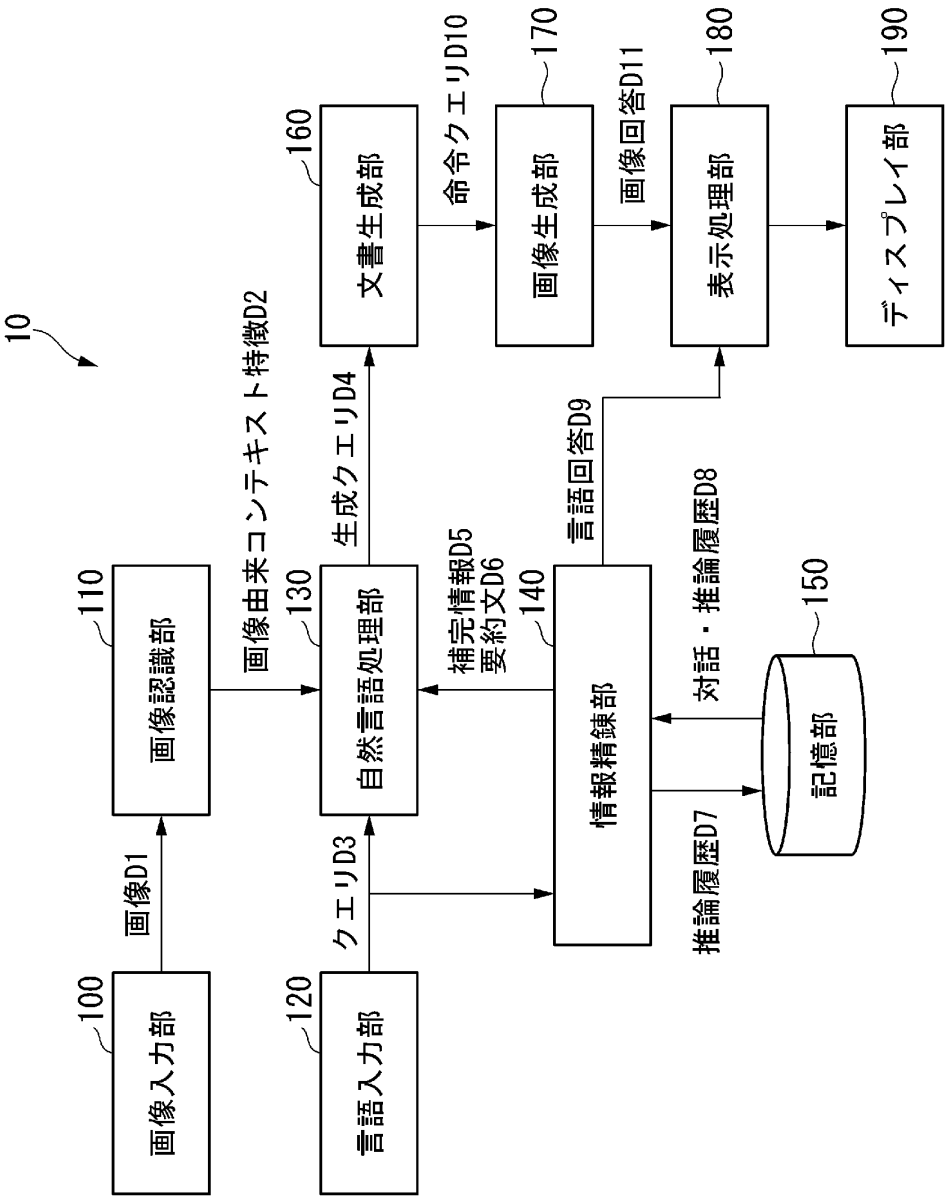
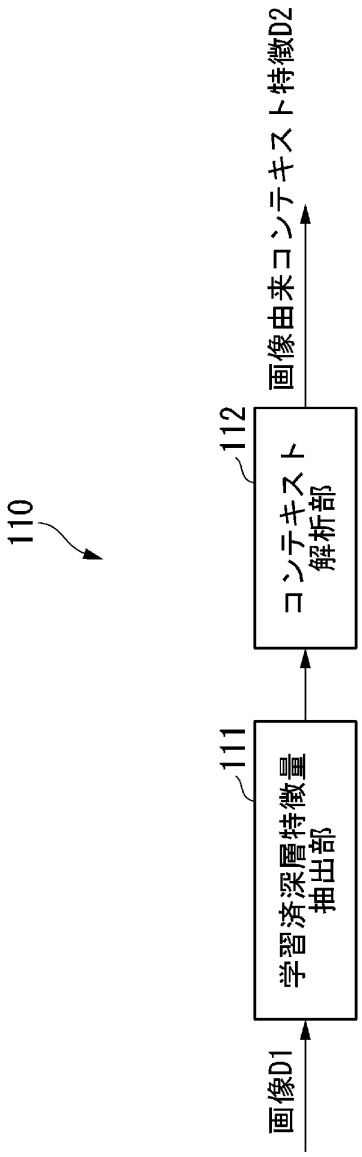
