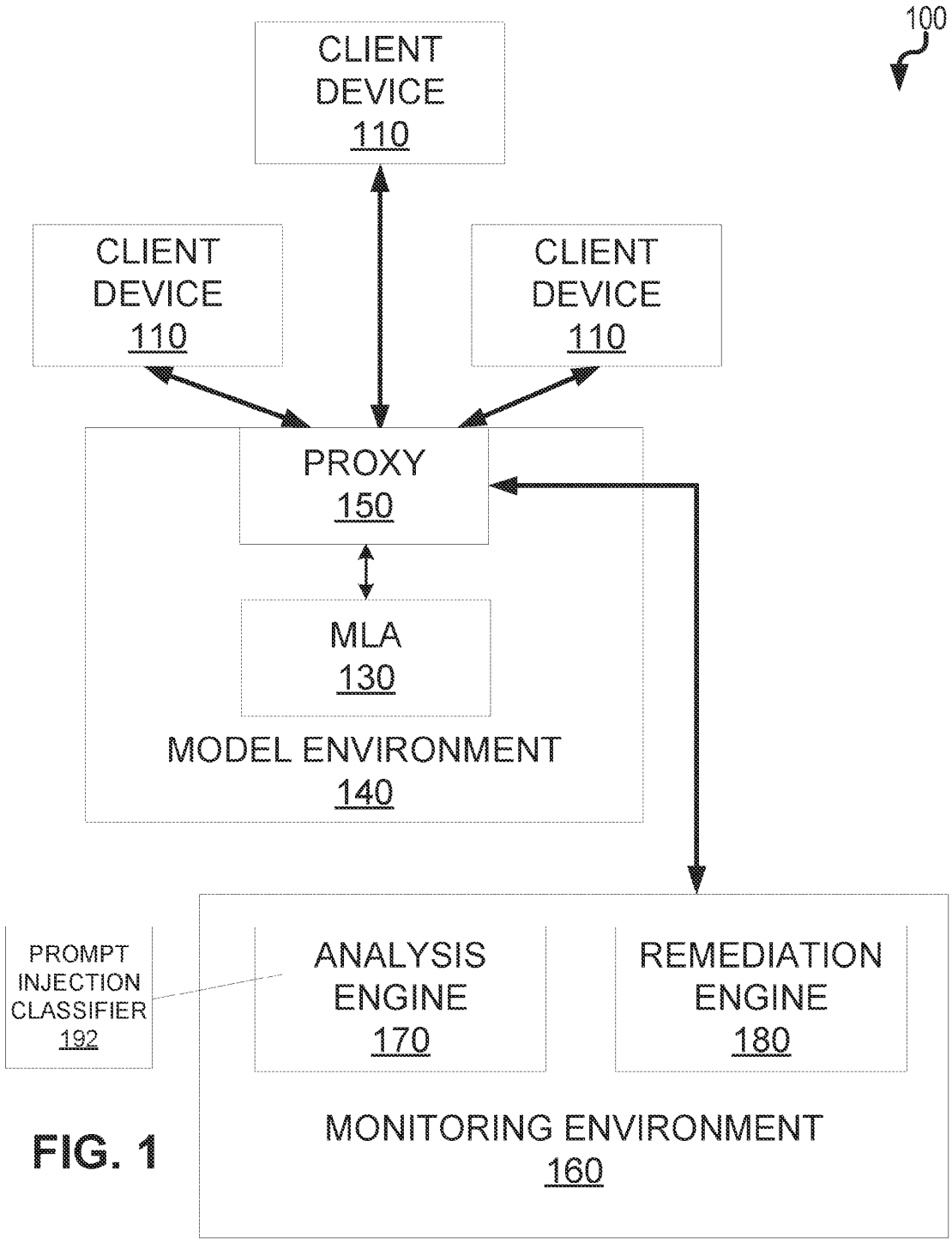

100
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROXY 150
MLA 130
MODEL ENVIRONMENT 140
PROMPT INJECTION CLASSIFIER 192
ANALYSIS ENGINE 170
REMEDIATION ENGINE 180
MONITORING ENVIRONMENT 160

