

## CLAIMS

### WHAT IS CLAIMED IS:

1. A computer-implemented method for performing marketing cost control, comprising:  
5 determining, by using a deep reinforcement learning system, a cost of a resource and a delivery method of the resource, wherein the deep reinforcement learning system comprises ~~an~~ a smart agent and ~~an~~ that utilizes one or more computer-implemented neural network algorithms and a computer-implemented execution environment, ~~the~~ communicably coupled to the smart agent ~~is configured to determine a marketing behavior;~~
- 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;
- 15 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;
- 20 constructing, by the computer-implemented execution environment and based on a marketing strategy and state information of the execution environment, and the method comprises: 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;
- 25 determining a cost of, by the marketing behavior;
- determining computer-implemented execution environment, a reward score of reinforcement learning based on at least the cost, so that the reward score is negatively correlated with the cost; and the first function;
- returning, by the computer-implemented execution environment, the reward score to the smart agent for;
- determining, by the smart agent to optimize its marketing, a strategy of determining the resource and the delivery 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 execution environment; and

5        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.

10        2. The computer-implemented method according to claim 1, wherein the marketing behavior delivery method comprises releasing a resource package, and the resource package comprises at least one of a red-envelope monetary gift, a coupon, or a bill-interest-free ticket installments; and the determining a cost of the marketing behavior comprises:

determining delivery method is determined as a resource share corresponding to the resource package, and using the resource share as the cost.

15        3. The computer-implemented method according to claim 1, wherein the determining a reward score of the deep reinforcement learning comprises:

obtaining a function constructed for the cost, and determining the reward score system is determined based on at least the function, wherein the a function is constructed as being generated to be negatively correlated with to the cost.

20        4. The computer-implemented method according to claim 1, before wherein the determining a reward score of the deep reinforcement learning, further comprising:

obtaining, from the execution environment, an execution result generated when a user responds to the marketing behavior system is determined by the agent, and a targeting effect corresponding to the marketing behavior.

25        5. The method according to claim 4, wherein the determining a reward score of reinforcement learning comprises:

determining the reward score by using a score evaluation function, wherein the score evaluation function uses the execution result and the cost as variables, and is designed to be positively correlated with a match rate between the execution result and the targeting targeted

effect and negatively correlated with the cost ~~of the marketing behavior~~.

6

5. The computer-implemented method according to claim 4, ~~wherein the determining a reward score of reinforcement learning comprises~~ 1 further comprising:

5 obtaining ~~a first function constructed based on the execution result, wherein the first function is designed to be positively correlated with a match rate between the execution result and the targeting effect;~~

~~obtaining~~, by the computer-implemented execution environment, a second function ~~constructed~~generated based on the cost, wherein the second function is ~~designed to be~~ positively correlated with the cost; and

10 ~~determining 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.

7

6. The computer-implemented method according to claim 6, wherein the ~~determining the reward score~~ based on at least the first function and the second function comprises:

~~determining the reward score is determined~~ based on a ratio ~~of the first function to the second function;~~ or

20 ~~determining the reward score based on~~ a difference between the first function and the second function.

8

7. The computer-implemented method according to claim 6, wherein the determining the reward score, by the computer-implemented execution environment, based on at least the first function and the second function comprises:

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

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

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

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

~~9. An apparatus for performing marketing cost control by using a deep~~

~~8. A non-transitory, computer-readable medium storing one or more instructions 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 is configured to determine a marketing behavior based on a marketing strategy and state information of the execution environment, and the apparatus comprises:~~

~~a first determining unit, configured to determine a cost of the marketing behavior;~~

~~a second 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 unit, configured to determine, by the computer-implemented execution environment, a reward score of reinforcement learning based on at least the cost, so that the reward score is negatively correlated with the cost; and the first function;~~

~~a returning unit, configured to return, by the computer-implemented execution environment, the reward score to the smart agent for;~~

determining, by the smart agent to optimize its marketing, 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.

