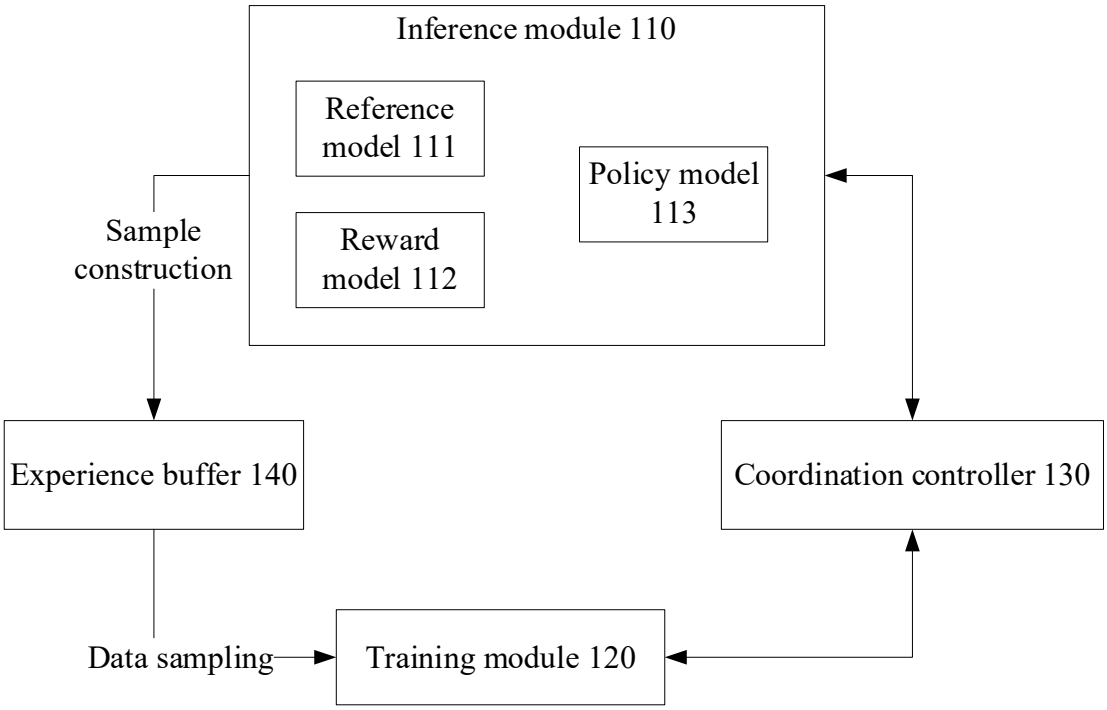

Inference module 110
Reference model 111
Policy model 113
Reward model 112
Sample construction
Experience buffer 140
Coordination controller 130
Data sampling
Training module 120

