(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)公開番号
特開2025-68754
(P2025-68754A)
(43)公開日　令和7年4月30日(2025.4.30)

(51)国際特許分類　FI
G06F 40/56 (2020.01)　G06F 40/56
G06F 3/16 (2006.01)　G06F 3/16 690
G06F 3/16 650

審査請求　未請求　請求項の数　38　OL　（全82頁）

(21)出願番号　特願2023-178738(P2023-178738)
(22)出願日　令和5年10月17日(2023.10.17)

（特許庁注：以下のものは登録商標）
１．ＰＹＴＨＯＮ

(71)出願人　000005810
マクセル株式会社
京都府乙訓郡大山崎町大山崎小泉１番地
(74)代理人　110002066
弁理士法人筒井国際特許事務所
(72)発明者　清水　拓也
京都府乙訓郡大山崎町大山崎小泉１番地
マクセル株式会社内
(72)発明者　鋪田　和夫
京都府乙訓郡大山崎町大山崎小泉１番地
マクセル株式会社内

(54)【発明の名称】　応答出力システムおよび応答出力装置

(57)【要約】　（修正有）
【課題】より好適な人工知能応答が可能な応答出力システムおよび応答出力装置を提供する。
【解決手段】応答出力システムにおいて、人工知能応答出力装置１００１０は、表示部１００１１と、音声出力部１１４０と、ユーザ入力にもとづいて指示文を生成し、大規模言語モデルへの指示文を送信する制御部と、を備える。制御部は、指示文に対する応答を生成して、インターネット１９０００を介して大規模言語モデルを保持する大規模言語モデルサーバ１９００１から大規模言語モデルを受信する。制御部による制御状態には、大規模言語モデルが生成した応答にもとづいて生成した出力を、表示部または音声出力部を介して出力する制御を行う状態がある。
【選択図】図１Ａ

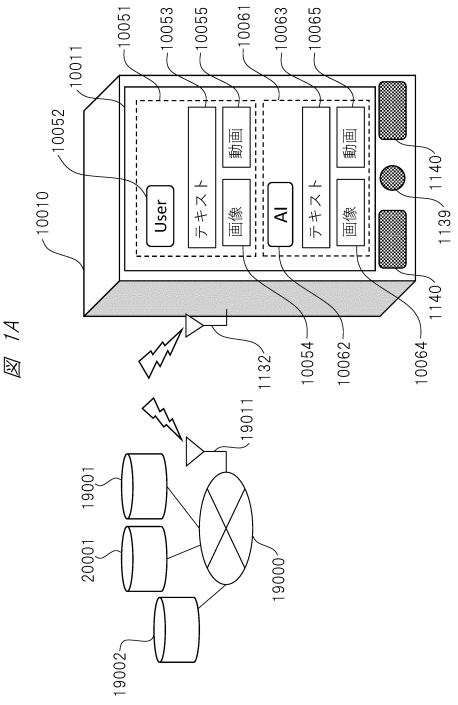


図 1A

【特許請求の範囲】
【請求項１】
表示部と、
音声出力部と、
ユーザ入力にもとづいて指示文を生成し、大規模言語モデルへの前記指示文を送信する制御部と、
前記指示文に対する応答を生成して前記制御部に送信する前記大規模言語モデルと、
を備え、
前記制御部による制御状態には、前記大規模言語モデルが生成した応答にもとづいて生成した出力を、前記表示部または前記音声出力部を介して出力する制御を行う状態がある、
応答出力システム。
【請求項２】
請求項１に記載の応答出力システムであって、
前記指示文または前記大規模言語モデルが生成した応答に不適切な内容が含まれる場合に前記大規模言語モデルが生成した応答にフィルタ処理を行い、前記フィルタ処理がなされた応答と前記フィルタ処理が行われたこと示すフラグとを前記大規模言語モデルから前記制御部へ送信する状態がある、
応答出力システム。
【請求項３】
請求項２に記載の応答出力システムであって、
前記フラグを受信した前記制御部は、前記フラグに応じた処理を実行し、該処理の結果に基づいたユーザへの出力を前記表示部または前記音声出力部を介して行う制御を行う、
応答出力システム。
【請求項４】
請求項３に記載の応答出力システムであって、
前記大規模言語モデルはインターネットに接続されたサーバに格納され、
前記制御部は前記サーバとは異なる装置に格納されるものであり、
前記大規模言語モデルと前記制御部は、前記インターネットを介して情報を送受信する、
応答出力システム。
【請求項５】
請求項３に記載の応答出力システムであって、
前記制御部は、前記ユーザ入力に対してキャラクターが応答する出力を、前記表示部または前記音声出力部を介して行う制御を行うものであり、
前記キャラクターが応答する出力は、前記制御部が受信した前記フラグに応じて生成されるものである、
応答出力システム。
【請求項６】
請求項１に記載の応答出力システムであって、
前記制御部は、ユーザ入力にもとづいて生成した指示文に変換処理を行って第２の指示文を生成し、前記第２の指示文を前記大規模言語モデルへ送信する、
応答出力システム。
【請求項７】
請求項６に記載の応答出力システムであって、
前記制御部が行う前記変換処理は、前記ユーザ入力にもとづいて生成した指示文の自然言語の情報に補助となる自然言語の情報を付加する処理である、
応答出力システム。
【請求項８】
請求項６に記載の応答出力システムであって、

