明　細　書

発明の名称 ： 資料生成装置、資料生成方法、およびプログラム

技術分野

[0001]　本発明は、資料生成装置、資料生成方法、およびプログラムに関する。

背景技術

[0002]　近年、OpenAI（登録商標）社のchatGPT（登録商標）、ミュンヘン大学のCompVisグループのStable Diffusionなど、深層学習を利用した生成AI（Generative Artificial Intelligence）が開発されている。特許文献１では、与えられた質問内容に対して深層学習によって導かれた回答により応答する際に、質問内容に関連したテキスト及び画像を含む複数の情報に基づいて回答を生成する情報処理装置が開示されている。

先行技術文献

特許文献

[0003] 特許文献1：日本国特開２０２２－１１２５４１号公報

発明の概要

発明が解決しようとする課題

[0004]　しかしながら、特許文献１に記載の情報処理装置においては、人物の特徴に応じた資料を得ることができないことがあるという問題がある。

[0005]　本開示は、このような事情に鑑みてなされたもので、人物の特徴に応じた資料を生成できる資料生成装置、資料生成方法、およびプログラムを提供する。

課題を解決するための手段

[0006]　この開示は上述した課題を解決するためになされたもので、本開示の一態様は、生成する資料を指示するクエリの入力を受け付ける入力部と、前記クエリに基づき、複数のＡＩ（Artificial Intelligence）エージェントの間で対話させることにより、前記クエリにより指示された資料を推論する対話管理部とを備え、前記複数のＡＩエージェントによる対話に、前記複数のＡＩ

エージェントのうち、少なくとも１つのＡＩエージェントに対応する人物の特徴を示す情報を含める、資料生成装置である。

[0007] また、この開示の他の一態様は、上述した資料生成装置であって、前記複数のＡＩエージェントは、複数の対話付き文書生成モデルであり、前記入力部は、前記クエリの入力を受け付ける言語入力部であり、前記対話管理部は、前記クエリを初期入力として、前記複数の対話付き文書生成モデルの間で相互に入出力を繰り返して対話させることにより、前記クエリにより指示された資料のうちの文書を推論する情報精錬部であり、前記情報精錬部による推論結果の文書に基づき、画像を生成する画像生成部を備え、前記複数の対話付き文書生成モデルによる対話に、特定の個人の嗜好あるいは思考パターンが含まれる嗜好情報を加える。

[0008] また、この開示の他の一態様は、上述した資料生成装置であって、前記生成する資料の事例となる画像の入力を受け付ける画像入力部と、前記画像入力部に入力された画像を、学習済みモデルに入力して、前記画像入力部に入力された画像の特徴量を推論させて、画像由来コンテキスト特徴とする画像認識部とをさらに備え、前記画像生成部は、前記情報精錬部による推論結果の文書に加えて、前記画像由来コンテキスト特徴に基づき、画像を生成する。

[0009] また、この開示の他の一態様は、上述した資料生成装置であって、前記クエリを、学習済みモデルに入力して、前記クエリの特徴量を推論させて言語由来コンテキスト特徴とし、前記クエリ、前記言語由来コンテキスト特徴、前記画像由来コンテキスト特徴、前記嗜好情報、前記情報精錬部による推論結果の文書の少なくとも一部を組み合わせた生成クエリを生成する自然言語処理部と、前記生成クエリを、対話付文書生成モデルに入力し、前記生成クエリにより要求された文書を推論し、推論結果を命令クエリとする文書生成部とをさらに備え、前記画像生成部は、前記命令クエリを学習済み画像生成モデルに入力して、前記命令クエリが示す画像を推論させる。

[0010] また、この開示の他の一態様は、上述した資料生成装置であって、前記複

数のＡＩエージェントによる対話は、１つのＡＩエージェントによる仮説の生成と、生成された前記仮説に対する他のＡＩエージェントによる批判または批評を含む。

[0011] また、この開示の他の一態様は、上述した資料生成装置であって、前記複数のＡＩエージェントは、一対一に複数の人物に対応し、前記複数のＡＩエージェントの各々は、前記複数の人物のうち、対応する人物の特徴を示す人物情報を対話に含める。

[0012] また、この開示の他の一態様は、上述した資料生成装置であって、前記複数の人物の少なくとも一人は、仮想の人物であり、前記仮想の人物の前記人物情報は、前記仮想の人物の役割を示す情報を含む。

[0013] また、この開示の他の一態様は、上述した資料生成装置であって、対話プロセスと、対話させる人物とを示す対話戦略情報を記憶する対話戦略記憶部を備え、前記対話管理部は、前記対話戦略記憶部から取得した対話戦略情報が示す前記対話プロセスに従い前記複数のＡＩエージェントから選択した前記ＡＩエージェントに対話させる。

[0014] また、この開示の他の一態様は、上述した資料生成装置であって、前記資料に関連する情報を記憶する記憶部を備え、前記複数のＡＩエージェントの各々は、前記記憶部から前記対話に関連する参照情報を取得し、取得した前記参照情報を前記対話に含める。

[0015] また、この開示の他の一態様は、上述した資料生成装置であって、前記入力部は、前記資料に関連するマルチモーダルなデータの入力を受け付け、前記対話管理部は、前記クエリおよび前記マルチモーダルな情報に基づき、前記複数のＡＩエージェントの間で対話させ、前記複数のＡＩエージェントによる対話は、マルチモーダルである。

[0016] また、この開示の他の一態様は、上述した資料生成装置であって、前記複数のＡＩエージェントによる対話の結果に基づき、前記資料を生成するアウトプット生成部と、前記アウトプット生成部が生成した資料を評価および検証する評価検証部とを備える。

[0017] また、この開示の他の一態様は、生成する資料を指示するクエリの入力を受け付ける第１のステップと、前記クエリに基づき、複数のＡＩエージェントの間で対話させることにより、前記クエリにより指示された資料を推論する第２のステップとを有し、前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する仮想を含む人物の特徴を示す情報を加える、資料生成方法である。

[0018] また、この開示の他の一態様は、コンピュータを、生成する資料を指示するクエリの入力を受け付ける入力部、前記クエリに基づき、複数のＡＩエージェントの間で対話させることにより、前記クエリにより指示された資料を推論する対話管理部として機能させるためのプログラムであって、前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する仮想を含む人物の特徴を示す情報を加える、プログラムである。

発明の効果

[0019] 本開示の資料生成装置、資料生成方法、およびプログラムは、人物の特徴に応じた資料を生成できる。

図面の簡単な説明

[0020] [図1]この開示の第１の実施形態による資料生成装置１０の構成を示す概略ブロック図である。

[図2]同実施形態における画像認識部１１０の構成例を示す概略ブロック図である。

[図3]同実施形態における自然言語処理部１３０の構成例を示す概略ブロック図である。

[図4]同実施形態における情報精錬部１４０の構成例を示す概略ブロック図である。

[図5]同実施形態における画像生成部１７０の構成例を示す概略ブロック図である。

[図6]同実施形態における資料生成装置１０の動作例を説明するフローチャー

トである。

[図7]同実施形態における資料生成装置１０のユースケースを示す模式図である。

[図8]この開示の第２の実施形態による資料生成装置２０の構成を示す概略ブロック図である。

[図9]同実施形態における仮説生成部２０８の構成を示す概略ブロック図である。

[図10]各実施形態に係る資料生成装置１０のハードウェア構成を説明する説明図である。

発明を実施するための形態

[0021] ＜第１の実施形態＞

以下、図面を参照して、本開示の第１の実施形態について説明する。図１は、この開示の第１の実施形態による資料生成装置１０の構成を示す概略ブロック図である。資料生成装置１０は、ユーザの要求に応じた文章と画像とからなる資料を生成する装置である。資料生成装置１０は、１つまたは複数のコンピュータがプログラムを読み込み実行することで実現されてもよい。該複数のコンピュータは、汎用あるいは専用のネットワークで相互に通信可能に接続されていてもよい。資料生成装置１０は、画像入力部１００、画像認識部１１０、言語入力部１２０（入力部）、自然言語処理部１３０、情報精錬部１４０（対話管理部）、記憶部１５０、文書生成部１６０、画像生成部１７０、表示処理部１８０、ディスプレイ部１９０を備える。