FIG. 1

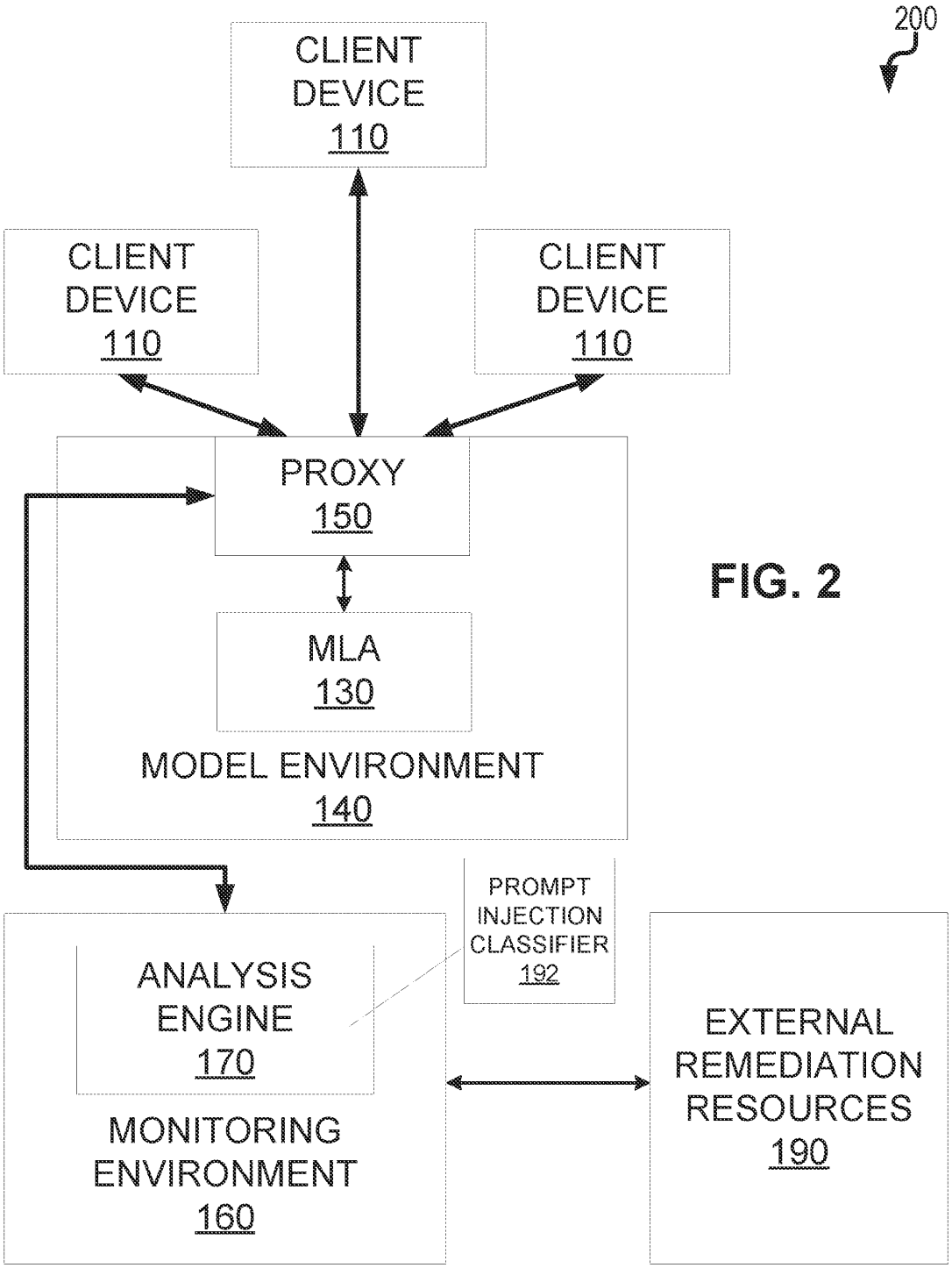

200
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROXY 150
MLA 130
MODEL ENVIRONMENT 140
PROMPT INJECTION CLASSIFIER 192
ANALYSIS ENGINE 170
MONITORING ENVIRONMENT 160
EXTERNAL REMEDIATION RESOURCES 190