前記制御部が行う前記変換処理は、前記ユーザ入力にもとづいて生成した指示文の非英語の自然言語の情報に補助となる英語の自然言語の情報を付加する処理である、
応答出力システム。
【請求項９】
請求項６に記載の応答出力システムであって、
前記制御部が行う前記変換処理は、前記ユーザ入力にもとづいて生成した指示文の非英語の自然言語の情報を英語の自然言語の情報に変換する処理である、
応答出力システム。
【請求項１０】
請求項６に記載の応答出力システムであって、
前記制御部が行う前記変換処理は、指示文のフォーマットにおいて、前記ユーザ入力にもとづいて生成した指示文の自然言語の情報が格納される第１のデータ領域とは異なる第２のデータ領域に自然言語またはコードである指示文を追加する処理である、
応答出力システム。
【請求項１１】
請求項１０に記載の応答出力システムであって、
前記第２のデータ領域は、実行される処理について予め取り決めがなされているフォーマットでデータを送受信するためのデータ領域である、
応答出力システム。
【請求項１２】
請求項１０に記載の応答出力システムであって、
前記第１のデータ領域に非英語による指示文が格納され、前記第２のデータ領域に英語による指示文が格納される、
応答出力システム。
【請求項１３】
表示部と、
音声出力部と、
制御部と、
第１の大規模言語モデルと、
第２の大規模言語モデルと、を備え、
前記制御部による制御状態には、ユーザ入力にもとづいて第１の指示文を生成し、前記第１の指示文を前記第１の大模言語モデルへ送信し、前記第１の指示文にもとづいて前記第１の大規模言語モデルが生成した第１の応答を取得し、前記第１の応答にもとづいて第２の指示文を生成し、前記第２の指示文を前記第２の大規模言語モデルへ送信し、前記第２の指示文にもとづいて前記第２の大規模言語モデルが生成した第２の応答を取得し、前記第１の応答および前記第２の応答の両者にもとづく処理を行い、前記第１の応答および前記第２の応答の両者にもとづく処理の結果に基づいて生成した出力を、前記表示部または前記音声出力部を介して出力する制御を行う状態がある、
応答出力システム。
【請求項１４】
請求項１３に記載の応答出力システムであって、
ストレージ部を備え、
前記第２の指示文の生成処理は、前記第１の応答の内容と前記ストレージ部に格納された定型文を組合せて前記第２の指示文を生成する処理である、
応答出力システム。
【請求項１５】
請求項１３に記載の応答出力システムであって、
前記第１の応答および前記第２の応答の両者にもとづく処理では、前記第２の応答に応じてユーザへの出力を切り替える処理を行う、
応答出力システム。