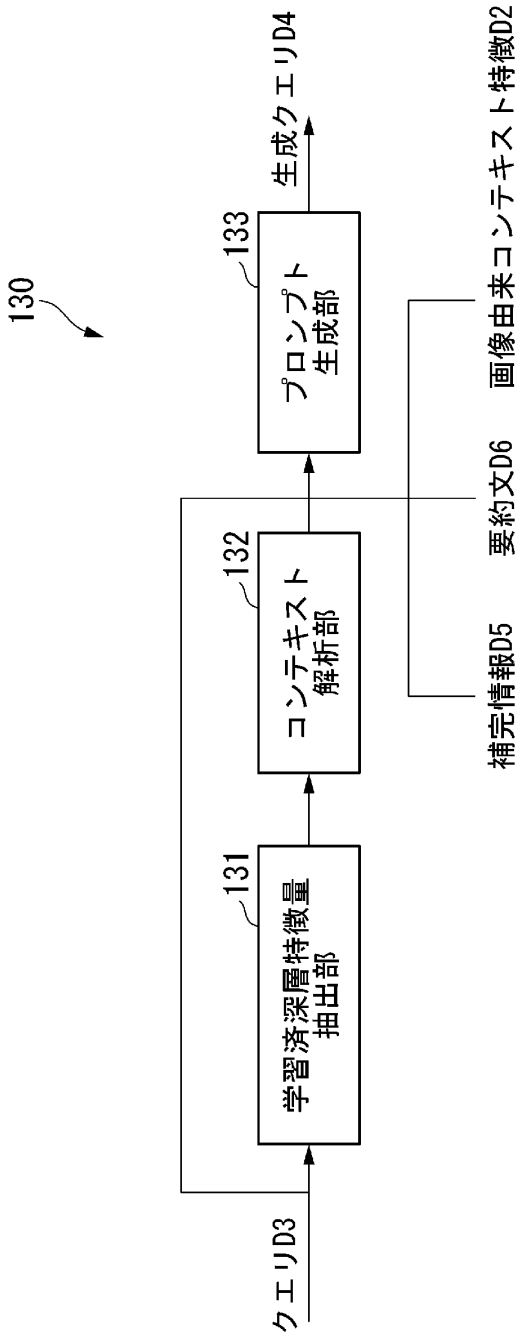
[図1]