FIG. 2

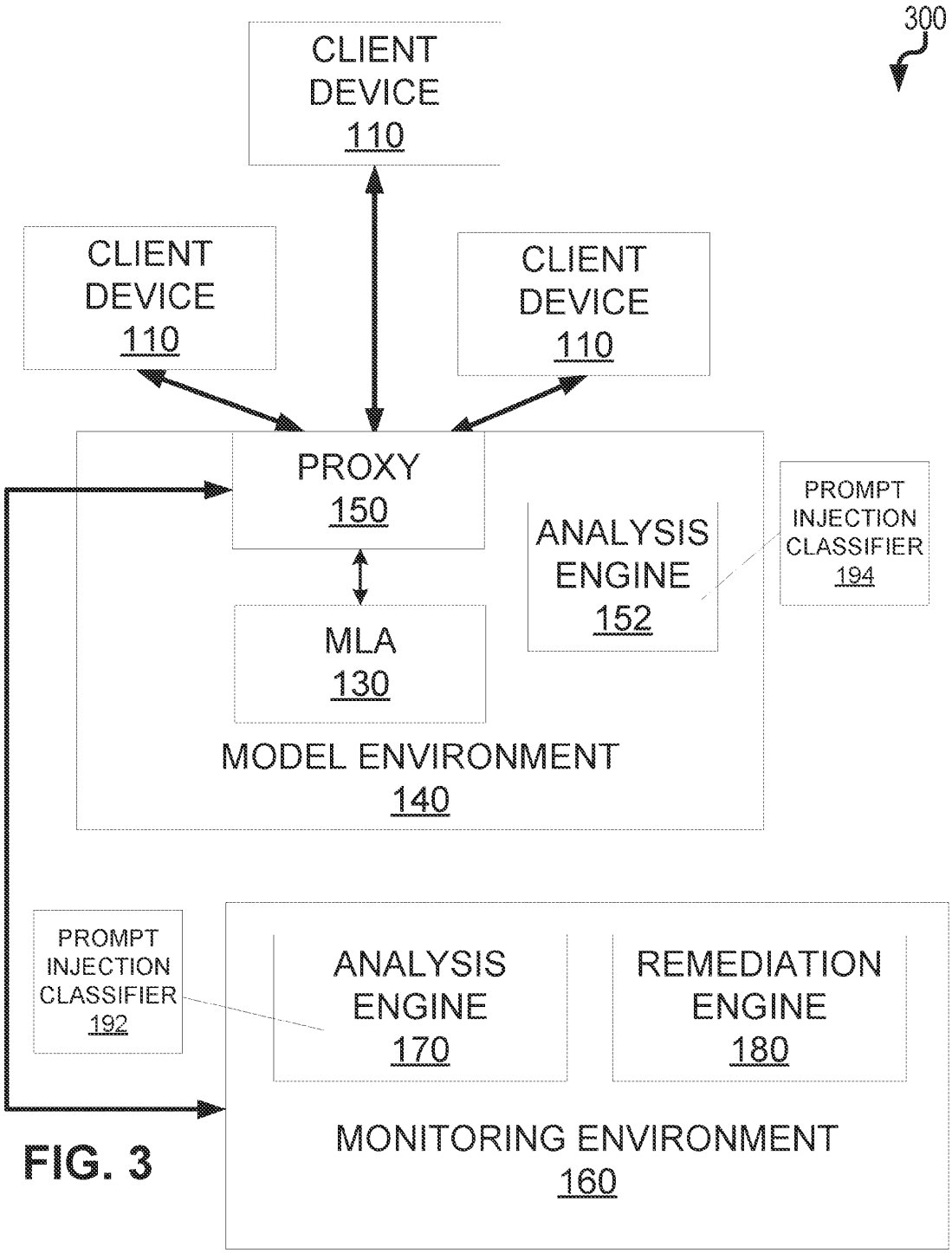

300
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROXY 150
ANALYSIS ENGINE 152
PROMPT INJECTION CLASSIFIER 194
MLA 130
MODEL ENVIRONMENT 140
PROMPT INJECTION CLASSIFIER 192
ANALYSIS ENGINE 170
REMEDIATION ENGINE 180
MONITORING ENVIRONMENT 160

FIG. 3

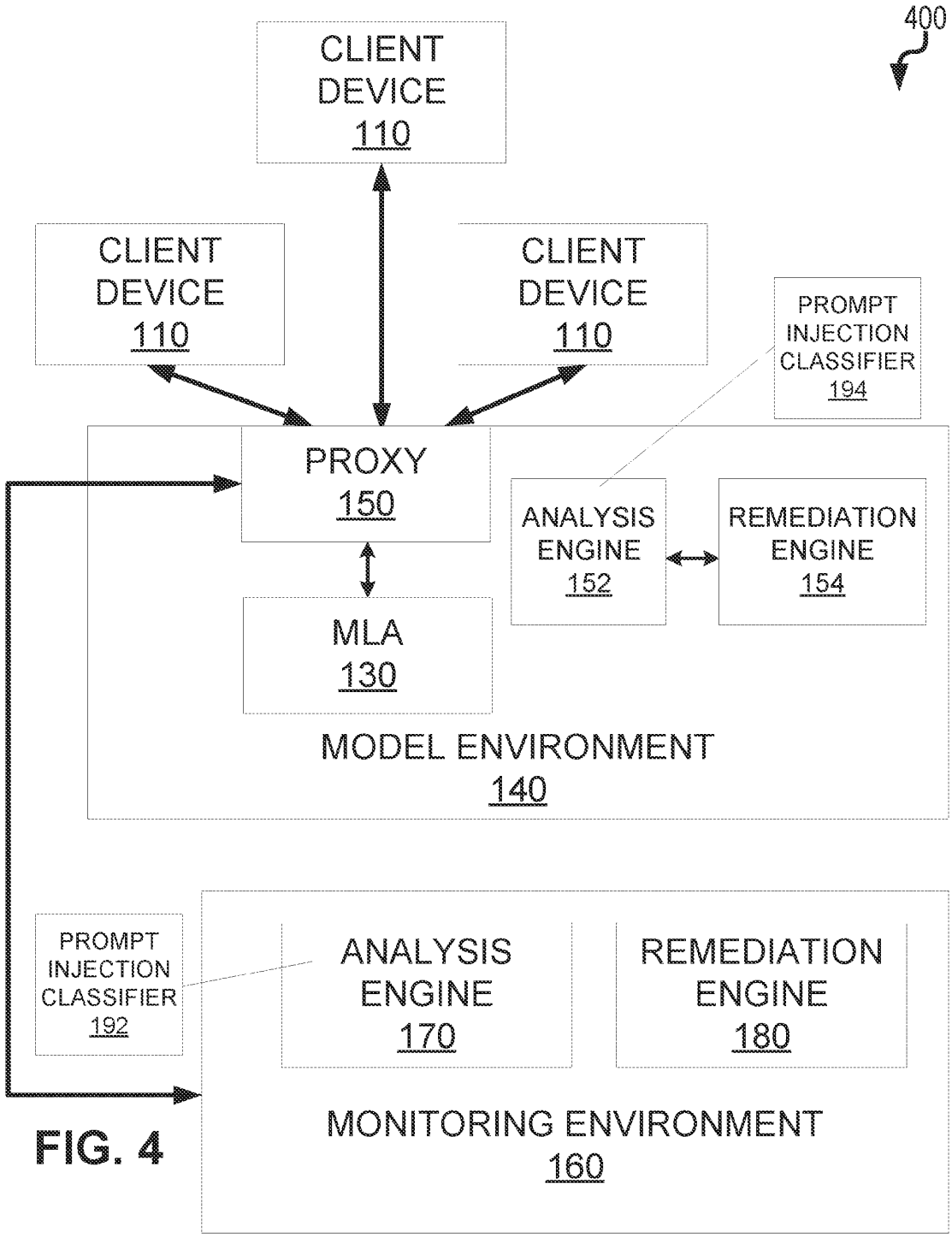

400
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROMPT INJECTION CLASSIFIER 194
PROXY 150
ANALYSIS ENGINE 152
REMEDIATION ENGINE 154
MLA 130
MODEL ENVIRONMENT 140
PROMPT INJECTION CLASSIFIER 192
ANALYSIS ENGINE 170
REMEDIATION ENGINE 180
MONITORING ENVIRONMENT 160