[0022] 画像入力部１００は、生成する資料の事例となる画像Ｄ１の、ユーザによる入力を受け付ける。なお、画像入力部１００は、キーボード、マウスなどの入力手段を用いてユーザによる入力を受け付けてもよいし、他装置から受信するなどして画像Ｄ１の入力を受け付けてもよい。画像認識部１１０は、画像入力部１００に入力された画像Ｄ１を、学習済みモデルに入力して、その特徴量を推論させて、画像由来コンテキスト特徴Ｄ２とする。なお、画像由来コンテキスト特徴Ｄ２は、推論結果の特徴量から重要な重みを持つ特徴

量を特定したものであってもよい。また、学習済みモデルは、例えば、OpenAI（登録商標）社のCLIP（Contrastive Language-Image Pre-training）のImage encoder、ViT（Vision Transformer）などの階層型ネットワークであってもよい。学習済みモデルは、画像認識部１１０が備えていてもよいし、ネットワークなどを介して資料生成装置１０と通信接続された他装置が備えていてもよい。また、重要な重みを持つ特徴量の特定は、マルチヘッドアテンションなどのアテンション機構により行われてもよい。

[0023] 言語入力部１２０は、生成する資料を指示する言語（クエリＤ３）の、ユーザによる入力を受け付ける。なお、言語入力部１２０は、キーボード、マウスなどの入力手段を用いてユーザによる入力を受け付けてもよいし、他装置から受信するなどしてクエリＤ３の入力を受け付けてもよい。このクエリＤ３は、テキストデータであり、生成する資料の事例となるテキストデータが含まれていてもよい。自然言語処理部１３０は、言語入力部１２０に入力されたクエリＤ３を、学習済みモデルに入力して、その特徴量を推論させて、言語由来コンテキスト特徴とする。なお、言語由来コンテキスト特徴は、推論結果の特徴量から重要な重みを持つ特徴量が特定されたものであってもよい。さらに、自然言語処理部１３０は、クエリＤ３、言語由来コンテキスト特徴、画像由来コンテキスト特徴Ｄ２、後述する嗜好情報である補完情報Ｄ５、情報精錬部１４０による推論結果の文書である要約文Ｄ６の少なくとも一部を組み合わせた言語（テキストデータ）である生成クエリＤ４を生成する。なお、生成クエリＤ４は、文書生成部１６０に、生成クエリＤ４が示す文書を推論させるためのプロンプトであってもよい。

[0024] 自然言語処理部１３０の学習済みモデルは、例えば、CLIPのtext encoderなどの階層型ネットワークであってもよい。学習済みモデルは、自然言語処理部１３０が備えていてもよいし、ネットワークなどを介して資料生成装置１０と通信接続された他装置が備えていてもよい。また、重要な重みを持つ特徴量の特定は、アテンション機構により行われてもよい。なお、自然言語処理部１３０の学習済みモデルは、画像認識部１１０の学習済みモデルとペ

アになっており、画像を説明する自然言語の特徴量として、画像認識部１１０の学習済みモデルが該画像から推論する特徴量と同様の特徴量を推論するように訓練されている。

[0025] 情報精錬部１４０は、クエリＤ３を初期入力として、２つの対話付き文書生成モデルの間で相互に入出力を繰り返して対話させることにより、クエリＤ３により指示された資料のうちの文書を推論する。情報精錬部１４０は、これらの２つの対話付き文書生成モデルによる対話に、特定の個人の嗜好あるいは思考パターンが含まれる嗜好情報を加える。これらの２つの対話付きの文書生成モデルは、例えば、OpenAI（登録商標）社のchatGPT（登録商標）などの対話付き文書生成ネットワークであってもよい。情報精錬部１４０は、推論した文書を要約し、その結果を要約文Ｄ６として、自然言語処理部１３０に入力する。また、情報精錬部１４０は、推論した文書の要約の結果を、言語回答Ｄ９とし、表示処理部１８０に入力する。

[0026] なお、これらの２つの対話付き文書生成モデルのうち、一方には詳細な文書の推論を要求し、もう一方には簡潔な文書の推論を要求してもよい。さらに、２つの対話付き文書生成モデルのうち、少なくとも一方への入力は、もう一方の出力に、記憶部１５０が記憶する対話・推論履歴Ｄ８が追加されたものであってもよい。対話・推論履歴Ｄ８は、情報精錬部１４０が、２つの対話付き文書生成モデルによる推論結果を、記憶部１５０に記憶させた推論履歴Ｄ７を含んでいてもよい。また、これらの対話付き文書生成モデルは、情報精錬部１４０が備えていてもよいし、ネットワークなどを介して資料生成装置１０と通信接続された他装置が備えていてもよい。

[0027] さらに、記憶部１５０が記憶する対話・推論履歴Ｄ８は、特定の個人の嗜好あるいは思考パターンが含まれる嗜好情報（例えば、該個人のＳＮＳ(Social Network Service)への投稿履歴）を含んでいてもよい。特定の個人は、資料の生成を要求したユーザであってもよいし、生成された資料を読むと想定される人物あるいは人物のグループであってもよい。また、情報精錬部１４０は、嗜好情報、もしくは嗜好情報からエンコーダーにより抽出された思考

パターンを高次元特徴として表現したもの、または該高次元特徴をデコードしたものを、補完情報Ｄ５として自然言語処理部１３０に入力してもよい。

[0028] 文書生成部１６０は、生成クエリＤ４を、対話付文書生成モデルに入力し、生成クエリＤ４により要求された文書を推論し、推論結果を命令クエリＤ１０とする。この対話付文書生成モデルは、例えば、chatGPT（登録商標）、BERTなどの大規模言語モデルであってもよい。また、この対話付き文書生成モデルは、文書生成部１６０が備えていてもよいし、ネットワークなどを介して資料生成装置１０と通信接続された他装置が備えていてもよい。なお、命令クエリＤ１０は、生成クエリＤ４により要求された文書に対応する画像を、画像生成部１７０に推論させるためのプロンプトであってもよい。

[0029] 画像生成部１７０は、少なくとも情報精錬部１４０による推論結果の文書に基づき、画像を生成する。さらに、この画像の生成は、画像由来コンテキスト特徴に基づいてもよい。具体的には、画像生成部１７０は、これらの情報を用いて生成されている命令クエリＤ１０を、学習済み画像生成モデルに入力して、命令クエリＤ１０が示す画像を推論させ、該推論結果を画像回答Ｄ１１とする。この学習済み画像生成モデルは、例えば、ミュンヘン大学のCompVisグループが開発したStable Diffusionなどのテキストから画像を生成する画像合成モデルであってもよいし、CLIPのtext encoderと、それに対応するdiffusion decoderとの組み合わせであってもよい。

[0030] 表示処理部１８０は、情報精錬部１４０から言語回答Ｄ９を、画像生成部１７０から画像回答Ｄ１１を取得し、これらをディスプレイ部１９０に表示させる。ディスプレイ部１９０は、液晶ディスプレイ、有機ＥＬ（Electro-Luminescence）ディスプレイなどのディスプレイである。

[0031] 図２は、本実施形態における画像認識部１１０の構成例を示す概略ブロック図である。画像認識部１１０は、学習済深層特徴量抽出部１１１、コンテキスト解析部１１２を備える。学習済深層特徴量抽出部１１１は、画像Ｄ１を、CLIPのImage encoder、ViTなどの学習済みモデルに入力して、その特徴量を推論させる。コンテキスト解析部１１２は、学習済深層特徴量抽出部１

１１が推論した特徴量の中から、アテンション機構により重要な重みを持つ特徴量を特定し、画像由来コンテキスト特徴Ｄ２とする。アテンション機構としては、セルフアテンション（Self-Attention）と、クロスアテンション（Cross-Attention）とを用いる。セルフアテンションは、シーケンスデータ内で異なる位置の要素間の関係性を学習し重みを割り当てる。すなわち、セルフアテンションは、同じ画像空間内の類似箇所や特徴を学習し重みを割り当てる。クロスアテンションは、異なる入力間の相関関係を学習でき、画像であれば入力で与えられた異なる画像間で類似する特徴を学習でき、言語であれば入力で与えられた複数のテキスト間の特徴を大局的に得ることができる。また、言語と画像などの異なるモダリティの重要な特徴の獲得もクロスアテンションで実現できる。なお、セルフアテンションもクロスアテンションもモデル自身が異なる要素間の関係性を自動で獲得し、情報の重要性を自動的に学習できるという特徴がある。

