WHAT IS CLAIMED IS:

1. A method comprising:

obtaining, using at least one processor, information associated with a current scenario to be evaluated by a machine learning/artificial intelligence (ML/AI) algorithm, the information comprising an initial reward function associated with the current scenario, wherein the information associated with the current scenario comprises a first Markov decision process including a first set of possible states and a first set of possible actions, the initial reward function defining rewards for transitions between the possible states due to the possible actions;

identifying, using the at least one processor, one or more policies associated with one or more prior scenarios that are analogous to the current scenario, wherein each of the one or more policies is a policy for selecting an action for the respective prior scenario and each of the one or more policies is associated with a respective reward function and comprises parameters that are optimized by the respective reward function;

accessing, using the at least one processor, a database configured to store one or more second Markov decision processes associated with the one or more prior scenarios, each of the one or more second Markov decision processes including a second set of possible states and a second set of possible actions, each of the one or more prior scenarios satisfying at least one statistical correlation criteria to the current scenario;

determining, using the at least one processor, one or more differences between parameters that are optimized by the initial reward function and the respective parameters that are optimized by the respective reward function of the one or more prior scenarios;

modifying, using the at least one processor, the initial reward function based on at least one of the one or more determined differences to generate a new reward function, wherein the new reward function comprises at least one new parameter from the respective reward function of the one or more prior scenarios, wherein the at least one new parameter is not included in the initial reward function; and

generating, using the at least one processor, a new policy for the current scenario based on the new reward function.

2. The method of Claim 1, further comprising:

applying, using the at least one processor, the new policy to the current scenario using the ML/AI algorithm in order to determine a selected course of action for the current scenario.

3. The method of Claim 1, further comprising:

repeating, using the at least one processor, at least some of the obtaining, identifying, determining, modifying, and generating operations using the new reward function as the initial reward function.

4. The method of Claim 1, wherein determining the one or more differences between the parameters that are optimized comprises the at least one processor using Inverse Reinforcement Learning to identify reasoning used in the one or more prior scenarios to be applied to the current scenario.

5. The method of Claim 1, wherein:

the at least one processor performs the obtaining, identifying, determining, modifying, and generating operations using multiple agents; and

the method further comprises:

comparing, using the at least one processor, the new policies generated by the multiple agents for conformance; and

modifying, using the at least one processor, in response to the new policies generated by the multiple agents not conforming, the initial reward function used by at least one of the multiple agents and repeating at least some of the obtaining, identifying, determining, modifying, and generating operations.

6. The method of Claim 1, wherein each of the initial reward function and the new reward function comprises:

a task-agnostic portion configured to enforce one or more ethical boundaries regardless of task; and

a task-dependent portion configured to drive contextual behavior of the associated reward function.

7. An apparatus comprising:

at least one memory configured to store information associated with a current scenario to be evaluated by a machine learning/artificial intelligence (ML/AI) algorithm, the information comprising an initial reward function associated with the current scenario, wherein the information associated with the current scenario comprises a first Markov decision process including a first set of possible states and a first set of possible actions, the initial reward function defining rewards for transitions between the possible states due to the possible actions; and

at least one processor configured to:

identify one or more policies associated with one or more prior scenarios that are analogous to the current scenario, wherein each of the one or more policies is a policy for selecting an action for the respective prior scenario and each of the one or more policies is associated with a respective reward function and comprises parameters that are optimized by the respective reward function;

access a database configured to store one or more second Markov decision processes associated with the one or more prior scenarios, each of the one or more second Markov decision processes including a second set of possible states and a second set of possible actions, each of the one or more prior scenarios satisfying at least one statistical correlation criteria to the current scenario;

determine one or more differences between parameters that are optimized by the initial reward function and the respective parameters that are optimized by the respective reward function of the one or more prior scenarios;

modify the initial reward function based on at least one of the one or more determined differences to generate a new reward function, wherein the new reward function comprises at least one new parameter from the respective reward function of the one or more prior scenarios, wherein the at least one new parameter is not included in the initial reward function; and

generate a new policy for the current scenario based on the new reward function.

8. The apparatus of Claim 7, wherein the at least one processor is further configured to apply the new policy to the current scenario using the ML/AI algorithm in order to determine a selected course of action for the current scenario.

9. The apparatus of Claim 7, wherein the at least one processor is further configured to use the new reward function as the initial reward function and repeat at least some of the identify, determine, modify, and generate operations.

10. The apparatus of Claim 7, wherein, to determine the one or more differences between the parameters that are optimized, the at least one processor is configured to use Inverse Reinforcement Learning to identify reasoning used in the one or more prior scenarios to be applied to the current scenario.

11. The apparatus of Claim 7, wherein the at least one processor is further configured to:

perform the identify, determine, modify, and generate operations using multiple agents;

compare the new policies generated by the multiple agents for conformance; and

in response to the new policies generated by the multiple agents not conforming, modify the initial reward function used by at least one of the multiple agents and repeat at least some of the obtain, identify, determine, modify, and generate operations.

12. The apparatus of Claim 7, wherein each of the initial reward function and the new reward function comprises:

a task-agnostic portion configured to enforce one or more ethical boundaries regardless of task; and

a task-dependent portion configured to drive contextual behavior of the associated reward function.

13. A non-transitory computer readable medium containing instructions that when executed cause at least one processor to perform the method of any of Claims 1-6.