FIG. 4

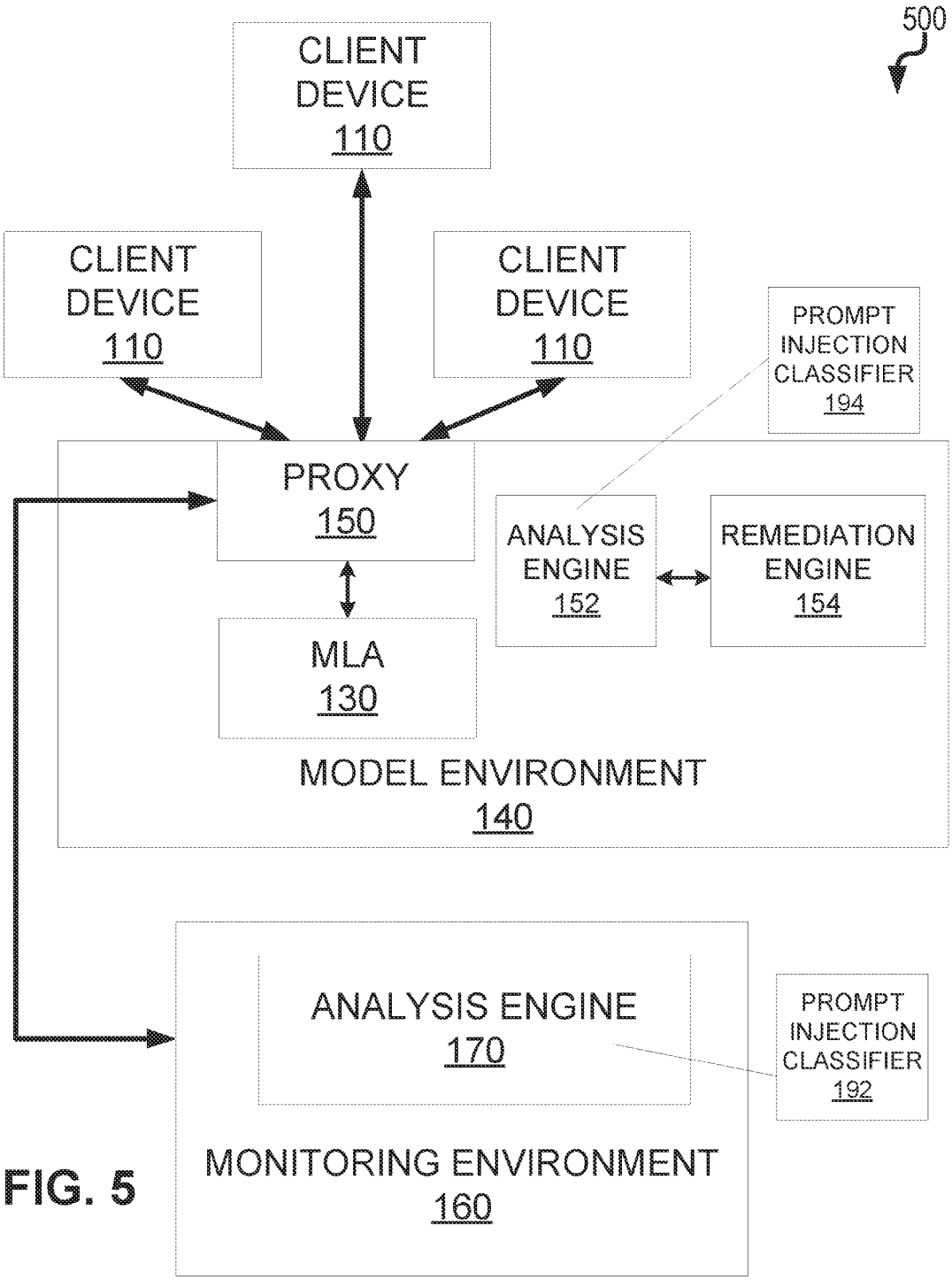

500
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROMPT INJECTION CLASSIFIER 194
PROXY 150
ANALYSIS ENGINE 152
REMEDIATION ENGINE 154
MLA 130
MODEL ENVIRONMENT 140
ANALYSIS ENGINE 170
PROMPT INJECTION CLASSIFIER 192
MONITORING ENVIRONMENT 160