FIG. 1

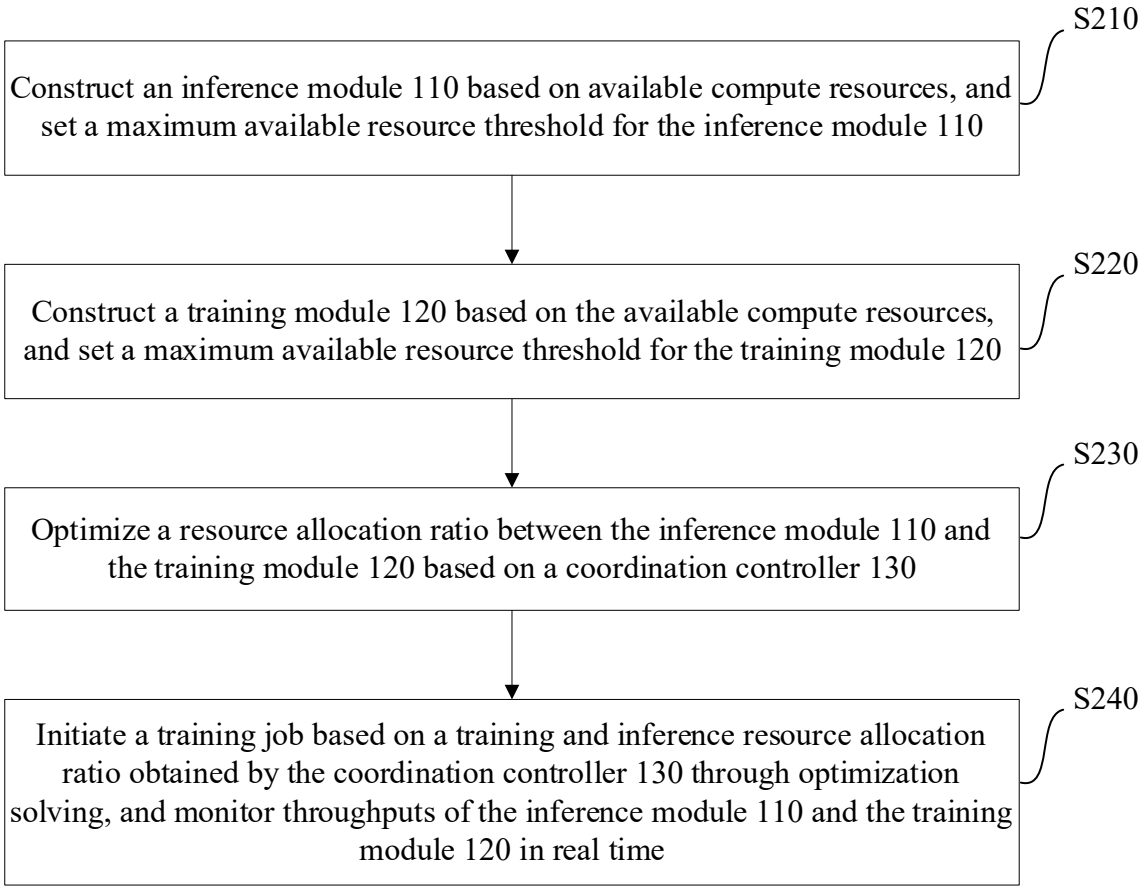

S210
Construct an inference module 110 based on available compute resources, and set a maximum available resource threshold for the inference module 110
S220
Construct a training module 120 based on the available compute resources, and set a maximum available resource threshold for the training module 120
S230
Optimize a resource allocation ratio between the inference module 110 and the training module 120 based on a coordination controller 130
S240
Initiate a training job based on a training and inference resource allocation ratio obtained by the coordination controller 130 through optimization solving, and monitor throughputs of the inference module 110 and the training module 120 in real time