【請求項１６】
請求項１５に記載の応答出力システムであって、
ストレージ部を備え、
前記第１の応答および前記第２の応答の両者にもとづく処理におけるユーザへの出力の生成処理では、前記第１の応答の内容も、前記第２の応答の内容も前記ユーザへの出力へ含めず、前記ストレージ部に格納された定型文で前記ユーザへの出力を生成する、
応答出力システム。
【請求項１７】
請求項１５に記載の応答出力システムであって、
ストレージ部を備え、
前記第１の応答および前記第２の応答の両者にもとづく処理におけるユーザへの出力の生成処理では、前記第１の応答の内容と前記ストレージ部に格納された定型文を組合せて前記ユーザへの出力を生成する、
応答出力システム。
【請求項１８】
請求項１５に記載の応答出力システムであって、
ストレージ部を備え、
前記第１の応答および前記第２の応答の両者にもとづく処理におけるユーザへの出力の生成処理では、前記第２の応答の内容と前記ストレージ部に格納された定型文を組合せて前記ユーザへの出力を生成する、
応答出力システム。
【請求項１９】
請求項１３に記載の応答出力システムであって、
前記制御部は、前記第１の大模言語モデルによる前記第１の指示文に対する応答のトークン数を制限する第１の制御情報を、前記第１の大模言語モデルへ送信し、前記第２の大模言語モデルによる前記第２の指示文に対する応答のトークン数を制限する第２の制御情報を、前記第２の大模言語モデルへ送信するものであり、
前記第２の制御情報のパラメータに示されるトークン数は、前記第１の制御情報のパラメータに示されるトークン数よりも小さい、
応答出力システム。
【請求項２０】
表示部と、
音声出力部と、
ユーザ入力にもとづいて指示文を生成し、大規模言語モデルへの前記指示文を送信する制御部と、
を備え、
前記制御部による制御状態には、前記指示文にもとづいて前記大規模言語モデルが生成した応答を取得し、前記応答にもとづいて生成した出力を、前記表示部または前記音声出力部を介して出力する制御を行う状態がある、
応答出力装置。
【請求項２１】
請求項２０に記載の応答出力装置であって、
前記制御部は、前記指示文または前記大規模言語モデルが生成した応答に不適切な内容が含まれる場合に、前記フィルタ処理がなされた応答と前記フィルタ処理が行われたこと示すフラグとを前記大規模言語モデルから取得する、
応答出力装置。
【請求項２２】
請求項２１に記載の応答出力装置であって、
前記フラグを受信した前記制御部は、前記フラグに応じた処理を実行し、該処理の結果に基づいたユーザへの出力を前記表示部または前記音声出力部を介して行う制御を行う、

応答出力装置。
【請求項２３】
請求項２２に記載の応答出力装置であって、
前記応答出力装置は、前記大規模言語モデルが格納されたサーバとインターネットを介して接続され、
前記応答出力装置の前記制御部は、前記インターネットを介して前記大規模言語モデルと情報を送受信する、
応答出力装置。
【請求項２４】
請求項２２に記載の応答出力装置であって、
前記制御部は、前記ユーザ入力に対してキャラクターが応答する出力を、前記表示部または前記音声出力部を介して行う制御を行うものであり、
前記キャラクターが応答する出力は、前記制御部が受信した前記フラグに応じて生成されるものである、
応答出力装置。
【請求項２５】
請求項２０に記載の応答出力装置であって、
前記制御部は、ユーザ入力にもとづいて生成した指示文に変換処理を行って第２の指示文を生成し、前記第２の指示文を前記大規模言語モデルへ送信する、
応答出力装置。
【請求項２６】
請求項２５に記載の応答出力装置であって、
前記制御部が行う前記変換処理は、前記ユーザ入力にもとづいて生成した指示文の自然言語の情報に補助となる自然言語の情報を付加する処理である、
応答出力装置。
【請求項２７】
請求項２５に記載の応答出力装置であって、
前記制御部が行う前記変換処理は、前記ユーザ入力にもとづいて生成した指示文の非英語の自然言語の情報に補助となる英語の自然言語の情報を付加する処理である、
応答出力装置。
【請求項２８】
請求項２５に記載の応答出力装置であって、
前記制御部が行う前記変換処理は、前記ユーザ入力にもとづいて生成した指示文の非英語の自然言語の情報を英語の自然言語の情報に変換する処理である、
応答出力装置。
【請求項２９】
請求項２５に記載の応答出力装置であって、
前記制御部が行う前記変換処理は、指示文のフォーマットにおいて、前記ユーザ入力にもとづいて生成した指示文の自然言語の情報が格納される第１のデータ領域とは異なる第２のデータ領域に自然言語またはコードである指示文を追加する処理である、
応答出力装置。
【請求項３０】
請求項２９に記載の応答出力装置であって、
前記第２のデータ領域は、実行される処理について予め取り決めがなされているフォーマットでデータを送受信するためのデータ領域である、
応答出力装置。
【請求項３１】
請求項２９に記載の応答出力装置であって、
前記第１のデータ領域に非英語による指示文が格納され、前記第２のデータ領域に英語による指示文が格納される、