10
100
画像入力部
110
画像認識部
120
言語入力部
130
自然言語処理部
140
情報精錬部
150
記憶部
160
文書生成部
170
画像生成部
180
表示処理部
190
ディスプレイ部
画像D1
クエリD3
画像由来コンテキスト特徴D2
生成クエリD4
補完情報D5
要約文D6
推論履歴D7
対話・推論履歴D8
言語回答D9
命令クエリD10
画像回答D11

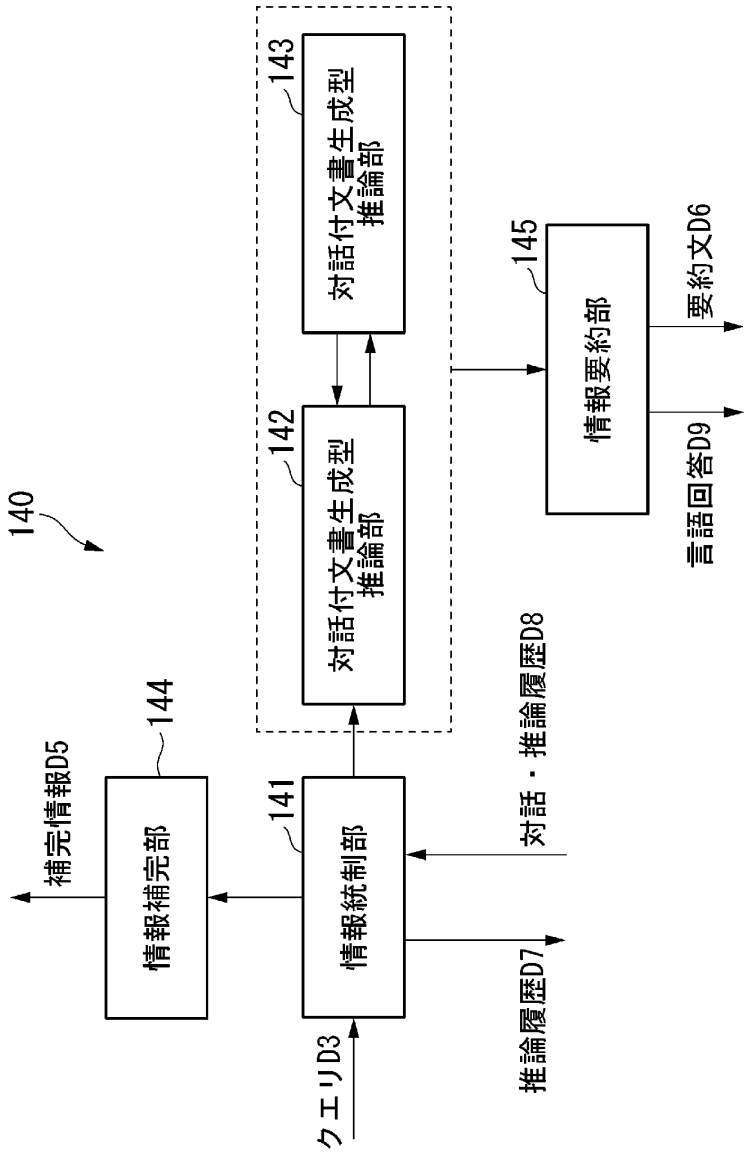
[図2]



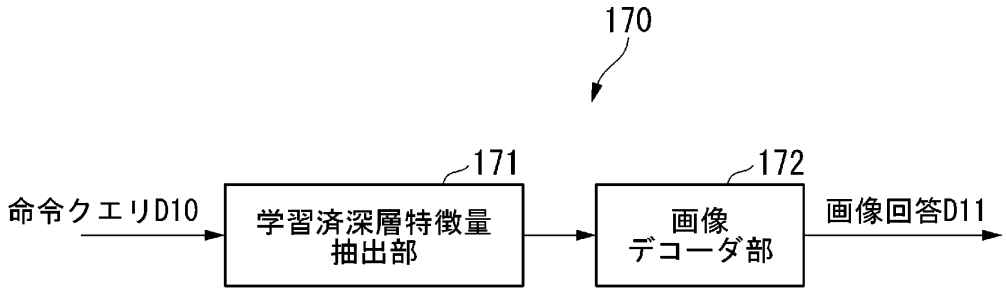
[図3]