[0032] 図３は、本実施形態における自然言語処理部１３０の構成例を示す概略ブロック図である。自然言語処理部１３０は、学習済深層特徴量抽出部１３１、コンテキスト解析部１３２、プロンプト生成部１３３を備える。学習済深層特徴量抽出部１３１は、クエリＤ３を、CLIPのtext encoderなどの学習済みモデルに入力して、その特徴量を推論させる。コンテキスト解析部１３２は、学習済深層特徴量抽出部１３１が推論した特徴量の中から、アテンション機構により重要な重みを持つ特徴量を特定し、言語由来コンテキスト特徴とする。アテンション機構としては、セルフアテンションと、クロスアテンションとを用いる。

[0033] プロンプト生成部１３３は、言語由来コンテキスト特徴と、画像由来コンテキスト特徴Ｄ２とを合成したコンテキスト特徴を、学習済みモデルに入力して該コンテキスト特徴を表すテキストデータであるコンテキスト言語を推論する。言語由来コンテキスト特徴と、画像由来コンテキスト特徴Ｄ２との合成は、例えば、言語由来コンテキスト特徴と、画像由来コンテキスト特徴Ｄ２との要素ごと(element-wise)の足し算や掛け算であってもよいし、異な

るテンソルの結合としてコンキャット(concatenate)してもよいし、クロスアテンションにより異なるモダリティの要素間の関係性を学習してもよい。

[0034] すなわち、言語由来コンテキスト特徴の要素がＴ１$_i$であり、画像由来コンテキスト特徴Ｄ２の要素がＴ２$_i$であり、ｉは０からＮ－１（Ｎは各特徴の要素数）の整数であるとして、合成したコンテキスト特徴であるＴ$_i$＝Ｔ１$_i$×Ｔ２$_i$であってもよいし、Ｔ$_i$＝Ｔ１$_i$＋Ｔ２$_i$であってもよい。学習済みモデルは、例えば、CLIPに対応するtext decoderなどの階層型ネットワークであってもよい。プロンプト生成部１３３は、クエリＤ３、補完情報Ｄ５、要約文Ｄ６、コンテキスト言語を組み合わせて、クエリＤ３により指示された資料を、補完情報Ｄ５、要約文Ｄ６、コンテキスト言語を参考にして、文書生成部１６０に命令クエリＤ１０の生成を指示するテキストデータである生成クエリＤ４を生成する。具体的には、プロンプト生成部１３３は、クエリＤ３、補完情報Ｄ５、要約文Ｄ６、コンテキスト言語を、高次元特徴のものはデコードしてテキスト情報にし、テキスト結合して生成クエリＤ４を生成する。クエリＤ３、補完情報Ｄ５、要約文Ｄ６、コンテキスト言語を、言語情報にし、テキスト結合することで、文書生成部１６０の対話付文書生成モデルに入力可能となる。例えば、画像認識部１１０が、画像をＶｉＴなどのトランスフォーマーエンコーダーを使って高次元特徴とした場合は、ここでデコードして言語情報に復元しておくことで、対話付文書生成モデルが理解できる言語情報に揃えておく。

[0035] 図４は、本実施形態における情報精錬部１４０の構成例を示す概略ブロック図である。情報精錬部１４０は、情報統制部１４１、対話付文書生成型推論部１４２、１４３、情報補完部１４４、情報要約部１４５を備える。情報統制部１４１は、対話付文書生成型推論部１４２、１４３の間で相互に入出力を繰り返して対話させる。このとき、情報統制部１４１は、クエリＤ３を初期入力とし、対話付文書生成型推論部１４２には、詳細な文書の推論を要求し、対話付文書生成型推論部１４３には、簡潔な文書の推論を要求する。なお、情報統制部１４１は、対話付文書生成型推論部１４２に詳細な文書の

推論を要求する際には、例えば、「以下の文章について詳細な情報を教えてください。」、「次のテキストに関する具体的な詳細を提供してください。」、「この文脈での次の情報を提供してください。」、「これについて詳しく説明してください。」、上記の内容を補足する情報を提供してください。」などの言語（テキスト）を用いてもよい。また、情報統制部１４１は、対話付文書生成型推論部１４２に簡潔な文書の推論を要求する際には、「以下の文章を簡潔にまとめてください。」、「次のテキストの要点を教えてください。」、「この文脈の要約を提供してください。」、「これについて簡潔に説明してください。」、「上記の内容の要点を教えてください。」などの言語を用いてもよい。

[0036] 情報統制部１４１は、対話の繰り返しを、予め決められた回数で終了させてもよいし、繰り返しても推論結果の変化がなくなったと判断したときに終了させてもよい。情報統制部１４１は、対話付文書生成型推論部１４２または対話付文書生成型推論部１４３による対話内容および記憶部１５０から嗜好情報を抽出するようにしてもよい。具体的には、情報統制部１４１は、対話付文書生成型推論部１４３あるいは１４２に文書の推論を要求する際に、対話付文書生成型推論部１４２あるいは１４３が出力したテキストに、Ａ．「ユーザの嗜好情報を抽出するために、SNS上の投稿から以下の要素を教えてください：［興味関心のあるトピック、好きな活動、好きな音楽や映画のジャンルなど］。」、Ｂ．「SNS上でのユーザの嗜好情報を教えてください。具体的には、好きなものや興味関心のあるトピックなどについて教えてください。」、Ｃ．「SNSからユーザの嗜好情報を抽出するために、好きなものや関心のあるトピック、よく投稿している内容などを教えてください。」、Ｄ「［趣味や興味関心］について説明し、それらが日常生活にどのように影響を与えているか教えてください。」、Ｅ「お気に入りの本や映画、音楽について語り、なぜ心に響くのか教えてください。」、Ｆ「製品やサービスを選ぶ際の意思決定プロセスと、どのような要因が最も重要なのかを説明してください。」、Ｇ「最近直面した課題について説明し、個人的な価値観に基づい

てどのように対処したかを教えてください。」、Ｈ「理想的な一日や完璧な経験を想像してください。それがどんな内容で、なぜそれがあなたにとって重要なのかを教えてください。」、Ｉ「最近の達成や成果について語り、その経験から学んだことを教えてください。」などの言語（テキスト）を付加して、特定の個人の嗜好あるいは思考パターンが含まれる嗜好情報が推論結果に含まれるようにする。

[0037] また、情報統制部１４１は、対話付文書生成型推論部１４３あるいは１４２に文書の推論を要求する際に、Ｊ．「最近の［特定のトピック］に関連する経験や思考を教えてください。」、Ｋ．「［最近のニュースやイベント］についての意見や感情を教えてください。」、Ｌ．「［特定の領域またはドメイン］での課題に対処する方法について説明してください。」、Ｍ．「ChatGPT（登録商標）や他のAIシステムを使う際に、最もモチベーションが高まる要因は何ですか？」、Ｎ．「ChatGPT（登録商標）との理想的な会話を想像してください。どのようなトピックを話し合い、その相互作用から何を期待しますか？」などの質問と、該質問へのユーザの回答のペアを付加して、特定の個人の思考パターンが含まれる嗜好情報が推論結果に含まれるようにする。なお、［］内の文言は、予め記憶されていてもよいし、記憶部１５０が記憶するＳＮＳへの投稿内容から、情報統制部１４１、対話付文書生成型推論部１４３あるいは１４２により抽出されてもよい。

[0038] なお、特定の個人の嗜好パターンは、バックエンドの履歴解析で取得することもでき、例えば以下のような方法での思考パターンの取得が考えられる。

１）チャット履歴の分析：特定の個人の過去のチャット履歴を分析して、特定のトピックへの興味関心や好み、特有の表現や言い回し、傾向を把握することができる。

２）自然言語処理技術：自然言語処理（NLP）の技術を活用して、特定の個人の文章から感情や態度、興味関心などを抽出することができる。

３）パーソナライゼーション：特定の個人の回答に基づいて、特定のトピッ

クや情報に対して個別化された応答を提供することで、特定の個人の思考パターンを把握しやすくなる。

４）クラスタリング：類似した回答を持つ個人をグループ化するクラスタリング技術を用いて、共通の特性や傾向を見つけることができる。

５）コンテキストの理解：特定の個人の会話のコンテキストを理解し、それに基づいて応答を生成することで、より適切な対話が可能になる。

[0039] また、クエリＤ３から抽出した言語に基づき、嗜好情報が抽出されるようにしてもよい。例えば、上記ＡやＤのかっこ内に含める言語については、クエリＤ３から抽出した言語であってもよい。画像入力部１００に画像などが入力されていた場合、情報統制部１４１が物体認識やVQA（visual question answering）を利用して該画像中の物体の固有名詞を予測し、クエリＤ３にその名称やコンテキストの説明を文書として付与して、その名称が上記Ａのかっこ内に記載されるようにしてもよい。これにより、情報統制部１４１は、自動で嗜好対象を指定することができる。また、クエリＤ３に固有名詞が含まれている場合、嗜好情報を含む固有名詞や文章のコンテキストが上記Ａのかっこ内に記載されるようにしてもよい。これにより、嗜好対象が与えられ、嗜好情報が自動で獲得できる。