9. The non-transitory, computer-readable medium ~~10. The apparatus~~ according to claim 98, wherein the ~~marketing behavior~~ delivery method comprises releasing a resource package, and the resource package comprises at least one of a ~~red envelope~~ monetary gift, a coupon, or a ~~bill~~ interest-free ~~ticket~~ installments; and the ~~first determining unit~~ cost of the delivery method is configured to:

~~determined~~ determined as a resource share corresponding to the resource package.

~~10. , and use the resource share as the cost.~~

~~11. The apparatus~~ non-transitory, computer-readable medium according to claim 98, wherein the ~~second determining unit~~ reward score the deep reinforcement learning system is configured to:

~~obtain a function constructed for the cost, and determine the reward score based on at least the function, wherein the function is constructed as being determined based on a function generated to be~~ negatively correlated ~~with~~ to the cost.

~~11. 12. The apparatus~~ non-transitory, computer-readable medium according to claim 98, wherein the ~~first determining unit is further configured to obtain, from~~ reward score of the execution environment, an execution result generated when a user responds to the marketing behavior deep reinforcement learning system is determined by the agent, and a targeting effect corresponding to the marketing behavior.

13. The apparatus according to claim 12, wherein the second determining unit is configured

to:

~~determine the reward score~~ by using a score evaluation function, wherein the score evaluation function uses the execution result and the cost as variables, and is ~~designed to be~~ positively correlated with a match rate between the execution result and the ~~targeting~~targeted effect and negatively correlated with the cost ~~of the marketing behavior~~.

~~12. 14.~~The ~~apparatus~~non-transitory, computer-readable medium according to claim ~~128~~, wherein the ~~second determining unit~~operations further comprises:

~~a first acquisition module, configured to obtain a first function constructed based on the obtaining, by the computer-implemented execution result, wherein the first function is designed to be positively correlated with a match rate between the execution result and the targeting effect;~~

~~a second acquisition module, configured to obtain a second function constructed~~environment, a second function generated based on the cost, wherein the second function is ~~designed to be~~ positively correlated with the cost; and

~~a wherein determining, by the computer-implemented execution environment, the reward score based on the first function comprises: determining module, configured to determine, by the computer-implemented execution environment, the reward score based on at least the first function and the second function.~~

~~15~~

~~13.~~ The ~~apparatus~~non-transitory, computer-readable medium according to claim ~~14~~12, wherein the ~~determining module is configured to:~~

~~determine the~~ reward score is determined based on a ratio ~~of the first function to the second function;~~ or

~~determine the reward score based on~~ a difference between the first function and the second function.

~~14. 16.~~The ~~apparatus~~non-transitory, computer-readable medium according to claim ~~14~~12, wherein the determining ~~module is configured to~~the reward score, by the computer-implemented execution environment, based on at least the first function and the second function comprises:

~~obtain pre-established~~obtaining, by the computer-implemented execution environment,

predetermined mapping relationships between a plurality of reward scores and ~~multiple-a~~ plurality of range combinations formed by ~~multiple-a plurality of~~ value ranges of the first function and ~~multiple-a plurality of~~ value ranges of the second function;

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

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

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

17

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 storage medium/media storing one or more instructions that, when executed by the one or more computers, perform one or more operations comprising:

15 determining, by a deep reinforcement learning system, a cost of a resource and a delivery method of the resource, wherein the ~~computer-readable storage medium stores a deep reinforcement learning system comprises a smart agent that utilizes one or more computer program,~~implemented neural network algorithms and when the computer ~~program runs on a-~~implemented execution environment communicably coupled to the smart agent;

20 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;

25 obtaining, by the computer~~is enabled to perform the method-~~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 ~~any one of claim 15~~, 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; and the cost of the delivery method is determined as a resource share corresponding to the resource package~~claims 1 to 8.~~

~~18. A computing device, comprising a memory and a processor, wherein the memory stores executable code, and when executing the executable code, the processor implements the method according to any one of claims 1 to 8.~~

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 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 positively correlated with a match rate between the execution result and the targeted effect and negatively correlated with the cost.

5        19. The computer-implemented system according to claim 15, wherein the operations further comprises:

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

10       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.

15       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.

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:

20       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;

25       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

30       determining, by the computer-implemented execution environment, the reward score

corresponding to the range combination based on the predetermined mapping relationships.