FIG. 2

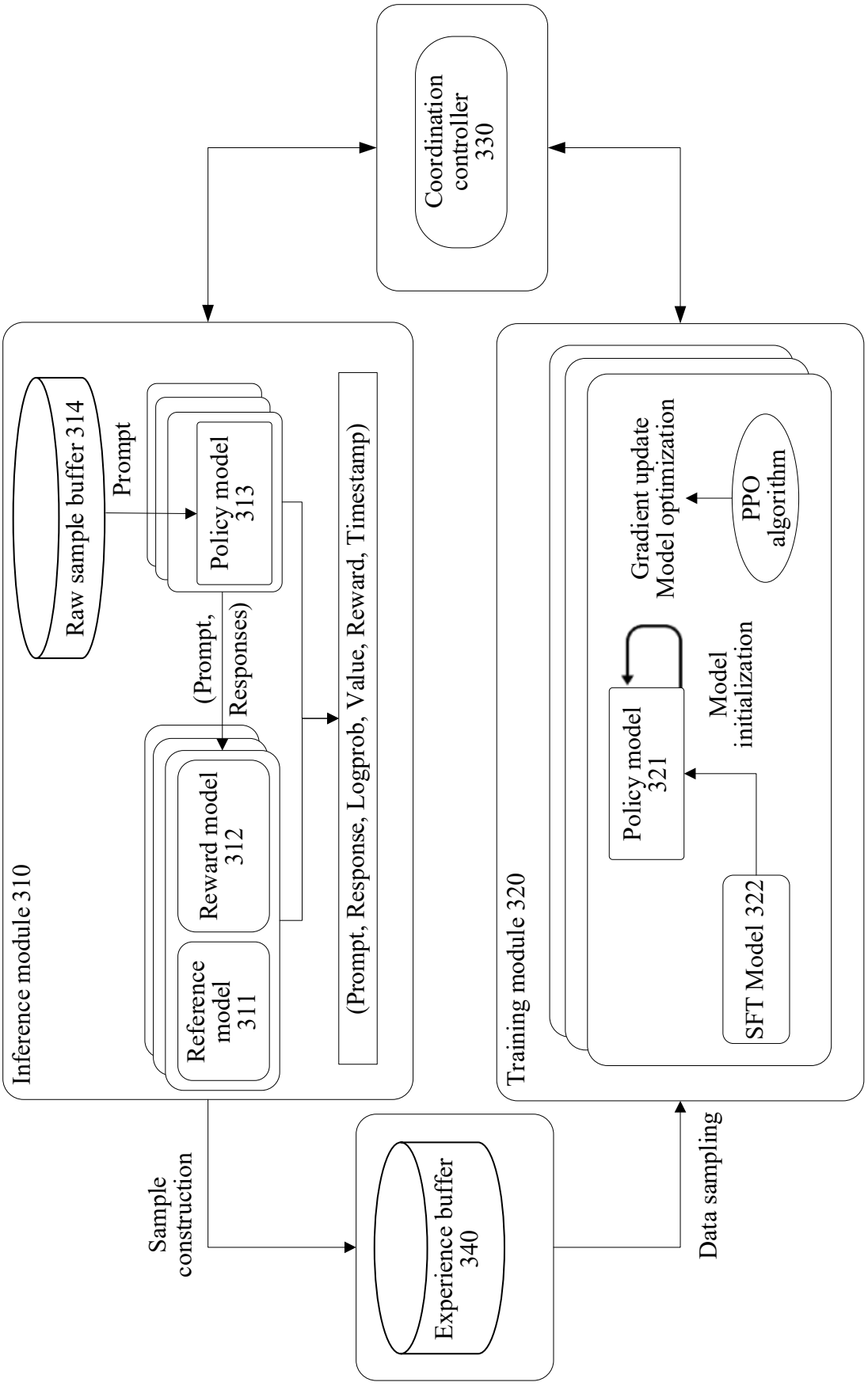

Coordination controller 330
Inference module 310
Raw sample buffer 314
Prompt
Policy model 313
(Prompt, Responses)
Reward model 312
Reference model 311
(Prompt, Response, Logprob, Value, Reward, Timestamp)
Training module 320
Gradient update Model optimization
PPO algorithm
Policy model 321
Model initialization
SFT Model 322
Sample construction
Experience buffer 340
Data sampling