[0040] ここで、嗜好は、例えば、興味関心のあるトピック、好きな活動、好きな音楽のジャンル、好きな映画およびドラマのジャンル、食べ物の好み、趣味、特技、観光地あるいは旅行先の好み、ファッションのスタイルである。また、思考パターンは、ユーザなど、生成された資料を読むことが想定される人物が取りやすい思考傾向を示す情報であり、例えば、論理的思考、直感的思考、具体的思考、嘲笑的思考、分析的思考、総合的思考のいずれかを示す情報である。論理的思考の人物は、論理的に情報を分析し、合理的な判断を下す傾向があり、論理的な課題解決や論証を好むという特徴がある。直感的思考の人物は、直感や感情に基づいて情報を処理し、直感的な判断を下す傾向があり、直感的な洞察や創造的な問題解決を好むという特徴がある。具体的思考の人物は、具体的な事実や詳細な情報に基づいて思考し、具体的なア

プローチを好み、具体的な例や具体的な手順に重点を置くという特徴がある。抽象的思考の人物は、大局的な視点や抽象的なアイデアに基づいて思考し、一般的な原則やパターンを見つけることに興味があるという特徴がある。分析的思考の人物は、問題を分析し、細かい部分や関連する要素を理解するために詳細に情報を調査し、論理的な推論やデータ分析に重点を置くという特徴がある。総合的思考の人物は、複数の視点や要素を総合的に考慮し、全体像を捉えることに興味があり、情報の統合や総合的な解決策を求めるという特徴がある。ＳＮＳは、例えば、Facebook（登録商標）、instagram（登録商標）、twitter（登録商標）、LINE（登録商標）などのいずれであってもよい。

[0041] 情報統制部１４１は、対話付文書生成型推論部１４２または対話付文書生成型推論部１４３に対して、記憶部１５０から嗜好情報を抽出するように指示してもよい。具体的には、情報統制部１４１は、対話付文書生成型推論部１４３あるいは１４２に、上記Ａ、Ｂ、Ｃなどのテキストを入力することにより、記憶部１５０が記憶するＳＮＳへの投稿内容から、特定の個人の嗜好情報を抽出する。情報統制部１４１は、記憶部１５０から嗜好情報を抽出するように指示する際に、クエリＤ３から抽出した語句を含むＳＮＳへの投稿内容を、対話付文書生成型推論部１４２または対話付文書生成型推論部１４３に入力するようにしてもよい。また、情報統制部１４１は、記憶部１５０ではなく、ＳＮＳのサーバなど外部の装置から嗜好情報を抽出してもよい。

[0042] 対話付文書生成型推論部１４２、１４３の各々は、他方の出力を、対話付き文書生成モデルに入力する。対話付文書生成型推論部１４２と、対話付文書生成型推論部１４３とで、使用する対話付き文書生成モデルは異なっていてもよい。情報補完部１４４は、嗜好情報を補完情報Ｄ５とする。なお、情報補完部１４４は、対話内容から抽出した嗜好情報と、記憶部１５０から抽出した嗜好情報とを結合して補完情報Ｄ５としてもよい。情報要約部１４５は、対話付文書生成型推論部１４２、１４３の推論結果の要約を作成し、要約文Ｄ６、言語回答Ｄ９とする。要約の作成は、要約作成用の学習済みモデ

ルを用いて行ってもよいし、対話付き文書生成モデルに要約の作成を指示することで行ってもよい。

[0043] 図５は、本実施形態における画像生成部１７０の構成例を示す概略ブロック図である。画像生成部１７０は、学習済深層特徴量抽出部１７１、画像デコーダ部１７２を備える。学習済深層特徴量抽出部１７１は、命令クエリＤ１０を、学習済みモデルに入力して、その特徴量を推論させる。この学習済みモデルは、例えば、CLIPのtext encoderなどの階層型ネットワークであってもよい。画像デコーダ部１７２は、学習済深層特徴量抽出部１７１による推論結果の特徴量を、学習済みモデルに入力して該特徴量を表す画像である画像回答Ｄ１１を推論する。この学習済みモデルは、例えば、CLIPに対応するdiffusion decoderなどの階層型ネットワークであってもよい。

[0044] なお、学習済深層特徴量抽出部１７１は、推論結果の特徴量から重要な重みを持つ特徴量を特定し、画像デコーダ部１７２は、該特定された特徴量を学習済みモデルに入力して該特定された特徴量を表す画像を推論し、画像回答Ｄ１１としてもよい。

[0045] 図６は、本実施形態における資料生成装置１０の動作例を説明するフローチャートである。まず、言語入力部１２０が、ユーザによるクエリＤ３の入力を受け付ける（ステップＳ１）。次に、画像入力部１００が、ユーザによる画像Ｄ１の入力を受け付ける（ステップＳ２）。

[0046] 次に、情報精錬部１４０が、ステップＳ１で言語入力部１２０に入力されたクエリＤ３を初期入力として、２つの対話付き文書生成モデルの間で相互に入出力を繰り返して対話させることにより、クエリＤ３により要求された文書を推論し、推論結果を要約して言語回答Ｄ９と、補完情報Ｄ５とする（ステップＳ３）。このとき、情報精錬部１４０は、記憶部１５０から読みだした特定の個人の嗜好情報が推論結果に反映されるように、２つの対話付き文書生成モデルのうち、一方に入力してもよい。また、情報精錬部１４０は、該嗜好情報を補完情報Ｄ５とする。

[0047] 次に、画像認識部１１０が、ステップＳ２で画像入力部１００に入力され

た画像Ｄ１を、学習済みモデルに入力して、その特徴量を推論させ、推論結果の特徴量から重要な重みを持つ特徴量を特定して、画像由来コンテキスト特徴Ｄ２とする（ステップＳ４）。

[0048] 次に、自然言語処理部１３０は、クエリＤ３を、学習済みモデルに入力して、その特徴量を推論させ、推論結果の特徴量から重要な重みを持つ特徴量（言語由来コンテキスト特徴）を特定する。さらに、自然言語処理部１３０は、クエリＤ３、言語由来コンテキスト特徴、画像由来コンテキスト特徴Ｄ２、補完情報Ｄ５、要約文Ｄ６を組み合わせて生成クエリＤ４を生成する（ステップＳ５）。

[0049] 次に、文書生成部１６０が、ステップＳ５で生成された生成クエリＤ４を、対話付文書生成モデルに入力し、生成クエリＤ４により要求された文書を推論し、推論結果を命令クエリＤ１０とする（ステップＳ６）。次に、画像生成部１７０が、ステップＳ６の命令クエリＤ１０を、学習済み画像生成モデルに入力して、命令クエリＤ１０が示す画像を推論させ、該推論結果を画像回答Ｄ１１とする（ステップＳ７）。

[0050] 次に、表示処理部１８０が、ステップＳ３の言語回答Ｄ９を情報精錬部１４０から、ステップＳ７の画像回答Ｄ１１を画像生成部１７０から取得し、これらをディスプレイ部１９０に表示させる（ステップＳ８）。なお、表示処理部１８０は、言語回答Ｄ９と画像回答Ｄ１１を電子ファイルとして出力してもよいし、他装置に送信してもよい。

[0051] 図７は、本実施形態における資料生成装置１０のユースケースを示す模式図である。図７では、ユーザにより、“I want to get a rabbit image with holding a carrot”というテキストＥ３がクエリＤ３として言語入力部１２０に入力される。また、ユーザにより、人参を抱えたウサギの画像Ｅ１が３つ、画像Ｄ１として画像入力部１００に入力される。ここで、画像Ｅ１におけるウサギは、灰色、ピンク、茶色である。