130
131
学習済深層特徴量
抽出部
132
コンテキスト
解析部
133
プロンプト
生成部
クエリD3
補完情報D5
要約文D6
画像由来コンテキスト特徴D2
生成クエリD4

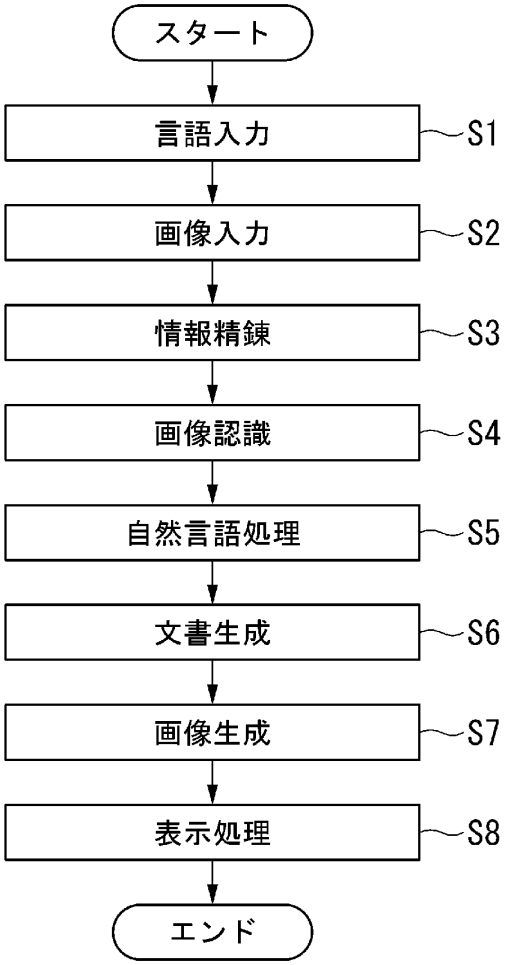
[図4]


140
143
対話付文書生成型
推論部
142
対話付文書生成型
推論部
145
情報要約部
要約文D6
言語回答D9
144
補完情報D5
情報補完部
141
情報統制部
対話・推論履歴D8
推論履歴D7
クエリD3