FIG. 3

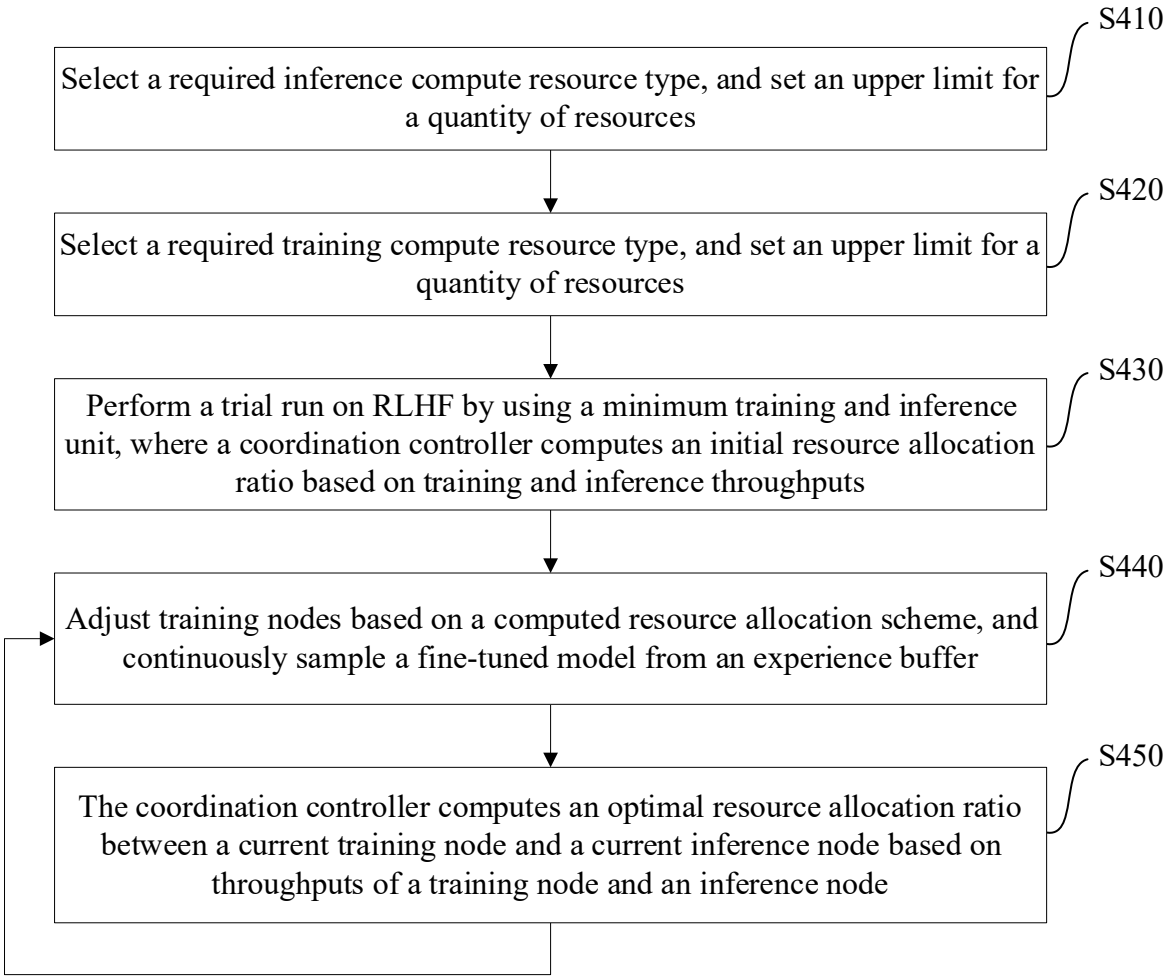

S410
Select a required inference compute resource type, and set an upper limit for a quantity of resources
S420
Select a required training compute resource type, and set an upper limit for a quantity of resources
S430
Perform a trial run on RLHF by using a minimum training and inference unit, where a coordination controller computes an initial resource allocation ratio based on training and inference throughputs
S440
Adjust training nodes based on a computed resource allocation scheme, and continuously sample a fine-tuned model from an experience buffer
S450
The coordination controller computes an optimal resource allocation ratio between a current training node and a current inference node based on throughputs of a training node and an inference node

FIG. 4

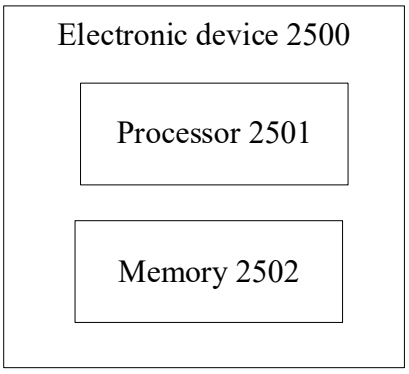

Electronic device 2500
Processor 2501
Memory 2502

FIG. 5