[0052] 情報精錬部１４０は、記憶部１５０が記憶するユーザのＳＮＳへの投稿履歴から、“She loves white rabbits”というテキストＥ８を嗜好情報として

抽出し、対話付文書生成型推論部１４２、１４３のいずれかに入力する。対話付文書生成型推論部１４２、１４３は、“What does look like a rabbit ?”、“Common rabbits are Lepus Timidus”というように対話Ｅ１２を行う。これらに基づき、情報要約部１４５が、“A white rabbit like Lepus Timidus with ears up, holding a tip up carrot and facing forward”というテキストＥ９を、言語回答Ｄ９あるいは要約文Ｄ６として生成する。

[0053] 文書生成部１６０が生成する命令クエリＤ１０には、テキストＥ９の内容が含まれているため、画像生成部１７０は、白いウサギが人参を抱えた画像Ｅ１１を、画像回答Ｄ１１として生成する。

[0054] ＜第２の実施形態＞

以下、図面を参照して、本開示の第２の実施形態について説明する。第１の実施形態では、人物の特徴を示す情報として、該人物の嗜好あるいは思考パターンが含まれる嗜好情報を用いた。第２の実施形態では、人物の特徴を示す情報として、人物の役割を示す情報を用いる。図８は、この開示の第２の実施形態による資料生成装置２０の構成を示す概略ブロック図である。資料生成装置１０は、ユーザの要求に応じたマルチモーダル（Multimodal）な資料を生成する装置である。資料生成装置１０は、ユーザインターフェース部２０１、言語解析部２０２、画像解析部２０３、マルチモーダル情報統合部２０４、ユーザ情報ＤＢ（Data Base）部２０５、提供者情報ＤＢ部２０６、外部情報ＤＢ部２０７、仮説生成部２０８、人工知能部２０９、表示処理部２１２を備える。資料生成装置１０は、１つまたは複数のコンピュータがプログラムを読み込み実行することで実現されてもよいし、その少なくとも一部がいわゆるクラウド上に配置されていてもよい。

[0055] ユーザインターフェース部２０１は、キーボード、マウス、タッチパネル、マイクロフォン、ディスプレイ、スピーカなどの入出力デバイスを備え、資料生成装置１０にデータを入力するためのユーザインターフェースを提供する。入力するデータは、テキストデータ、音声データ、画像データ、ＰＤＦファイルなどの文書データ、ＵＲＬなどの参照先データを含むマルチモー

ダルなデータである。また、入力するデータは、生成する資料を指示するクエリ、事前に記憶させておき、資料を生成する際に参照させるユーザ情報、提供者情報を含む。ユーザ情報は、資料生成装置１０が提供する資料生成サービスのユーザが有する情報であり、例えば、ユーザの有するノウハウなどである。提供者情報は、資料生成装置１０による資料生成サービスを提供する提供者が有する情報であり、例えば、提供者の有するノウハウなどである。

[0056] 言語解析部２０２は、ユーザインターフェース部２０１に入力されたテキストデータを処理する。テキストデータは、ＰＤＦファイルなどの文書データに含まれるテキストデータであってもよいし、参照先データが示すコンテンツ（Ｗｅｂページコンテンツなど）に含まれるテキストデータであってもよい。また、言語解析部２０２は、ユーザインターフェース部２０１に入力された音声データを音声認識処理して、テキストデータに変換してもよい。

[0057] 画像解析部２０３は、ユーザインターフェース部２０１に入力された画像データを解析して、特徴量を抽出する。画像データは、ＰＤＦファイルなどの文書データに含まれる画像データであってもよいし、参照先データが示すコンテンツに含まれる画像データであってもよい。

[0058] マルチモーダル情報統合部２０４は、言語解析部２０２、画像解析部２０３、によるテキストデータと画像データの各特徴量を統合し、ベクトル表現に変換する。マルチモーダル情報統合部２０４は、ベクトル表現に変換したクエリについては、仮説生成部２０８に入力する。また、マルチモーダル情報統合部２０４は、ベクトル表現に変換したユーザ情報は、ユーザ情報ＤＢ部２０５に記憶させ、ベクトル表現に変換した提供者情報は、提供者情報ＤＢ部２０６に記憶させる。

[0059] ユーザ情報ＤＢ部２０５、提供者情報ＤＢ部２０６、外部情報ＤＢ部２０７は、資料に関連する情報を記憶する。ユーザ情報ＤＢ部２０５は、ベクトル表現に変換されたユーザ情報を記憶するデータベースである。提供者情報ＤＢ部２０６は、ベクトル表現に変換された提供者情報を記憶するデータベ

ースである。外部情報ＤＢ部２０７は、インターネットなど外部に配置されたデータベース、または外部から取得したデータを記憶するデータベースである。

[0060] 仮説生成部２０８は、マルチモーダル情報統合部２０４から入力されたベクトル表現に変換したクエリが示す資料を生成する。仮説生成部２０８は、複数のＡＩエージェントの間で対話させることにより、クエリにより指示された資料を推論する。仮説生成部２０８は、複数のＡＩエージェントによる対話に、複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する人物の特徴を示す情報を含める。この対話は、仮説生成（Generation）側と批判（Discrimination）側との対話であってもよい。例えば、仮説生成部２０８は、１つのＡＩエージェントには、仮説を生成させ、他のＡＩエージェントには、該仮説を批判または批評させてもよい。これにより、仮説生成部２０８は、仮説の精度を上げて、生成する資料の質を上げられる。さらに、仮説生成部２０８は、批判または批評の結果に基づき、仮説を生成したＡＩエージェントに仮説を修正させてもよい。

[0061] また、仮説生成部２０８は、資料を生成する際に、仮説生成部２０８は、ユーザ情報ＤＢ部２０５、提供者情報ＤＢ部２０６、外部情報ＤＢ部２０７が記憶する情報を参照する。また、資料を生成する際に、仮説生成部２０８は、人工知能の機能を有する人工知能部２０９を用いる。

[0062] 人工知能部２０９は、推論、判断等の知的な機能を備えた人工知能およびその動作環境をいう。人工知能部２０９は、モデル制御部２１０と学習済モデル記憶部２１１とにより構成される。人工知能部２０９は、プロンプトが入力されると、プロンプトに対する回答を出力するように構成されたモデルおよびその動作環境である。人工知能部２０９は、仮説生成部２０８からプロンプトが入力されると、プロンプト、および後述する学習済モデルに基づいて、回答を出力する。

[0063] 学習済モデル記憶部２１１は、学習済モデルを記憶する。

[0064] 学習済モデルは、後述するモデルの情報を含む。学習済モデルは、例えば

、制約条件、重みづけ変数および評価関数等、モデルの挙動を規定する情報であるモデルパラメータを含んでいてもよい。

[0065] モデルは、例えば、ＮＮ（Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋ）、ＣＮＮ（Ｃｏｎｖｏｌｕｔｉｏｎａｌ Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋ）、ＲＮＮ（Ｒｅｃｕｒｒｅｎｔ Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋ）、ＶＡＥ（Ｖａｒｉａｔｉｏｎａｌ Ａｕｔｏｅｎｃｏｄｅｒ）、ＧＡＮ（Ｇｅｎｅｒａｔｉｖｅ Ａｄｖｅｒｓａｒｉａｌ Ｎｅｔｗｏｒｋｓ）、拡散モデル（Ｄｉｆｆｕｓｉｏｎ ｍｏｄｅｌ）、Ｔｒａｎｓｆｏｒｍｅｒ、ＬＬＭ（Ｌａｒｇｅ Ｌａｎｇｕａｇｅ Ｍｏｄｅｌ）、ＶＬＭ（Ｖｉｓｕａｌ Ｌａｎｇｕａｇｅ Ｍｏｄｅｌ）、ＢＥＲＴ（Ｂｉｄｉｒｅｃｔｉｏｎａｌ Ｅｎｃｏｄｅｒ Ｒｅｐｒｅｓｅｎｔａｔｉｏｎｓ ｆｒｏｍ Ｔｒａｎｓｆｏｒｍｅｒｓ）、ＧＰＴ（Ｇｅｎｅｒａｔｉｖｅ Ｐｒｅ－ｔｒａｉｎｅｄ Ｔｒａｎｓｆｏｒｍｅｒ）、ＣＬＩＰ（Ｃｏｎｔｒａｓｔｉｖｅ Ｌａｎｇｕａｇｅ ＩｍａｇｅＰｒｅ－ｔｒａｉｎｉｎｇ）と呼ばれるモデルであってもよい。なお、上述したモデルは排他的な物ではなく、例えば、ＬＬＭ、ＶＬＭ、ＢＥＲＴ、ＧＰＴは、Ｔｒａｎｓｆｏｒｍｅｒに含まれる。また、例えば、Ｔｒａｎｓｆｏｒｍｅｒは、ＮＮに含まれる。また、学習アルゴリズムおよびモデルは複数の種類を組み合わせたものであってもよい。モデルには、複数の異なる種類のデータを組み合わせて学習したマルチモーダルモデルと呼ばれるものも含まれる。