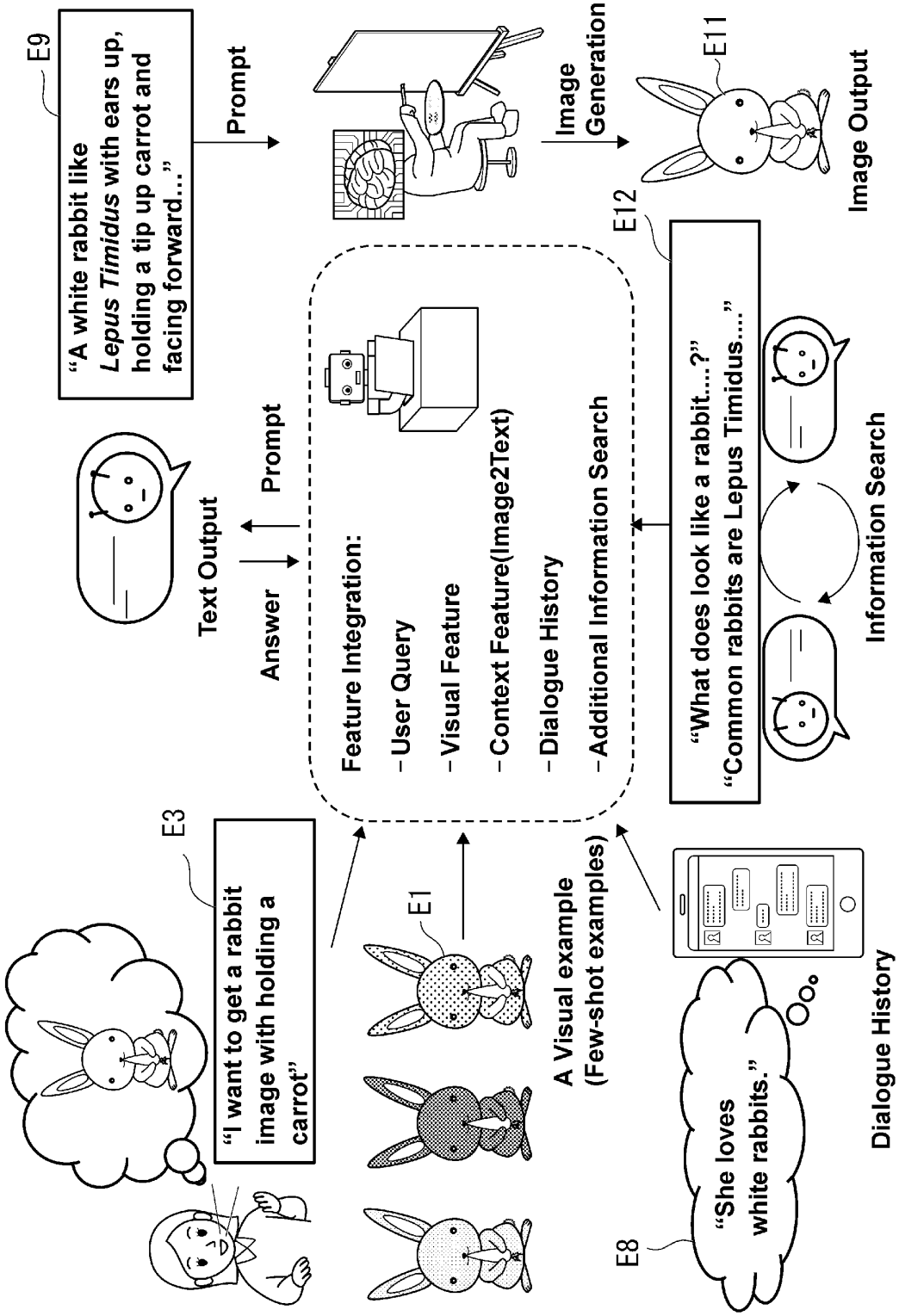
[図5]


