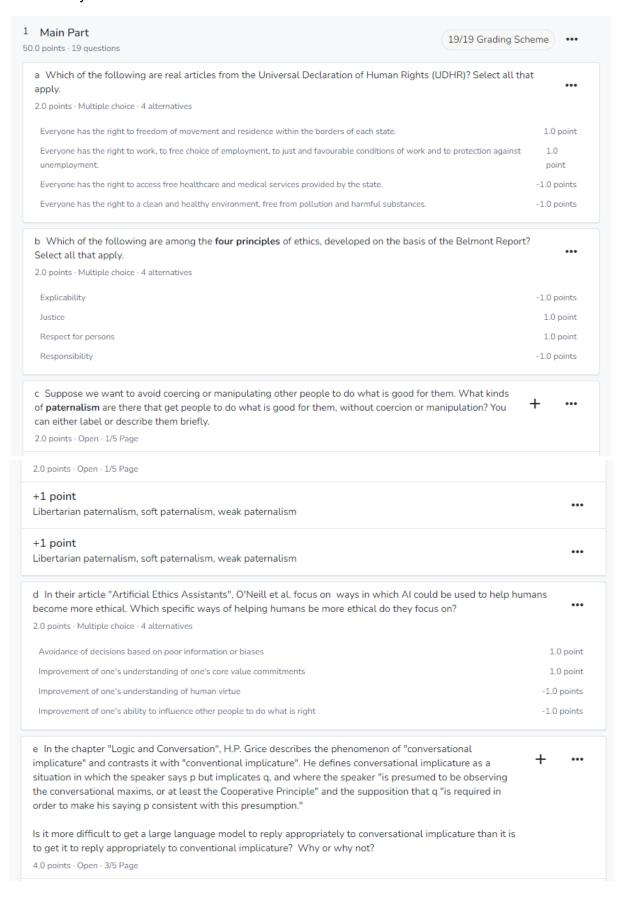
## **Answer Key**



e In the chapter "Logic and Conversation", H.P. Grice describes the phenomenon of "conversational implicature" and contrasts it with "conventional implicature". He defines conversational implicature as a situation in which the speaker says p but implicates q, and where the speaker "is presumed to be observing the conversational maxims, or at least the Cooperative Principle" and the supposition that q "is required in order to make his saying p consistent with this presumption." Is it more difficult to get a large language model to reply appropriately to conversational implicature than it is to get it to reply appropriately to conventional implicature? Why or why not? 4.0 points · Open · 3/5 Page +1 point Def of conversational implicature requires awareness of context to recover speaker's intended meaning through logical inference. (Also acceptable: argument in terms of maxims) +1 point Def. of conventional implicature does not require awareness of context or use of logic, it is a matter of conventional (though not literal) meaning. +1 point LLMs cannot learn context from their source data, and they cannot make appropriate logical inference to respond without the context. +1 point LLMs can learn conventional associations between p & q because these will be represented in source data texts. f According to Goodall's argument, why is it necessary to program ethics into automated vehicles? Select all that apply.

2.0 points  $\cdot$  Multiple choice  $\cdot$  4 alternatives

There is no way to encode human morals into software.

Automated vehicles will be able to avoid all crashes with perfect sensors and algorithms.

Automated vehicles will almost certainly crash, even in ideal conditions.

Automated vehicles can predict various crash trajectory alternatives and select a path with the lowest damage or likelihood of collision.

All students awarded 2 points

g Write a short statement reflecting on whether Awad et al.'s recommended approach to programming moral machines, as discussed in their article "Crowdsourcing Moral Machines," is consistent with Goodall's plan for introducing moral machines. In your response, consider the methodologies and principles each author advocates for. Use the following quote from Goodall's article to help frame your reflection:

"Although artificial intelligence approaches allows computers to loan human othics without the need for

+ ••

"Although artificial intelligence approaches allow computers to learn human ethics without the need for humans to perform the difficult task of articulating ethics as code, they produce actions that cannot be justified or explained in an understandable way."