[0066] モデル制御部２１０は、プロンプトを取得すると、プロンプト、および学習済モデルに基づいて、プロンプトに対応する回答を出力する。すなわち、モデル制御部は２１０、プロンプトを取得すると、学習済モデルにより示されるモデルを用いて、プロンプトに対応する回答を生成して出力する。

[0067] なお、学習済モデル、および人工知能部２０９が用いるその他の情報は、予め用意されたものであってもよいし、必要に応じて、ネットワークを介して取得されるものであってもよい。

[0068] 表示処理部２１２は、液晶ディスプレイ、有機ＥＬ（Electro Luminescence）ディスプレイなどのディスプレイを備え、仮説生成部２０８が生成した資

料を表示する。なお、表示処理部２１２が備えるディスプレイは、ユーザインターフェース部２０１が備えるディスプレイと同一であってもよい。

[0069] 図９は、本実施形態における仮説生成部２０８の構成を示す概略ブロック図である。仮説生成部２０８は、３つのＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃ（ＡＩエージェント）、ＡＩエージェント対話管理部３０２（対話管理部）、対話履歴知識ＤＢ部３０３、対話戦略ＤＢ部３０４（対話戦略記憶部）、アウトプット生成部３０５、評価検証部３０６を備える。図９では、３つのＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃを示したが、仮説生成部２０８が備えるＡＩエージェント部の数は、２つであってもよいし、４つ以上であってもよい。

[0070] ３つのＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃは、一対一に３人の人物に対応する。以降、３つのＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃから選択されたＡＩエージェント部など、任意のＡＩエージェント部をＡＩエージェント部３０１ともいう。３つのＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの各々は、該３人の人物のうち、対応する人物の特徴を示す人物情報を対話に含める。該３人の人物の少なくとも一人は、仮想の人物であってもよい。仮想の人物の人物情報は、該仮想の人物の役割を示す情報を含んでもよい。仮想の人物の役割は、例えば、優秀なビジネスアナリスト、デジタル化を推進する提供者の優秀なコンサルタント、ユーザのＸ事業部の経営幹部であってもよい。なお、３人の人物に、仮想ではない特定の人物が含まれている場合、該特定の人物の人物情報は、該特定の人物の役割を示す情報を含んでもよいし、該特定の人物の著作物、または該特定の人物に関する著作物を含んでもよい。

[0071] ＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの各々は、ユーザ情報ＤＢ部２０５、提供者情報ＤＢ部２０６、外部情報ＤＢ部２０７、対話履歴知識ＤＢ部３０３から対話に関連する参照情報を取得し、取得した参照情報を対話に含める。例えば、ＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの各々は、ユーザ情報ＤＢ部２０５、提供者情報ＤＢ部２０６、外部情報

ＤＢ部２０７、対話履歴知識ＤＢ部３０３から取得する参照情報と、タスクとを、ＡＩエージェント対話管理部３０２から指定される。ＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの各々は、指定された参照情報を、ユーザ情報ＤＢ部２０５、提供者情報ＤＢ部２０６、外部情報ＤＢ部２０７、対話履歴知識ＤＢ部３０３から取得する。ＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの各々は、取得した参照情報と、指定されたタスクと、対応する人物情報とを含むプロンプトを生成し、生成したプロンプトを人工知能部２０９のモデル制御部２１０に入力する。すなわち、ＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの各々は、指定されたタスクを、人物情報が示す人物が、取得した参照情報を参照して行った場合の回答を、モデル制御部２１０に要求する。ＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの各々は、モデル制御部２１０からの回答を、対話履歴知識ＤＢ部３０３に記憶させる。タスクは、仮説の生成、生成された仮説の批判または批評、批判または批評に基づく仮説の修正を含み、例えば、「参照情報を分析し、詳細なレポートを作成せよ」、「前のステップで出力されたレポートについて実現可能性を評価せよ」、「実現可能性の評価に基づきレポートを修正せよ」などである。また、タスクには、回答の具体的なフォーマットの指定も含まれていてもよい。

[0072] ＡＩエージェント対話管理部３０２は、対話戦略ＤＢ部３０４から取得した対話戦略情報が示す対話プロセスに従い３つのＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃから選択したＡＩエージェント部３０１に対話させる。ＡＩエージェント対話管理部３０２は、対話の結果をアウトプット生成部３０５に入力する。対話および対話の結果は、テキスト以外に、画像、図、表、グラフ、音声、動画などを含むマルチモーダルなデータであってもよい。対話プロセスは、複数のステップを含み、各ステップにおけるタスクと、ＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃのうちのいずれかの指定と、参照情報の指定とを含んでもよい。なお、ＡＩエージェント対話管理部３０２は、人工知能部２０９のモデル制御部２１０を用いて、対話戦略情

報に基づいた３つのＡＩエージェント部３０１ａ、３０１ｂ、３０１ｃの対話を行うようにしてもよい。例えば、ＡＩエージェント対話管理部３０２は、それまでのステップの実施履歴と、対話戦略情報とともに、次に実施すべきステップの内容を問い合わせるプロンプトをモデル制御部２１０に入力し、該プロンプトに対応する回答に従い、ＡＩエージェント部３０１に対話させてもよい。

[0073] ＡＩエージェント対話管理部３０２は、マルチモーダル情報統合部２０４から入力されたベクトル表現に変換されたクエリ（以降、クエリ情報ともいう）に基づき、対話戦略情報を対話戦略ＤＢ部３０４から取得してもよい。例えば、ＡＩエージェント対話管理部３０２は、クエリ情報で対話戦略ＤＢ部３０４を検索して、検索結果の対話戦略情報を取得するようにしてもよい。また、対話戦略の各ステップは、再度の対話戦略ＤＢ部３０４の検索による対話戦略情報の取得と、該検索に用いる情報の指定とを含んでいてもよい。検索に用いる情報は、例えば、クエリ情報、参照情報などである。

[0074] 対話履歴知識ＤＢ部３０３は、ＡＩエージェント部３０１がモデル制御部２１０から取得した回答を対話履歴として記憶する。対話履歴知識ＤＢ部３０３は、モデル制御部２１０に入力されたプロンプトも記憶していてもよい。

[0075] 対話戦略ＤＢ部３０４は、対話戦略情報を記憶する。対話戦略情報は、該対話戦略の目的と、対話プロセスとを含んでいてもよい。対話戦略ＤＢ部３０４が記憶する対話戦略情報は、例えば、「業務プロセスの改善提案を作成する。対話プロセスは、３つのステップからなる。１つめのステップは、ビジネスアナリストのＡＩエージェントに、ユーザの業務プロセスを収集および分析させ、その結果のレポートを提出させる。レポートのフォーマットは、．．．」というような自然言語による文章を、ベクトル表現に変換したものであってもよい。

[0076] アウトプット生成部３０５は、ＡＩエージェント対話管理部３０２から入力された対話の結果に基づき、資料を生成する。資料のフォーマットは、対

話戦略情報に基づき、ＡＩエージェント対話管理部３０２により指定されたものであってもよい。例えば、アウトプット生成部３０５は、ＡＩエージェント対話管理部３０２から入力された対話の結果と、資料のフォーマットとともに、資料の生成を指示するプロンプトをモデル制御部２１０に入力し、該プロンプトに対応する回答として資料を取得することで、資料を生成してもよい。該資料は、テキスト以外に、画像、図、表、グラフなどが含まれるマルチモーダルなデータであってもよい。

[0077] また、アウトプット生成部３０５は、生成した資料に対する評価検証部３０６による評価と検証の結果に基づき、該資料を修正してもよい。例えば、アウトプット生成部３０５は、生成した資料に対する評価検証部３０６による評価と検証の結果とともに、資料の修正を指示するプロンプトをモデル制御部２１０に入力し、該プロンプトに対応する回答として資料を取得することで、資料を修正してもよい。アウトプット生成部３０５は、修正した資料を表示処理部２１２に入力する。なお、アウトプット生成部３０５は、評価検証部３０６による評価および検証と、修正とを複数回繰り返してから、修正した資料を表示処理部２１２に入力してもよい。