170
171
命令クエリD10
学習済深層特徴量
抽出部
172
画像
デコーダ部
画像回答D11

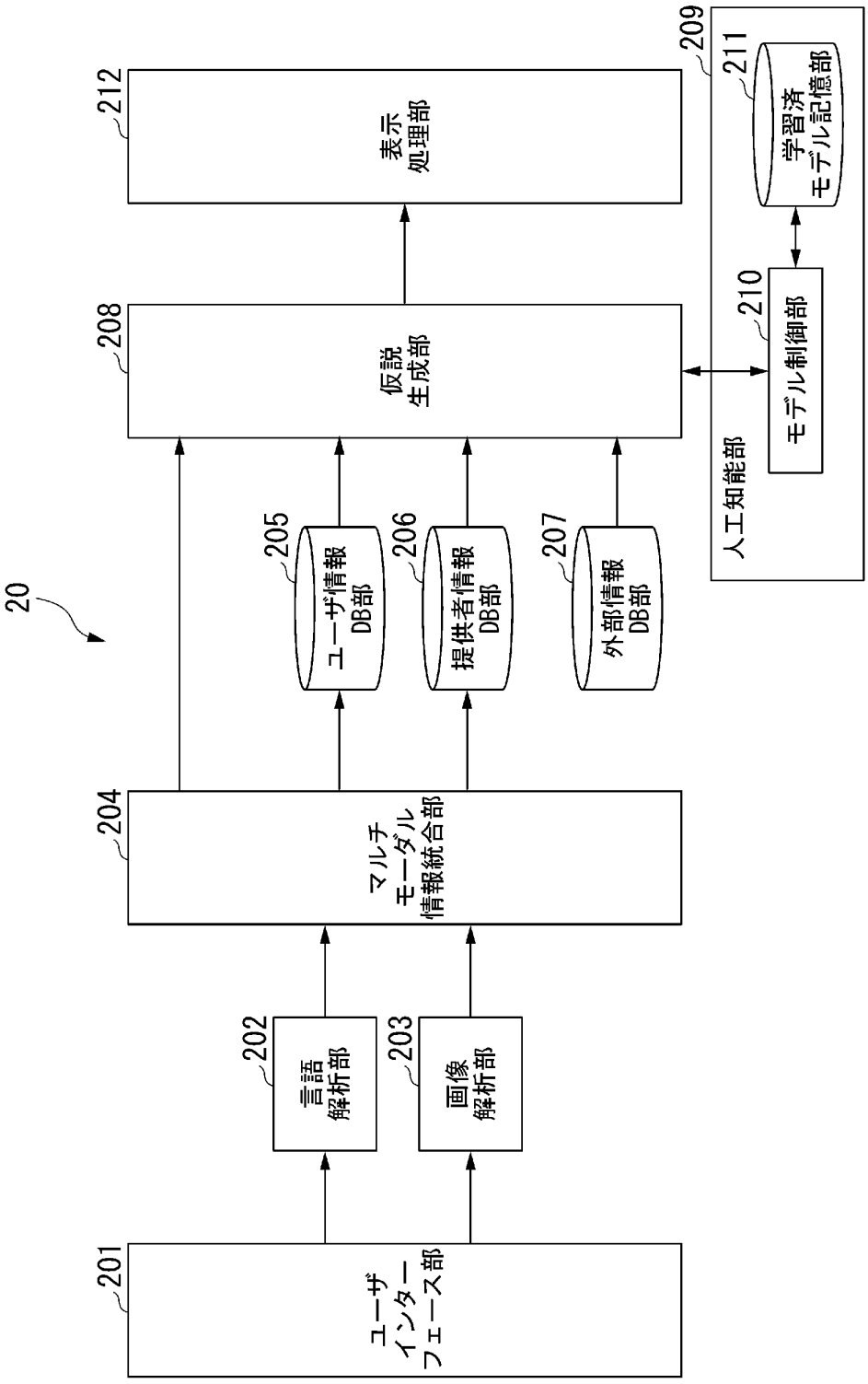
[図6]


