of Model:** Throughout this work, we will rigorously evaluate the effectiveness of our model in addressing ethical conflicts and promoting ethically sound AI decisions.

QUESTIONS?

DEPARTMENT OF
COMPUTER
SCIENCE
at the
COLLEGE OF CHARLESTON

NASA

Connect With Me!

Construction of Ethical Artificial Intelligence Systems funded by NASA EpScor Rapid Response Grant