[0078] 評価検証部３０６は、アウトプット生成部３０５から入力された資料を評価および検証する。例えば、評価検証部３０６は、アウトプット生成部３０５から入力された資料とともに、資料の評価および検証を指示するプロンプトをモデル制御部２１０に入力し、該プロンプトに対応する回答として評価および検証の結果を取得してもよい。また、評価検証部３０６は、プロンプトに、例えば、校閲者などを示す人物情報を含めてもよい。プロンプトに含める人物情報は、評価および検証の度に変化してもよい。

[0079] このように、資料生成装置２０において、複数のＡＩエージェント部３０１は、一対一に複数の人物に対応し、複数のＡＩエージェント部３０１の各々は、複数の人物のうち、対応する人物の特徴を示す人物情報を対話に含める。これにより、資料生成装置２０は、複数の人物の特徴に応じた資料を生成できる。

[0080] 図１０は、本実施形態に係る資料生成装置１０、２０のハードウェア構成を説明する説明図である。

資料生成装置１０、２０は、入出力モジュールＩ、記憶モジュールＭ、及び制御モジュールＰを含んで構成される。入出力モジュールＩは、通信モジュールＨ１１、接続モジュールＨ１２、ポインティングデバイスＨ２１、キーボードＨ２２、ディスプレイＨ２３、ボタンＨ３、マイクＨ４１、スピーカＨ４２、カメラＨ５１、又はセンサＨ５２の一部或いは全部を含んで実現される。記憶モジュールＭは、ドライブＨ７を含んで実現される。記憶モジュールＭは、さらに、メモリＨ８の一部或いは全部を含んで構成されてもよい。制御モジュールＰは、メモリＨ８及びプロセッサＨ９を含んで実現される。これらのハードウェア構成要素は、バス（Ｂｕｓ）を介して、相互に通信可能に接続されるとともに、電源Ｈ６から電力を供給されている。

[0081] 接続モジュールＨ１２は、ＵＳＢ（Ｕｎｉｖｅｒｓａｌ Ｓｅｒｉｕｌ Ｂｕｓ）等のデジタル入出力ポートである。携帯機器の場合、ポインティングデバイスＨ２１、キーボードＨ２２、及びディスプレイＨ２３は、タッチパネルである。センサＨ５２は、加速度センサ、ジャイロセンサ、ＧＰＳ受信モジュール、近接センサ等である。電源Ｈ６は、各装置を動かすために必要な電気を供給する電源ユニットである。携帯機器の場合、電源Ｈ６は、バッテリーである。ドライブＨ７は、ハードディスクドライブやソリッドステートドライブ等の補助記憶媒体である。ドライブＨ７は、ＥＥＰＲＯＭやフラッシュメモリ等の不揮発性メモリ、又は、光磁気ディスクドライブやフレキシブルディスクドライブであってもよい。また、ドライブＨ７は、例えば、各装置に内蔵されるものに限らず、接続モジュールＨ１２のコネクタに接続された外付け型の記憶装置でもよい。メモリＨ８は、ランダムアクセスメモリ等の主記憶媒体である。なお、メモリＨ８は、キャッシュメモリであってもよい。メモリＨ８は、一又は複数のプロセッサＨ９によって命令が実行されるときに、これらの命令を格納する。プロセッサＨ９は、ＣＰＵ（中央演算装置）である。プロセッサＨ９は、ＭＰＵ（マイクロプロセッシングユ

ニット）又はＧＰＵ（グラフィックスプロセッシングユニット）であってもよい。プロセッサＨ９は、メモリＨ８を介してドライブＨ７から、プログラム及び各種データを読み出して演算を行うことで、一又は複数のメモリＨ８に格納した命令を実行する。

[0082] なお、本明細書等において、資料生成装置１０、２０との記載は、制御モジュールＰとの記載に置き換えられてもよい。

[0083] また、以下のような実施形態であってもよい。

（１）本開示の一実施形態は、生成する資料を指示するクエリの入力を受け付ける入力部と、前記クエリに基づき、複数のＡＩエージェントの間で対話させることにより、前記クエリにより指示された資料を推論する対話管理部とを備え、前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する人物の特徴を示す情報を含める資料生成装置である。

[0084] これにより、資料生成装置は、人物の特徴に応じた資料を生成できる。

[0085] （２）また、本開示の他の一実施形態は、（１）に記載の資料生成装置であって、前記複数のＡＩエージェントは、２つの対話付き文書生成モデルであり、前記入力部は、前記クエリの入力を受け付ける言語入力部であり、前記対話管理部は、前記クエリを初期入力として、２つの対話付き文書生成モデルの間で相互に入出力を繰り返して対話させることにより、前記クエリにより指示された資料のうちの文書を推論する情報精錬部であり、前記情報精錬部による推論結果の文書に基づき、画像を生成する画像生成部を備え、前記２つの対話付き文書生成モデルによる対話に、特定の個人の嗜好または思考パターンが含まれる嗜好情報を加える、資料生成装置である。

[0086] これにより、資料生成装置は、特定の個人の嗜好または思考パターンが反映された資料を生成することができる。

[0087] （３）本開示の他の一実施形態は、（２）に記載の資料生成装置であって、前記生成する資料の事例となる画像の入力を受け付ける画像入力部と、前記画像入力部に入力された画像を、学習済みモデルに入力して、前記画像入力

部に入力された画像の特徴量を推論させて、画像由来コンテキスト特徴とする画像認識部とをさらに備え、前記画像生成部は、前記情報精錬部による推論結果の文書に加えて、前記画像由来コンテキスト特徴に基づき、画像を生成する、資料生成装置である。

[0088] これにより、資料生成装置は、資料の事例となる画像が反映された画像を生成することができる。

[0089] （４）本開示の他の一実施形態は、（３）に記載の資料生成装置であって、前記クエリを、学習済みモデルに入力して、前記クエリの特徴量を推論させて言語由来コンテキスト特徴とし、前記クエリ、前記言語由来コンテキスト特徴、前記画像由来コンテキスト特徴、前記嗜好情報、前記情報精錬部による推論結果の文書の少なくとも一部を組み合わせた生成クエリを生成する自然言語処理部と、前記生成クエリを、対話付文書生成モデルに入力し、前記生成クエリにより要求された文書を推論し、推論結果を命令クエリとする文書生成部とをさらに備え、前記画像生成部は、前記命令クエリを学習済み画像生成モデルに入力して、前記命令クエリが示す画像を推論させる、資料生成装置である。

[0090] これにより、資料生成装置は、特定の個人の嗜好または思考パターンと、資料の事例となる画像とが反映された画像を生成することができる。

[0091] （５）本開示の他の一実施形態は、（１）に記載の資料生成装置であって、前記複数のＡＩエージェントによる対話は、１つのＡＩエージェントによる仮説の生成と、生成された前記仮説に対する他のＡＩエージェントによる批判または批評を含む。

[0092] これにより、資料生成装置は、生成する資料の質を上げられる。

[0093] （６）本開示の他の一実施形態は、（１）または（５）に記載の資料生成装置であって、前記複数のＡＩエージェントは、一対一に複数の人物に対応し、前記複数のＡＩエージェントの各々は、前記複数の人物のうち、対応する人物の特徴を示す人物情報を対話に含める。

[0094] これにより、資料生成装置は、複数の人物の特徴に応じた資料を生成でき

る。

[0095] （７）本開示の他の一実施形態は、（６）に記載の資料生成装置であって、前記複数の人物の少なくとも一人は、仮想の人物であり、前記仮想の人物の前記人物情報は、前記仮想の人物の役割を示す情報を含む。

[0096] これにより、資料生成装置は、仮想の人物の役割に応じた資料を生成できる。

[0097] （８）本開示の他の一実施形態は、（１）から（７）のいずれかに記載の資料生成装置であって、対話プロセスと、対話させる人物とを示す対話戦略情報を記憶する対話戦略記憶部を備え、前記対話管理部は、前記対話戦略記憶部から取得した対話戦略情報が示す前記対話プロセスに従い前記複数のＡＩエージェントから選択した前記ＡＩエージェントに対話させる。