4.0 points · Open · 1/2 Page

+1 point

Awad et al basic methodology (role of social science in collecting public views specific to a context)

•••

+1 point

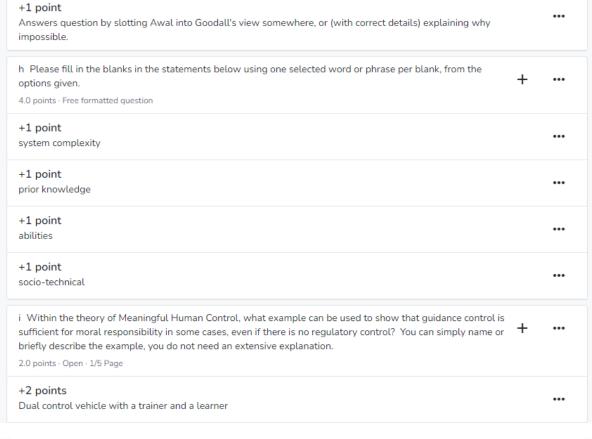
Goodall basic methodology (three stages: logic, ML/AI, XAI)

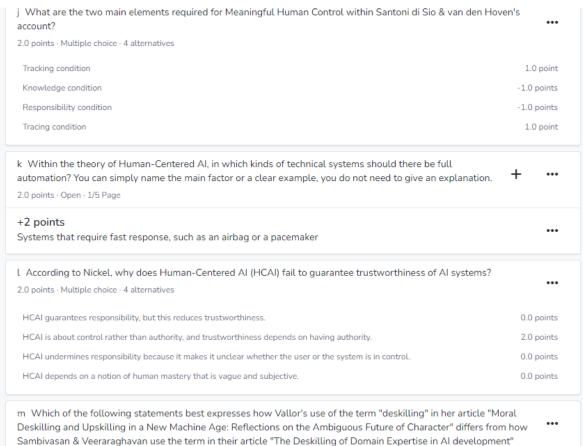
•••

+1 point

Sophisticated understanding of Awal and or Goodall. Goodall: "The neural network could be trained on a combination of simulation and recordings of crashes and near crashes, with human feedback on the ethical response." Awal: "We very much agree that regulations of ethical trade-offs should be left to policy experts, rather than resolved by referendum. But we also believe that policy experts will best serve the public interest when they are well informed about citizens' preferences, regardless of whether they ultimately decide to accommodate these preferences

•••





2.0 points · Multiple choice · 4 alternatives Vallor uses "deskilling" to describe the loss of moral and practical skills due to automation, while Sambasivan & Veeraraghavan use it to 2.0 describe the reduction of domain experts to mere data collectors in AI development. Vallor uses "deskilling" to highlight the economic devaluation of traditional skills, whereas Sambasivan & Veeraraghavan use it to 0.0 emphasize the lack of technical skills needed in AI development. Vallor uses "deskilling" to argue for the complete elimination of human skills by machines, while Sambasivan & Veeraraghavan use it to 0.0 discuss the reduced need for expensive, expert employees in automated data collection practices. Vallor uses "deskilling" to focus on the enhancement of moral virtues brought about through technological advancements, whereas Sambasivan & Veeraraghavan use it to criticize the ethical implications of AI development. n Consider the following incomplete argument: 1) Skill displacement will occur if the technological autonomy ideal is pursued. \_[MISSING PREMISE]\_ 3) We should avoid significant loss of value within lives, unless it is offset by comparative benefits. 4) We can achieve similar benefits by pursuing human-centered AI as by pursuing the technological autonomy ideal, but without skill displacement. 5) Therefore, we should not pursue the technological autonomy ideal. What proposition should premise (2) specify in order to make this argument deductively valid? (A single sentence is a sufficient answer.) 2.0 points · Open · 1/5 Page +2 points Skill displacement involves significant loss of value within lives. +1 point Skill is a value