スタート
言語入力
S1
画像入力
S2
情報精錬
S3
画像認識
S4
自然言語処理
S5
文書生成
S6
画像生成
S7
表示処理
S8
エンド

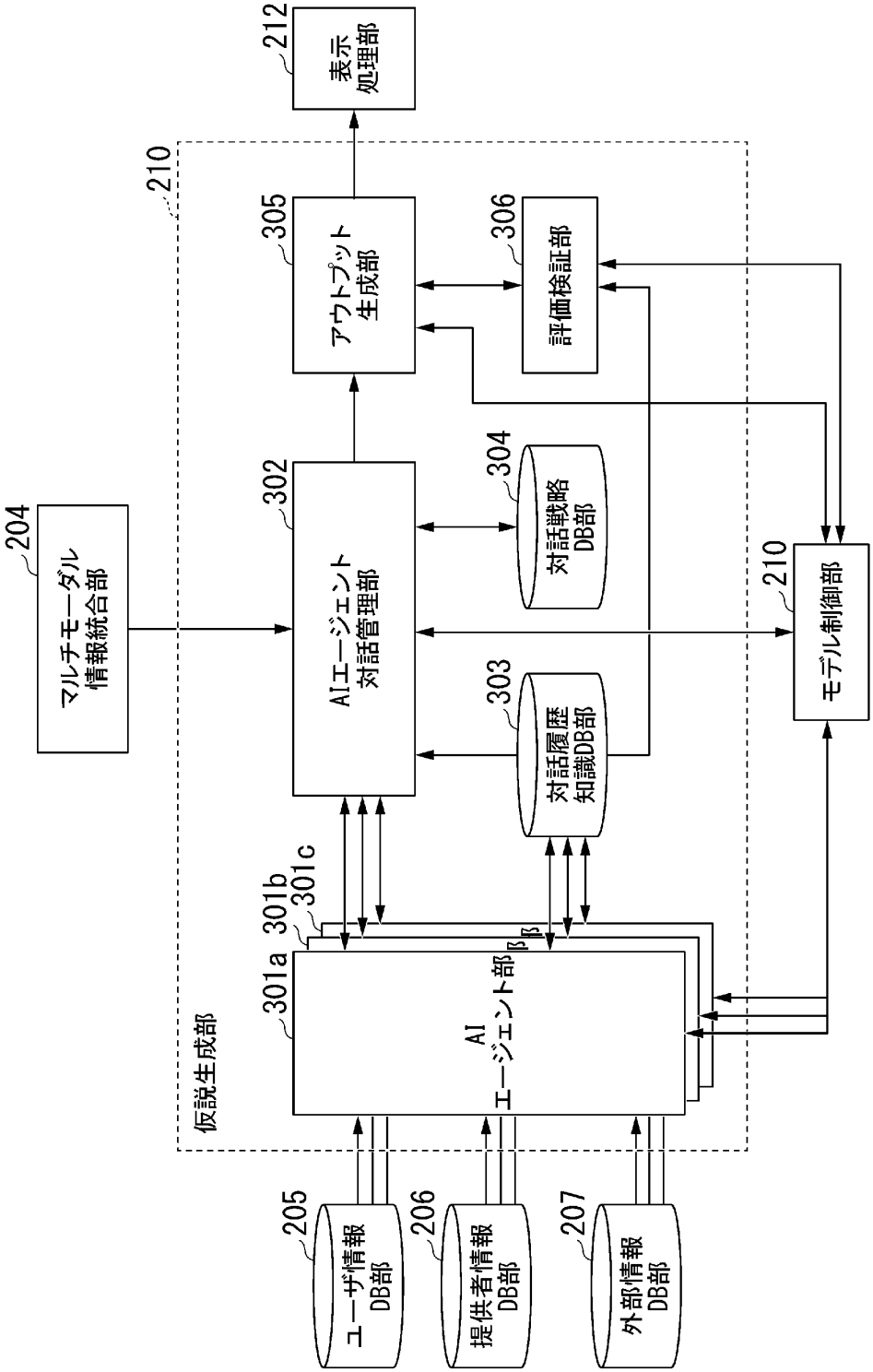
[図7]


E9
"A white rabbit like *Lepus Timidus* with ears up, holding a tip up carrot and facing forward..."
Prompt
Image Generation
E11
Image Output
E12
"What does look like a rabbit....?"
"Common rabbits are Lepus Timidus...."
Information Search
Text Output
Answer
Prompt
Feature Integration:
-User Query
-Visual Feature
-Context Feature(Image2Text)
-Dialogue History
-Additional Information Search
E3
"I want to get a rabbit image with holding a carrot"
E1
A Visual example
(Few-shot examples)
E8
"She loves white rabbits."
Dialogue History