[0098] これにより、資料生成装置は、対話プロセスを柔軟にすることができる。

[0099] （９）本開示の他の一実施形態は、（１）から（８）のいずれかに記載の資料生成装置であって、前記資料に関連する情報を記憶する記憶部を備え、前記複数のＡＩエージェントの各々は、前記記憶部から前記対話に関連する参照情報を取得し、取得した前記参照情報を前記対話に含める。

[0100] これにより、資料生成装置は、対話に関連する参照情報に基づき資料を生成することができる。

[0101] （１０）本開示の他の一実施形態は、（１）から（９）のいずれかに記載の資料生成装置であって、前記入力部は、前記資料に関連するマルチモーダルなデータの入力を受け付け、前記対話管理部は、前記クエリおよび前記マルチモーダルな情報に基づき、前記複数のＡＩエージェントの間で対話させ、前記複数のＡＩエージェントによる対話は、マルチモーダルである。

[0102] これにより、資料生成装置は、マルチモーダルな情報に基づき資料を生成することができる。

[0103] （１１）本開示の他の一実施形態は、（１）から（１０）のいずれかに記載の資料生成装置であって、前記複数のＡＩエージェントによる対話の結果に基づき、前記資料を生成するアウトプット生成部と、前記アウトプット生成

部が生成した資料を評価および検証する評価検証部とを備える。

[0104] これにより、資料生成装置は、評価および検証された資料を生成できる。

[0105] （１２）本開示の他の一実施形態は、生成する資料を指示するクエリの入力を受け付ける第１のステップと、前記クエリに基づき、複数のＡＩエージェントの間で対話させることにより、前記クエリにより指示された資料を推論する第２のステップとを有し、前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する仮想を含む人物の特徴を示す情報を加える、資料生成方法である。

[0106] （１３）本開示の他の一実施形態は、コンピュータを、生成する資料を指示するクエリの入力を受け付ける入力部、前記クエリに前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する仮想を含む人物の特徴を示す情報を加える、プログラムである。

[0107] また、図１における資料生成装置１０の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより資料生成装置１０を実現してもよい。なお、ここでいう「コンピュータシステム」とは、ＯＳや周辺機器等のハードウェアを含むものとする。

[0108] また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ＲＯＭ、ＣＤ－ＲＯＭ等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含むものとする。また上記プログラムは、前述した機能の一部を実現するためのものであっても良く、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムと

の組み合わせで実現できるものであっても良い。

[0109] 以上、この開示の実施形態を図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この開示の要旨を逸脱しない範囲の設計変更等も含まれる。

符号の説明

[0110] １０、２０　資料生成装置

１００　画像入力部

１１０　画像認識部

１１１　学習済深層特徴量抽出部

１１２　コンテキスト解析部

１２０　言語入力部

１３０　自然言語処理部

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

１４０　情報精錬部

１４１　情報統制部

１４２、１４３　対話付文書生成型推論部

１４４　情報補完部

１４５　情報要約部

１５０　記憶部

１６０　文書生成部

１７０　画像生成部

１７１　学習済深層特徴量抽出部

１７２　画像デコーダ部

１８０　表示処理部

１９０　ディスプレイ部

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

３０１ａ、３０１ｂ、３０１ｃ　ＡＩエージェント部

３０２　ＡＩエージェント対話管理部

３０３　対話履歴知識ＤＢ部

３０４　対話戦略ＤＢ部

３０５　アウトプット生成部

３０６　評価検証部

請求の範囲

[請求項1] 生成する資料を指示するクエリの入力を受け付ける入力部と、

前記クエリに基づき、複数のＡＩ（Artificial Intelligence）エージェントの間で対話させることにより、前記クエリにより指示された資料を推論する対話管理部と

を備え、

前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する人物の特徴を示す情報を含める、

資料生成装置。

[請求項2] 前記複数のＡＩエージェントは、複数の対話付き文書生成モデルであり、

前記入力部は、前記クエリの入力を受け付ける言語入力部であり、

前記対話管理部は、前記クエリを初期入力として、前記複数の対話付き文書生成モデルの間で相互に入出力を繰り返して対話させることにより、前記クエリにより指示された資料のうちの文書を推論する情報精錬部であり、

前記情報精錬部による推論結果の文書に基づき、画像を生成する画像生成部を備え、

前記複数の対話付き文書生成モデルによる対話に、特定の個人の嗜好あるいは思考パターンが含まれる嗜好情報を加える、

請求項１に記載の資料生成装置。

[請求項3] 前記生成する資料の事例となる画像の入力を受け付ける画像入力部と、

前記画像入力部に入力された画像を、学習済みモデルに入力して、前記画像入力部に入力された画像の特徴量を推論させて、画像由来コンテキスト特徴とする画像認識部と

をさらに備え、

前記画像生成部は、前記情報精錬部による推論結果の文書に加えて、前記画像由来コンテキスト特徴に基づき、画像を生成する、

請求項２に記載の資料生成装置。

[請求項4] 前記クエリを、学習済みモデルに入力して、前記クエリの特徴量を推論させて言語由来コンテキスト特徴とし、前記クエリ、前記言語由来コンテキスト特徴、前記画像由来コンテキスト特徴、前記嗜好情報、前記情報精錬部による推論結果の文書の少なくとも一部を組み合わせた生成クエリを生成する自然言語処理部と、

前記生成クエリを、対話付文書生成モデルに入力し、前記生成クエリにより要求された文書を推論し、推論結果を命令クエリとする文書生成部と

をさらに備え、

前記画像生成部は、前記命令クエリを学習済み画像生成モデルに入力して、前記命令クエリが示す画像を推論させる、

請求項３に記載の資料生成装置。

[請求項5] 前記複数のＡＩエージェントによる対話は、１つのＡＩエージェントによる仮説の生成と、生成された前記仮説に対する他のＡＩエージェントによる批判または批評を含む、請求項１に記載の資料生成装置。

[請求項6] 前記複数のＡＩエージェントは、一対一に複数の人物に対応し、

前記複数のＡＩエージェントの各々は、前記複数の人物のうち、対応する人物の特徴を示す人物情報を対話に含める、

請求項１に記載の資料生成装置。

[請求項7] 前記複数の人物の少なくとも一人は、仮想の人物であり、

前記仮想の人物の前記人物情報は、前記仮想の人物の役割を示す情報を含む、請求項６に記載の資料生成装置。

[請求項8] 対話プロセスと、対話させる人物とを示す対話戦略情報を記憶する対話戦略記憶部を備え、

前記対話管理部は、前記対話戦略記憶部から取得した対話戦略情報が示す前記対話プロセスに従い前記複数のＡＩエージェントから選択した前記ＡＩエージェントに対話させる、

請求項１に記載の資料生成装置。

[請求項9] 前記資料に関連する情報を記憶する記憶部を備え、

前記複数のＡＩエージェントの各々は、前記記憶部から前記対話に関連する参照情報を取得し、取得した前記参照情報を前記対話に含める、

請求項１に記載の資料生成装置。

[請求項10] 前記入力部は、前記資料に関連するマルチモーダルなデータの入力を受け付け、

前記対話管理部は、前記クエリおよび前記マルチモーダルな情報に基づき、前記複数のＡＩエージェントの間で対話させ、

前記複数のＡＩエージェントによる対話は、マルチモーダルである、

請求項１に記載の資料生成装置。

[請求項11] 前記複数のＡＩエージェントによる対話の結果に基づき、前記資料を生成するアウトプット生成部と、

前記アウトプット生成部が生成した資料を評価および検証する評価検証部と

を備える、請求項１に記載の資料生成装置。

[請求項12] 生成する資料を指示するクエリの入力を受け付ける第１のステップと、

前記クエリに基づき、複数のＡＩエージェントの間で対話させることにより、前記クエリにより指示された資料を推論する第２のステップと

を有し、

前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージ

ェントのうち、少なくとも１つのＡＩエージェントに対応する仮想を含む人物の特徴を示す情報を加える、

資料生成方法。

[請求項13] コンピュータを、

生成する資料を指示するクエリの入力を受け付ける入力部、

前記クエリに基づき、複数のＡＩエージェントの間で対話させることにより、前記クエリにより指示された資料を推論する対話管理部

として機能させるためのプログラムであって、

前記複数のＡＩエージェントによる対話に、前記複数のＡＩエージェントのうち、少なくとも１つのＡＩエージェントに対応する仮想を含む人物の特徴を示す情報を加える、

プログラム。