応答出力装置。
【請求項３２】
第１の大規模言語モデルとデータを送受信可能であり、第２の大規模言語モデルとデータを送受信可能である応答出力装置であって、
表示部と、
音声出力部と、
制御部と、
を備え、
前記制御部による制御状態には、ユーザ入力にもとづいて第１の指示文を生成し、前記第１の指示文を前記第１の大模言語モデルへ送信し、前記第１の指示文にもとづいて前記第１の大規模言語モデルが生成した第１の応答を取得し、前記第１の応答にもとづいて第２の指示文を生成し、前記第２の指示文を前記第２の大規模言語モデルへ送信し、前記第２の指示文にもとづいて前記第２の大規模言語モデルが生成した第２の応答を取得し、前記第１の応答および前記第２の応答の両者にもとづく処理を行い、前記第１の応答および前記第２の応答の両者にもとづく処理の結果に基づいて生成した出力を、前記表示部または前記音声出力部を介して行う制御状態がある、
応答出力装置。
【請求項３３】
請求項３２に記載の応答出力装置であって、
ストレージ部を備え、
前記第２の指示文の生成処理は、前記第１の応答の内容と前記ストレージ部に格納された定型文を組合せて前記第２の指示文を生成する処理である、
応答出力装置。
【請求項３４】
請求項３２に記載の応答出力装置であって、
前記第１の応答および前記第２の応答の両者にもとづく処理では、前記第２の応答に応じてユーザへの出力を切り替える処理を行う、
応答出力装置。
【請求項３５】
請求項３２に記載の応答出力装置であって、
ストレージ部を備え、
前記第１の応答および前記第２の応答の両者にもとづく処理におけるユーザへの出力の生成処理では、前記第１の応答の内容も、前記第２の応答の内容も前記ユーザへの出力へ含めず、前記ストレージ部に格納された定型文で前記ユーザへの出力を生成する、
応答出力装置。
【請求項３６】
請求項３２に記載の応答出力装置であって、
ストレージ部を備え、
前記第１の応答および前記第２の応答の両者にもとづく処理におけるユーザへの出力の生成処理では、前記第１の応答の内容と前記ストレージ部に格納された定型文を組合せて前記ユーザへの出力を生成する、
応答出力装置。
【請求項３７】
請求項３２に記載の応答出力装置であって、
ストレージ部を備え、
前記第１の応答および前記第２の応答の両者にもとづく処理におけるユーザへの出力の生成処理では、前記第２の応答の内容と前記ストレージ部に格納された定型文を組合せて前記ユーザへの出力を生成する、
応答出力装置。
【請求項３８】

請求項３２に記載の応答出力装置であって、
前記制御部は、前記第１の大模言語モデルによる前記第１の指示文に対する応答のトークン数を制限する第１の制御情報を、前記第１の大模言語モデルへ送信し、前記第２の大模言語モデルによる前記第２の指示文に対する応答のトークン数を制限する第２の制御情報を、前記第２の大模言語モデルへ送信するものであり、
前記第２の制御情報のパラメータに示されるトークン数は、前記第１の制御情報のパラメータに示されるトークン数よりも小さい、
応答出力装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、応答出力システムおよび応答出力装置に関する。
【背景技術】
【０００２】
言語モデルなどの人工知能を用いた応答出力技術については、例えば、特許文献１に開示されている。
【先行技術文献】
【特許文献】
【０００３】
【特許文献１】特表２０１９－５２８５１２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の開示では、人工知能を用いた応答出力技術をユーザにより好適に提供するための構成などについての考慮は十分ではなかった。
【０００５】
本発明の目的は、より好適な応答出力技術を提供することにある。
【課題を解決するための手段】
【０００６】
上記課題を解決するために、例えば特許請求の範囲に記載の構成を採用する。本願は上記課題を解決する手段を複数含んでいるが、その一例を挙げるならば、応答出力システムであって、表示部と、音声出力部と、ユーザ入力にもとづいて指示文を生成し、大規模言語モデルへの前記指示文を送信する制御部と、前記指示文に対する応答を生成して前記制御部に送信する前記大規模言語モデルと、を備え、前記制御部による制御状態には、前記大規模言語モデルが生成した応答にもとづいて生成した出力を、前記表示部または前記音声出力部を介して出力する制御を行う状態がある、ように構成すればよい。
【発明の効果】
【０００７】
本発明によれば、より好適な応答出力技術を提供できる。これ以外の課題、構成および効果は、以下の実施形態の説明において明らかにされる。
【図面の簡単な説明】
【０００８】
【図１Ａ】本発明の一実施例に係る人工知能応答出力装置およびシステムの一例を示す図である。
【図１Ｂ】本発明の一実施例に係る人工知能応答出力装置の一例を示す図である。
【図１Ｃ】本発明の一実施例に係る人工知能応答出力装置およびシステムの動作の一例を示す図である。
【図２Ａ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの一例の説明図である。
【図２Ｂ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システ