FIG. 5

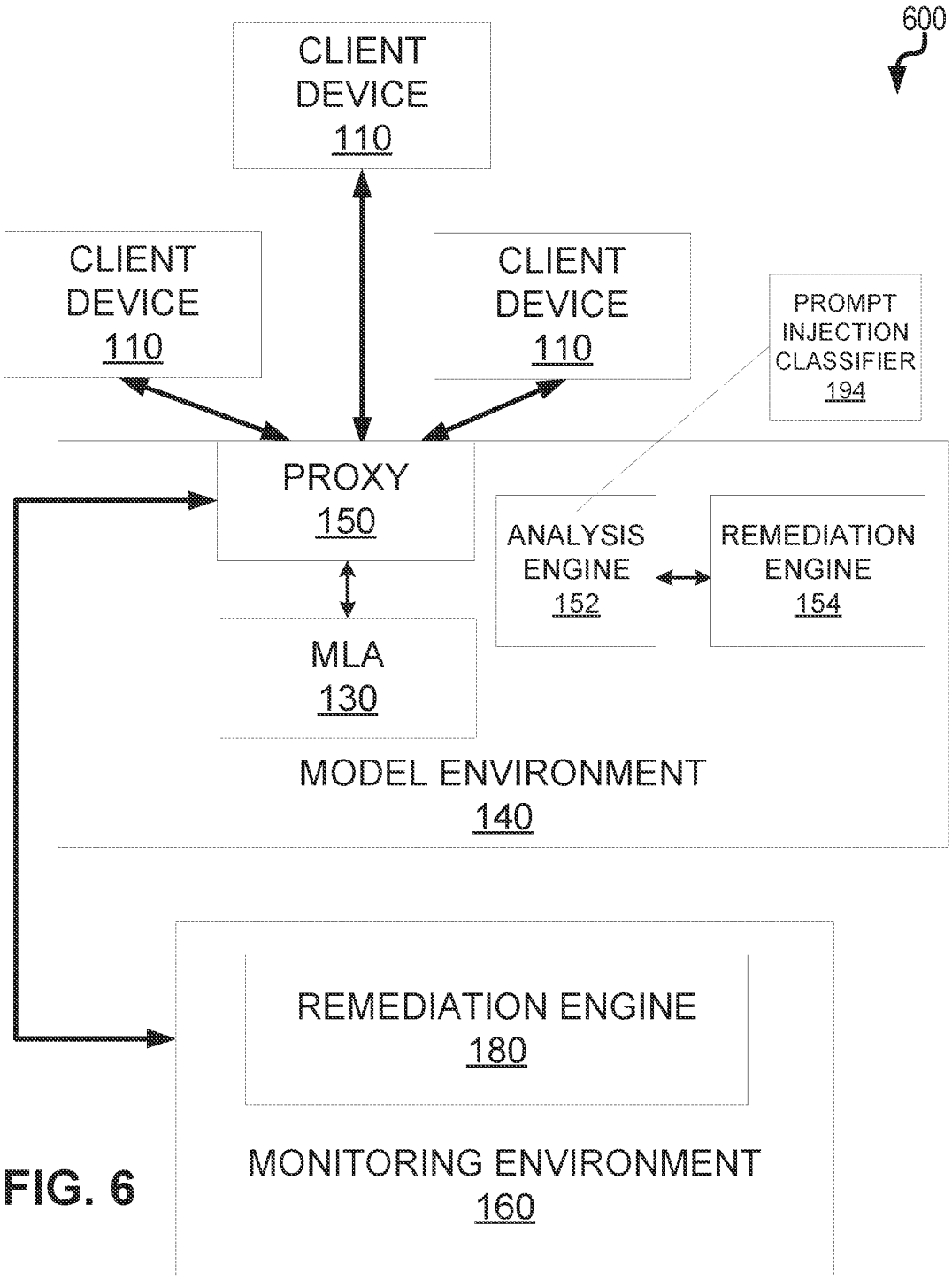

600
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROMPT INJECTION CLASSIFIER 194
PROXY 150
ANALYSIS ENGINE 152
REMEDIATION ENGINE 154
MLA 130
MODEL ENVIRONMENT 140
REMEDIATION ENGINE 180
MONITORING ENVIRONMENT 160

