

CLAIMS

1. A computer-implemented method, comprising:

determining, by a deep reinforcement learning system, a cost of a resource and a delivery
5 method of the resource, wherein the deep reinforcement learning system comprises a smart agent
that utilizes one or more computer-implemented neural network algorithms and a computer-
implemented execution environment communicably coupled to the smart agent, and wherein the
delivery method comprises releasing a resource package, and the resource package comprises at
least one of a monetary gift, a coupon, or interest-free installments;

10 determining, by the computer-implemented execution environment, an execution result
generated in response to a user responding to the resource and the delivery method of the
resource, wherein the resource and the delivery method of the resource is determined by the
smart agent;

obtaining, by the computer-implemented execution environment, a targeted effect
15 corresponding to the resource and the delivery method of the resource determined by the smart
agent;

constructing, by the computer-implemented execution environment and based on the
execution result, a first function that is positively correlated to a match rate between the
execution result and the targeted effect and negatively correlated to the cost;

20 determining, by the computer-implemented execution environment, a reward score of
reinforcement learning based on the first function;

returning, by the computer-implemented execution environment, the reward score to the
smart agent;

determining, by the smart agent, a strategy of determining the resource and the delivery
25 method of the resource based on the reward score;

based on the strategy and status of the computer-implemented execution environment of
the deep reinforcement learning system, updating, by the smart agent and using a deep learning
algorithm, the resource and the delivery method of the resource and providing the updated
resource and updated delivery method of the resource as feedback to the computer-implemented
30 execution environment; and

sending, by the computer-implemented execution environment, the updated resource based on the updated delivery method of the resource, to a computer platform for delivery to one or more users.

5 2. The computer-implemented method according to claim 1, wherein the cost of the delivery method is determined as a resource share corresponding to the resource package.

 3. The computer-implemented method according to claim 1, wherein the reward score of the deep reinforcement learning system is determined based on a function generated to be
10 negatively correlated to the cost.

 4. The computer-implemented method according to claim 1, wherein the reward score the deep reinforcement learning system is determined by using a score evaluation function, wherein the score evaluation function uses the execution result and the cost as variables, and is positively
15 correlated with a match rate between the execution result and the targeted effect and negatively correlated with the cost.

 5. The computer-implemented method according to claim 1 further comprising:
 obtaining, by the computer-implemented execution environment, a second function
20 generated based on the cost, wherein the second function is positively correlated with the cost;
 and

 wherein determining, by the computer-implemented execution environment, the reward score based on the first function comprises: determining, by the computer-implemented execution environment, the reward score based on at least the first function and the second
25 function.

 6. The computer-implemented method according to claim 5, wherein the reward score is determined based on a ratio or a difference between the first function and the second function.

30 7. The computer-implemented method according to claim 5, wherein the determining the reward score, by the computer-implemented execution environment, based on at least the first

function and the second function comprises:

obtaining, by the computer-implemented execution environment, predetermined mapping relationships between a plurality of reward scores and a plurality of range combinations formed by a plurality of value ranges of the first function and a plurality of value ranges of the second
5 function;

determining, by the computer-implemented execution environment and based on the execution result, a first range that a function value of the first function falls;

determining, by the computer-implemented execution environment and based on the cost, a second range that a function value of the second function falls, wherein the first range and the
10 second range form a range combination; and

determining, by the computer-implemented execution environment, the reward score corresponding to the range combination based on the predetermined mapping relationships.

8. A non-transitory, computer-readable medium storing one or more instructions
15 executable by a computer system to perform operations comprising:

determining, by a deep reinforcement learning system, a cost of a resource and a delivery method of the resource, wherein the deep reinforcement learning system comprises a smart agent that utilizes one or more computer-implemented neural network algorithms and a computer-implemented execution environment communicably coupled to the smart agent, and wherein the
20 delivery method comprises releasing a resource package, and the resource package comprises at least one of a monetary gift, a coupon, or interest-free installments;

determining, by the computer-implemented execution environment, an execution result generated in response to a user responding to the resource and the delivery method of the resource, wherein the resource and the delivery method of the resource is determined by the
25 smart agent;

obtaining, by the computer-implemented execution environment, a targeted effect corresponding to the resource and the delivery method of the resource determined by the smart agent;

constructing, by the computer-implemented execution environment and based on the
30 execution result, a first function that is positively correlated to a match rate between the execution result and the targeted effect and negatively correlated to the cost;

determining, by the computer-implemented execution environment, a reward score of reinforcement learning based on the first function;

returning, by the computer-implemented execution environment, the reward score to the smart agent;

5 determining, by the smart agent, a strategy of determining the resource and the delivery method of the resource based on the reward score; and

 based on the strategy and status of the computer-implemented execution environment of the deep reinforcement learning system, updating, by the smart agent and using a deep learning algorithm, the resource and the delivery method of the resource and providing the update as
10 feedback to the computer-implemented execution environment; and

 sending, by the computer-implemented execution environment, the updated resource based on the updated delivery method of the resource, to a computer platform for delivery to one or more users.

15 9. The non-transitory, computer-readable medium according to claim 8, wherein the cost of the delivery method is determined as a resource share corresponding to the resource package.