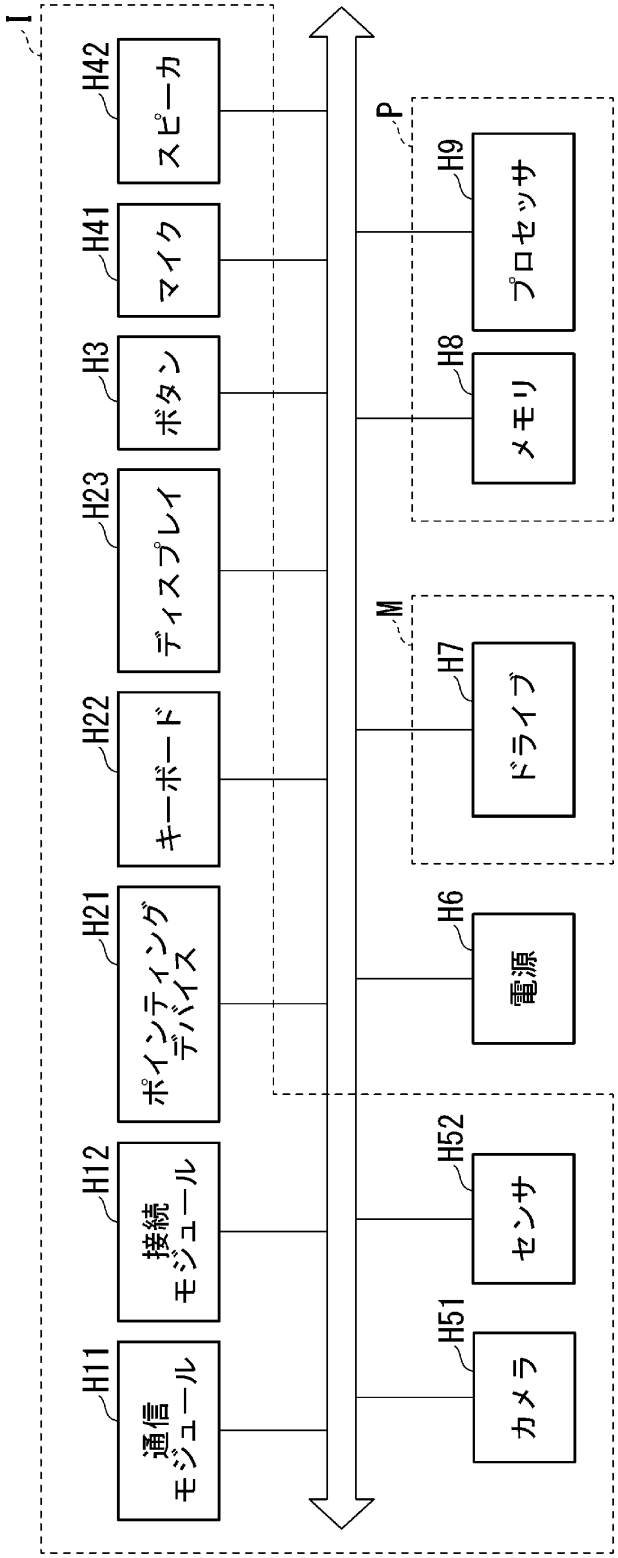
[図8]


20
201
ユーザ
インター
フェース部
202
言語
解析部
203
画像
解析部
204
マルチ
モーダル
情報統合部
205
ユーザ情報
DB部
206
提供者情報
DB部
207
外部情報
DB部
208
仮説
生成部
209
人工知能部
210
モデル制御部
211
学習済
モデル記憶部
212
表示
処理部

[図9]


212
表示
処理部
210
305
アウトプット
生成部
306
評価検証部
304
対話戦略
DB部
302
AIエージェント
対話管理部
204
マルチモーダル
情報統合部
303
対話履歴
知識DB部
210
モデル制御部
301b
301c
301a
AI
エージェント部
β
仮説生成部
205
ユーザ情報
DB部
206
提供者情報
DB部
207
外部情報
DB部

[図10]


I
H42
スピーカ
H41
マイク
H3
ボタン
H23
ディスプレイ
H22
キーボード
H21
ポインティングデバイス
H12
接続モジュール
H11
通信モジュール
P
H9
プロセッサ
H8
メモリ
M
H7
ドライブ
H6
電源
H52
センサ
H51
カメラ