FIG. 6

FIG. 7

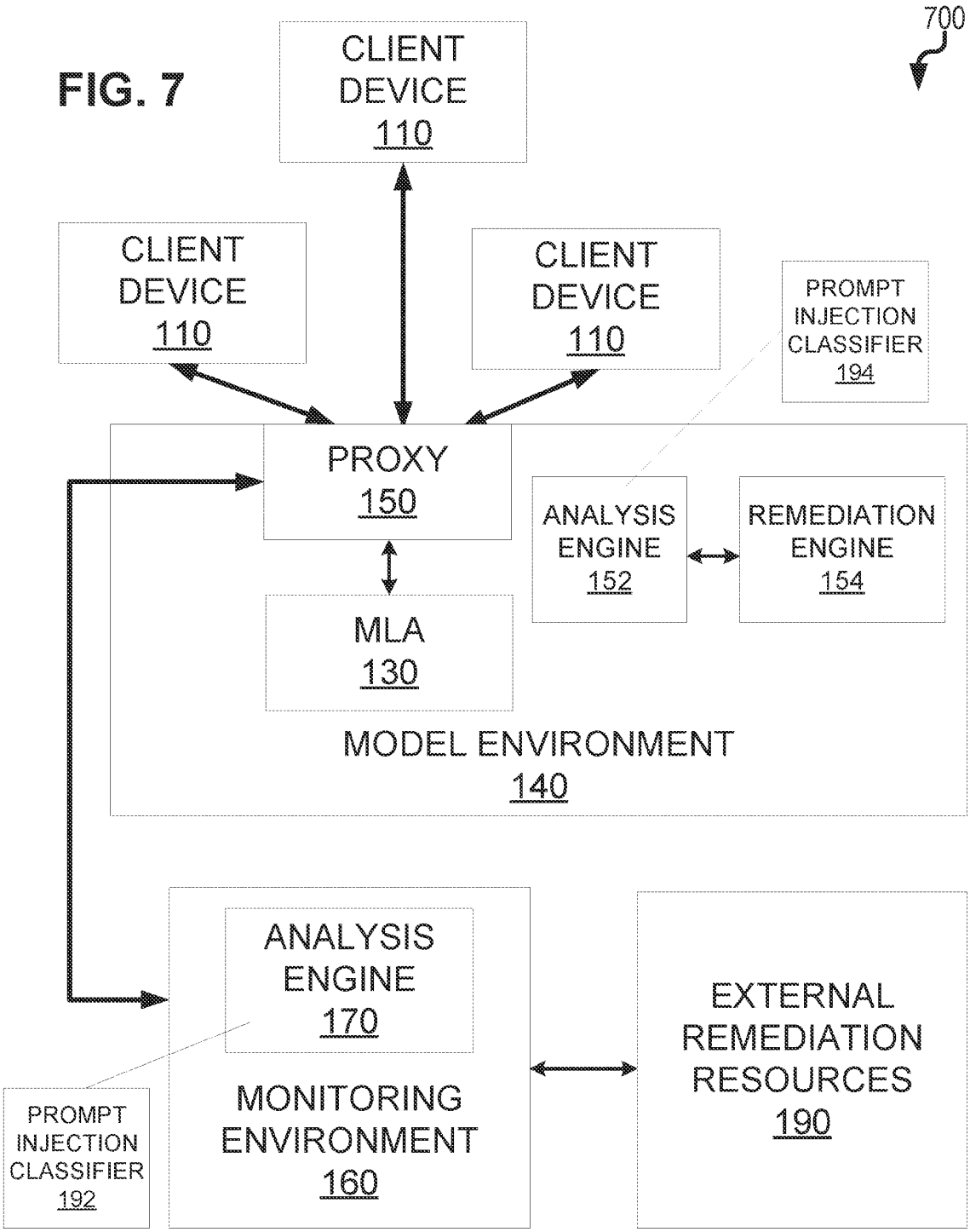

700
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROMPT INJECTION CLASSIFIER 194
PROXY 150
ANALYSIS ENGINE 152
REMEDIATION ENGINE 154
MLA 130
MODEL ENVIRONMENT 140
ANALYSIS ENGINE 170
MONITORING ENVIRONMENT 160
PROMPT INJECTION CLASSIFIER 192
EXTERNAL REMEDIATION RESOURCES 190

FIG. 8

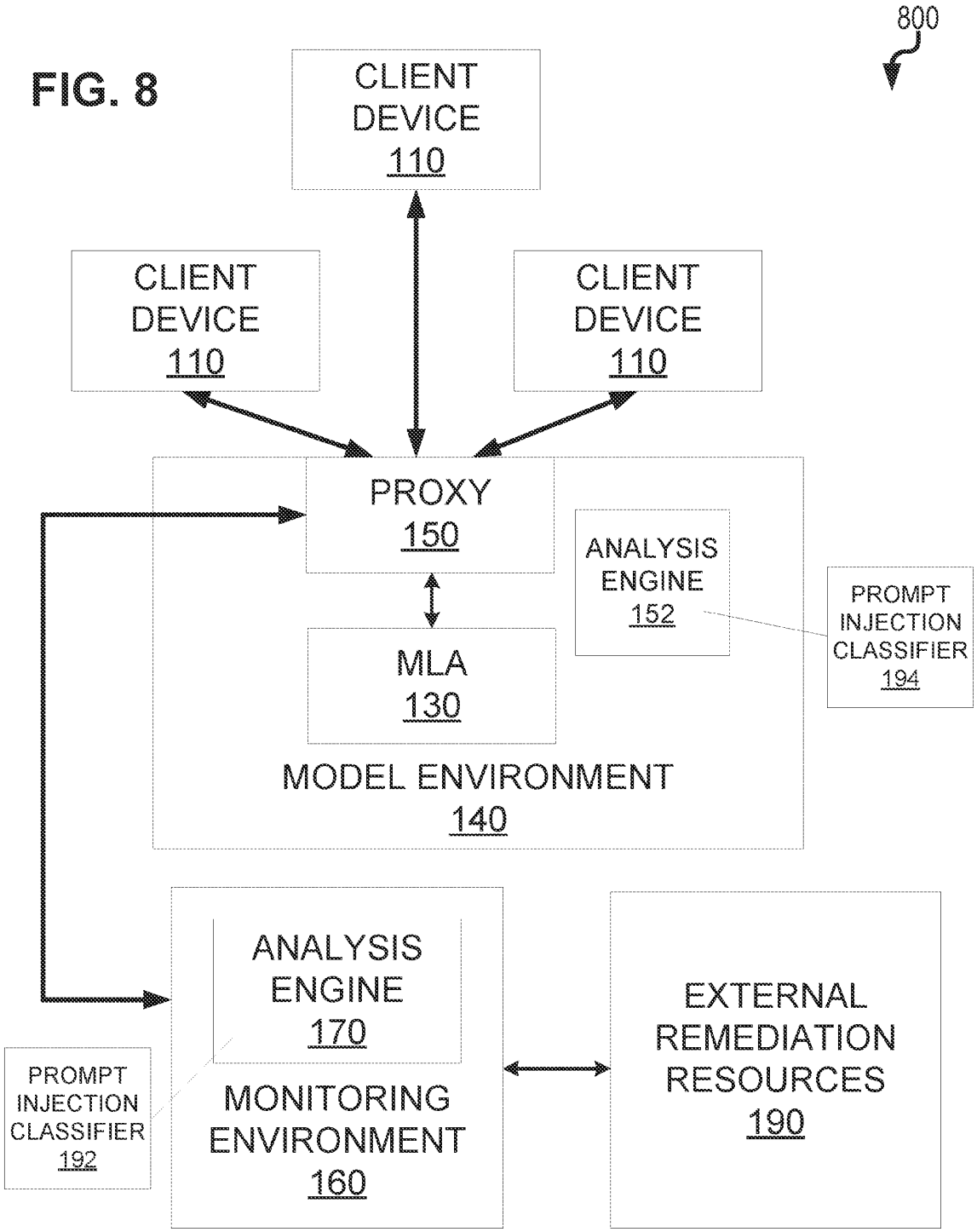

800
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROXY 150
ANALYSIS ENGINE 152
PROMPT INJECTION CLASSIFIER 194
MLA 130
MODEL ENVIRONMENT 140
ANALYSIS ENGINE 170
PROMPT INJECTION CLASSIFIER 192
MONITORING ENVIRONMENT 160
EXTERNAL REMEDIATION RESOURCES 190