 10. The non-transitory, computer-readable medium according to claim 8, wherein the reward score the deep reinforcement learning system is determined based on a function
20 generated to be negatively correlated to the cost.

 11. The non-transitory, computer-readable medium according to claim 8, wherein the reward score of the deep reinforcement learning system is determined by using a score evaluation function, wherein the score evaluation function uses the execution result and the cost
25 as variables, and is positively correlated with a match rate between the execution result and the targeted effect and negatively correlated with the cost.

 12. The non-transitory, computer-readable medium according to claim 8, wherein the operations further comprises:

30 obtaining, by the computer-implemented execution environment, a second function generated based on the cost, wherein the second function is positively correlated with the cost;

and

wherein determining, by the computer-implemented execution environment, the reward score based on the first function comprises: determining, by the computer-implemented execution environment, the reward score based on at least the first function and the second function.

13. The non-transitory, computer-readable medium according to claim 12, wherein the reward score is determined based on a ratio or a difference between the first function and the second function.

14. The non-transitory, computer-readable medium according to claim 12, wherein the determining the reward score, by the computer-implemented execution environment, based on at least the first function and the second function comprises:

obtaining, by the computer-implemented execution environment, predetermined mapping relationships between a plurality of reward scores and a plurality of range combinations formed by a plurality of value ranges of the first function and a plurality of value ranges of the second function;

determining, by the computer-implemented execution environment and based on the execution result, a first range that a function value of the first function falls;

determining, by the computer-implemented execution environment and based on the cost, a second range that a function value of the second function falls, wherein the first range and the second range form a range combination; and

determining, by the computer-implemented execution environment, the reward score corresponding to the range combination based on the predetermined mapping relationships.

15. A computer-implemented system, comprising:

one or more computers; and

one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing one or more instructions that, when executed by the one or more computers, perform one or more operations comprising:

determining, by a deep reinforcement learning system, a cost of a resource and a delivery method of the resource, wherein the deep reinforcement learning system comprises a smart agent that utilizes one or more computer-implemented neural network algorithms and a computer-implemented execution environment communicably coupled to the smart agent, and wherein the delivery method comprises releasing a resource package, and the resource package comprises at least one of a monetary gift, a coupon, or interest-free installments;

determining, by the computer-implemented execution environment, an execution result generated in response to a user responding to the resource and the delivery method of the resource, wherein the resource and the delivery method of the resource is determined by the smart agent;

obtaining, by the computer-implemented execution environment, a targeted effect corresponding to the resource and the delivery method of the resource determined by the smart agent;

constructing, by the computer-implemented execution environment and based on the execution result, a first function that is positively correlated to a match rate between the execution result and the targeted effect and negatively correlated to the cost;

determining, by the computer-implemented execution environment, a reward score of reinforcement learning based on the first function;

returning, by the computer-implemented execution environment, the reward score to the smart agent;

determining, by the smart agent, a strategy of determining the resource and the delivery method of the resource based on the reward score; and

based on the strategy and status of the computer-implemented execution environment of the deep reinforcement learning system, updating, by the smart agent and using a deep learning algorithm, the resource and the delivery method of the resource and providing the update as feedback to the computer-implemented execution environment; and

sending, by the computer-implemented execution environment, the updated resource based on the updated delivery method of the resource, to a computer platform for delivery to one or more users.

16. The computer-implemented system according to claim 15, wherein the cost of the

delivery method is determined as a resource share corresponding to the resource package.

17. The computer-implemented system according to claim 15, wherein the reward score of the deep reinforcement learning system is determined based on a function generated to be
5 negatively correlated to the cost.

18. The computer-implemented system according to claim 15, wherein the reward score of the deep reinforcement learning system is determined by using a score evaluation function, wherein the score evaluation function uses the execution result and the cost as variables, and is
10 positively correlated with a match rate between the execution result and the targeted effect and negatively correlated with the cost.

19. The computer-implemented system according to claim 15, wherein the operations further comprises:
15 obtaining, by the computer-implemented execution environment, a second function generated based on the cost, wherein the second function is positively correlated with the cost; and

wherein determining, by the computer-implemented execution environment, the reward score based on the first function comprises: determining, by the computer-implemented
20 execution environment, the reward score based on at least the first function and the second function.

20. The computer-implemented system according to claim 19, wherein the reward score is determined based on a ratio or a difference between the first function and the second function.
25

21. The computer-implemented system according to claim 19, wherein the determining the reward score based on at least the first function and the second function comprises:

wherein the determining the reward score, by the computer-implemented execution environment, based on at least the first function and the second function comprises:

30 obtaining, by the computer-implemented execution environment, predetermined mapping relationships between a plurality of reward scores and a plurality of range combinations formed

by a plurality of value ranges of the first function and a plurality of value ranges of the second function;

determining, by the computer-implemented execution environment and based on the execution result, a first range that a function value of the first function falls;

5 determining, by the computer-implemented execution environment and based on the cost, a second range that a function value of the second function falls, wherein the first range and the second range form a range combination; and

determining, by the computer-implemented execution environment, the reward score corresponding to the range combination based on the predetermined mapping relationships.

10