1a. Human rights

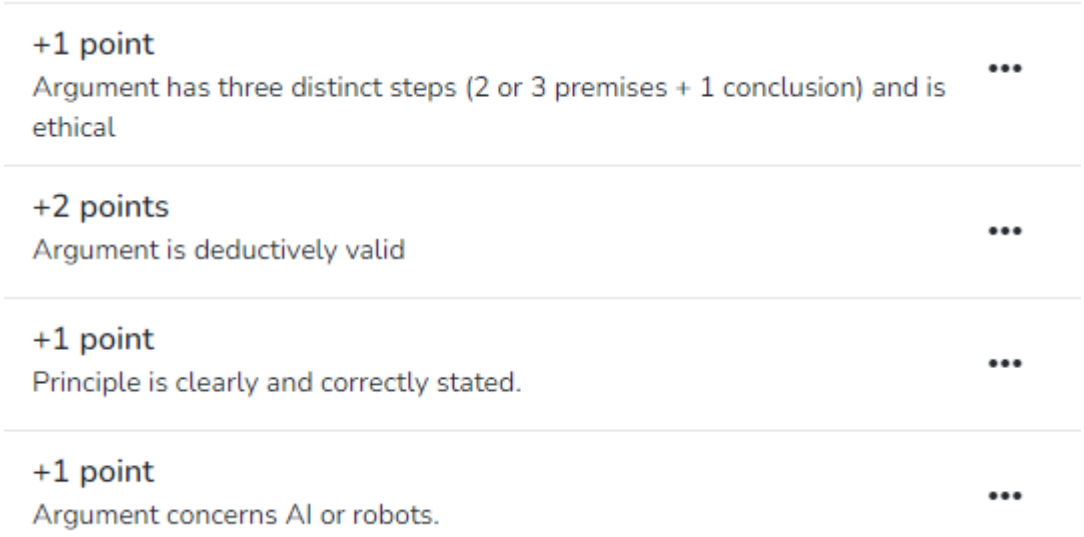
Algorithmic decision making that targets individuals for arrest without making the reasoning explicit, can be considered to violate Art. 9. Black box systems that create risk scores, if used by police to justify arrest or detention, could be considered a form of arbitrary arrest if they are based on arbitrary features (race, sex, location).

Art. 11 is somewhat similar, but with the idea going beyond arrest and covering determinations of criminal violation. The Dutch tax fraud scandal ("toeslagenaffaire") concerned administrative violations, not penal offences. However, we can imagine a comparable case where defendants do not have sufficient access to the reasons of an opaque system used as evidence, and therefore do not have "all the guarantees necessary" to defend themselves.

COMPAS was an example of an opaque system used in sentencing, hence it does not fall strictly under Articles 9 or 11. However, a similar system used as evidence fits the description.

Answers should specify which of the two articles is targeted. The main duty holder of human rights is typically nation states (signatories to the UDHR) and their governments.

1b. Four principles



1c. Goodall

Rule-based ethics and machine learning based on human ethical judgments, because this allows gradual progress. 2 points

1d. HCAI

Camera or one of Shneiderman’s other examples

Elevator or one of Shneiderman’s other examples

1e. Human skills and work

+2 points
Zoller's argument basically correct.

+1 point
Refined understanding of Zoller's arguments

+2 points
Cogent application of "third way" going beyond exam text.

+1 point
Soundness correctly understood

1f. Environmental impact of AI

+1 point
Measuring and reporting climate impact

+1 point
Set standards for sustainable design of AI

+1 point
Clean energy for data centers

+1 point
Policy not feasible or equivalent

+1 point
Policy not desirable or equivalent

+1 point
Explanation

1g. Self-driving cars

+1 point
Validity and right conclusion

+1 point
Simple

+1 point
Ballpark of responsibility condition

+1 point
Specific and precise responsibility condition

1h. explainability

Stakeholders do not have the required understanding of how complex data-driven decisions are made 1pt

People who operate data science systems are secretive about how these systems work 1pt

1i. explainability

Different stakeholders have different needs for knowledge and explanation with regard to algorithmic decision-making. 2 points

1j. Medical AI

+1 point
Interpretable defined in a plausible, non-circular way (e.g., in terms of avoiding the black box problem, or using explainability methods, or as "understandable to a human" from 09-22 lecture)

+1 point
Precise definition compatible with Hatherley et al's rejection of post hoc explanation (e.g., as ex ante interpretability using simple rule based classifiers)

+1 point
Advantage 1 from above list or maybe Zednik

+1 point
Advantage 2 from above list or maybe Zednik

+1 point
Advantage 3 from above list (specific to the clinical perspective as in Hatherley et al/ lecture on Medical AI)

1k. AI in military applications

+1 point
MHC definition in the ballpark (e.g., same as HCAI)

+1 point
TA definition in the ballpark

+1 point
Definitions strongly grounded in the literature

+1 point
Just warfare principle in ballpark

+1 point
Just warfare principle precise and sophisticated

+1 point
Cogent argumentation grounded in literature

1l. Generative AI

Originality

Time and effort invested

1m. Generative AI

True

1n. Generative AI

True