FIG. 9

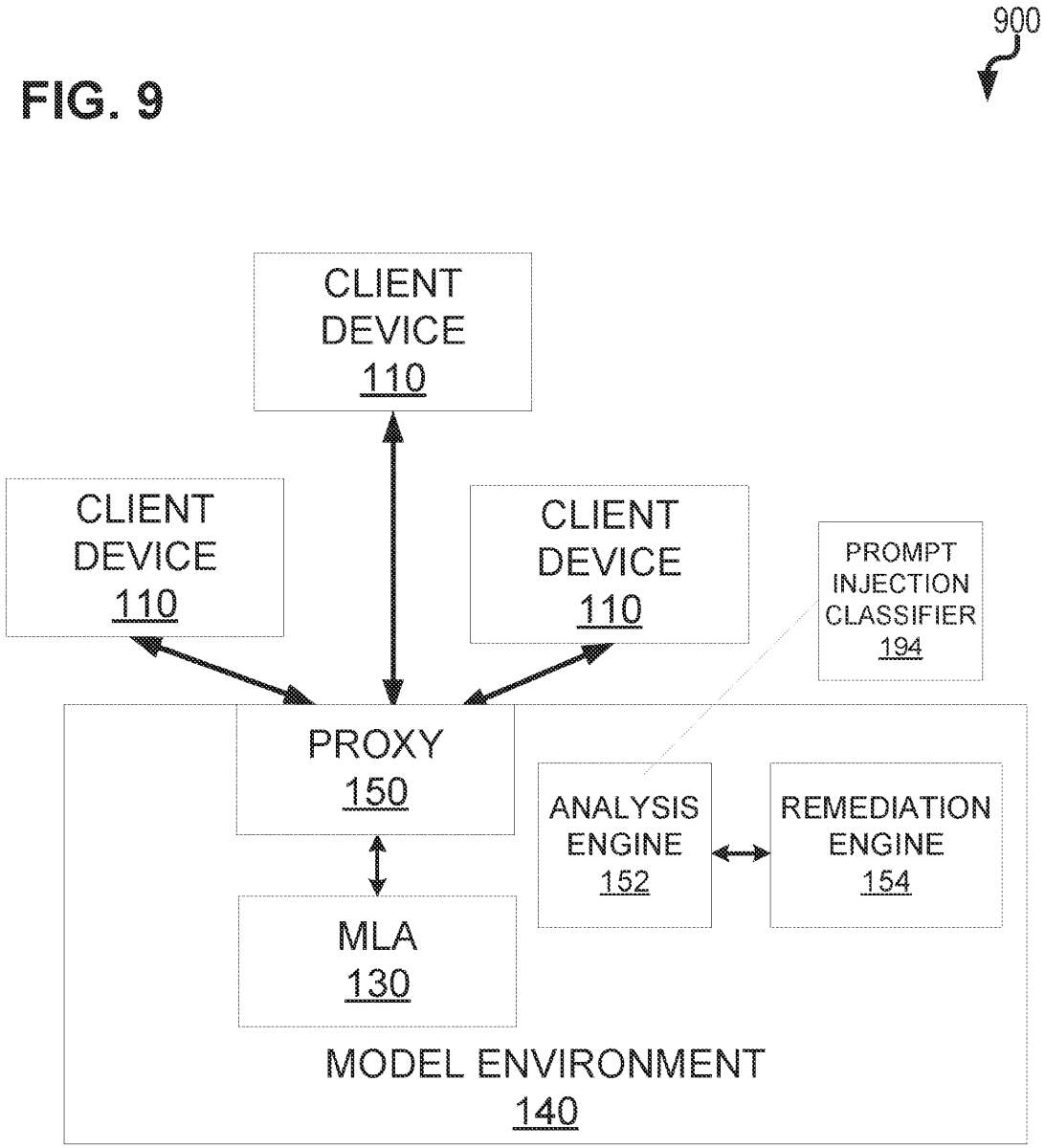
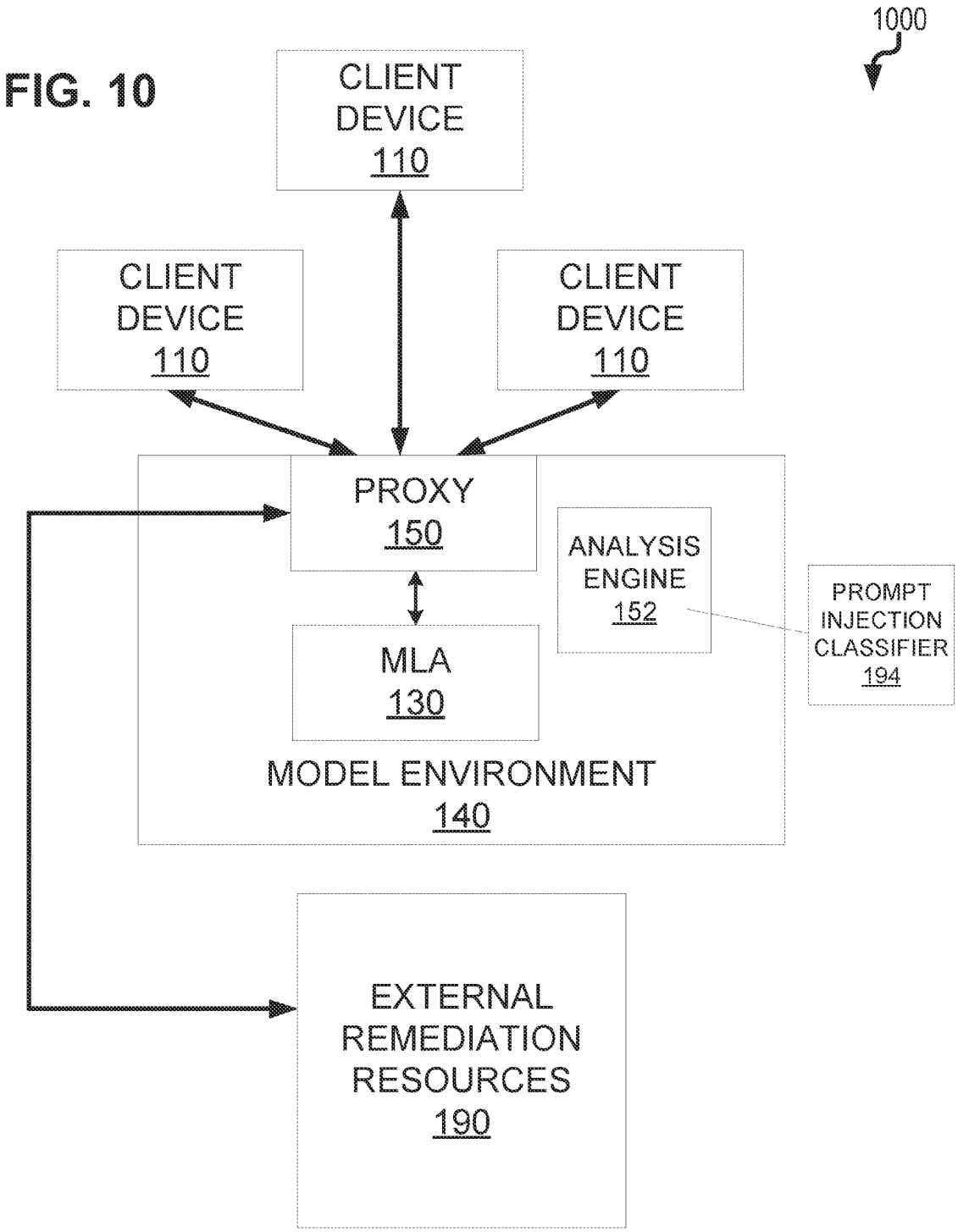

900
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROMPT INJECTION CLASSIFIER 194
PROXY 150
ANALYSIS ENGINE 152
REMEDIATION ENGINE 154
MLA 130
MODEL ENVIRONMENT 140

FIG. 10


1000
CLIENT DEVICE 110
CLIENT DEVICE 110
CLIENT DEVICE 110
PROXY 150
ANALYSIS ENGINE 152
PROMPT INJECTION CLASSIFIER 194
MLA 130
MODEL ENVIRONMENT 140
EXTERNAL REMEDIATION RESOURCES 190

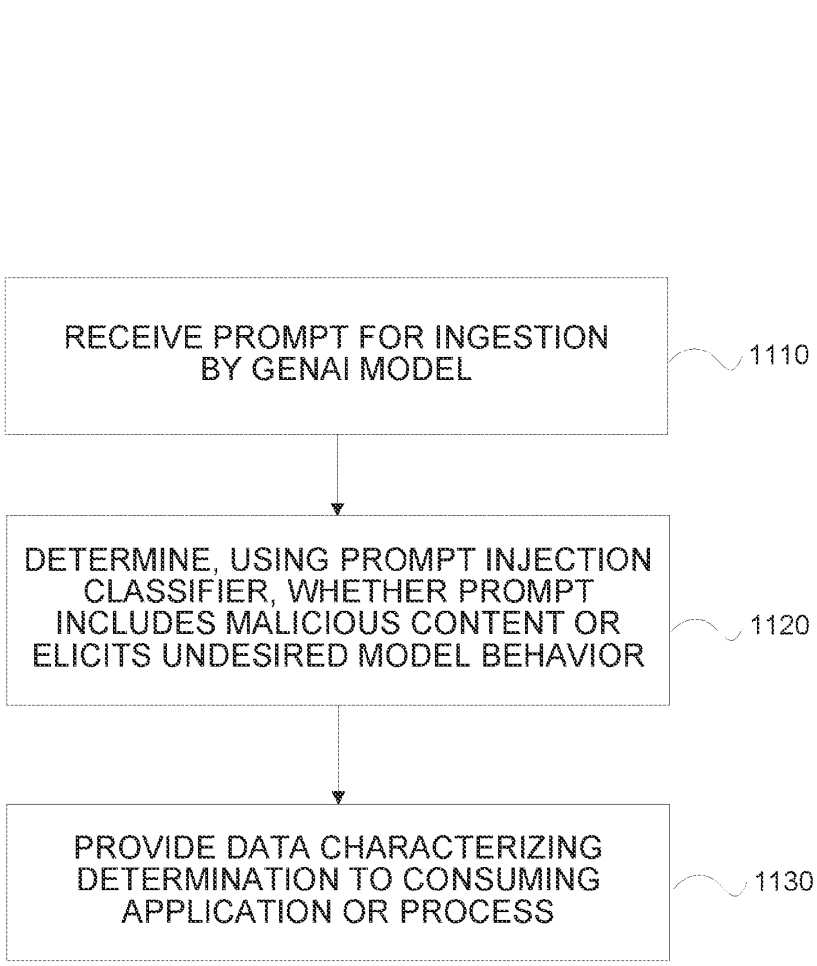

1100
RECEIVE PROMPT FOR INGESTION BY GENAI MODEL
1110
DETERMINE, USING PROMPT INJECTION CLASSIFIER, WHETHER PROMPT INCLUDES MALICIOUS CONTENT OR ELICITS UNDESIRED MODEL BEHAVIOR
1120
PROVIDE DATA CHARACTERIZING DETERMINATION TO CONSUMING APPLICATION OR PROCESS
1130

FIG. 11