ムの動作の一例の説明図である。
【図２Ｃ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｄ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムにける会話の一例の説明図である。
【図２Ｅ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｆ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｇ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｈ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｉ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｊ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｋ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図２Ｌ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図３Ａ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの一例の説明図である。
【図３Ｂ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図３Ｃ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図３Ｄ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムにける会話の一例の説明図である。
【図３Ｅ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図３Ｆ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図３Ｇ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図３Ｈ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図３Ｉ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図４Ａ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図４Ｂ】本発明の一実施例に係るキャラクター会話装置およびキャラクター会話システムの動作の一例の説明図である。
【図５Ａ】本発明の一実施例に係る人工知能応答出力装置の動作の一例の説明図である。
【図５Ｂ】本発明の一実施例に係る人工知能応答出力装置の表示例の一例の説明図である。
【図５Ｃ】本発明の一実施例に係る人工知能応答出力装置の表示例の一例の説明図である。
【図５Ｄ】本発明の一実施例に係る人工知能応答出力装置の表示例の一例の説明図である。

【図６】本発明の一実施例に係る人工知能応答出力装置の応答生成処理の一例の説明図である。
【図７Ａ】本発明の一実施例に係る人工知能応答出力装置の応答出力処理の一例の説明図である。
【図７Ｂ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図７Ｃ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図７Ｄ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図７Ｅ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図８Ａ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図８Ｂ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図８Ｃ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図９Ａ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【図９Ｂ】本発明の一実施例に係る人工知能応答出力装置の処理の一例の説明図である。
【発明を実施するための形態】
【０００９】
以下、本発明の実施の形態を図面に基づいて詳細に説明する。なお、本発明は実施例の説明に限定されるものではなく、本明細書に開示される技術的思想の範囲内において当業者による様々な変更および修正が可能である。また、本発明を説明するための全図において、同一の機能を有するものには、同一の符号を付与し、その繰り返しの説明は省略する場合がある。
【００１０】
なお、本発明の各実施例に係る人工知能応答出力装置が表示画面を有する場合は、表示装置と呼んでもよい。人工知能応答出力装置が音声出力機能を有する場合は、音声出力装置と呼んでもよい人工知能応答出力装置は単に情報処理装置と呼んでもよい。人工知能応答出力装置と、大規模言語モデルを保持する大規模言語モデルサーバを含むシステムを人工知能応答出力システムと呼んでもよい。また、人工知能応答出力装置が人工知能である大規模言語モデルの応答サービスをユーザに提供し、ユーザの助力になる場合は、人工知能応答出力装置または人工知能応答出力装置の表示出力は、ユーザにとって人工知能（ＡＩ）アシスタントとなることができる。よって、この場合、人工知能応答出力装置はＡＩアシスタント装置またはＡＩアシスタント表示装置と呼んでもよい。同様に、この場合、人工知能応答出力装置と、大規模言語モデルを保持する大規模言語モデルサーバを含むシステムをＡＩアシスタントシステムまたはＡＩアシスタント表示システムと呼んでもよい。また、この場合、人工知能応答出力装置は、ユーザと人工知能の間のインタフェースとなるので、人工知能インタフェース装置と呼んでもよい。この場合、人工知能応答出力装置と、大規模言語モデルを保持する大規模言語モデルサーバを含むシステムを人工知能インタフェースシステムと呼んでもよい。
【００１１】
＜実施例１＞
本発明の実施例１として、大規模言語モデル人工知能からの応答を出力する人工知能応答出力装置およびそのシステムについて、説明する。
【００１２】
図１Ａを用いて、本発明の人工知能応答出力装置１００１０の一例について説明する。また、当該人工知能応答出力装置１００１０が大規模言語モデルサーバ１９００１と通信などにより連携する場合について、人工知能応答出力装置１００１０が大規模言語モデルサーバ１９００１および／またはマルチモーダルな大規模言語モデルサーバ２０００１とを含むシステムの一例について説明する。
【００１３】
図１Ａの例では、人工知能応答出力装置１００１０は、表示部１００１１を有する。図１Ａの例では、表示部１００１１は、平面ディスプレイでもよく、背面から映像を投影するスクリーンでもよく、光学像を空中に結像する空中浮遊映像でもよい。表示部１００１

１はが平面ディスプレイの場合は、液晶パネルとバックライトを有する液晶ディスプレイでもよい。また、表示部１００１１は、プラズマディスプレイでもよい。表示部１００１１は、画素が自発光する有機ＥＬディスプレイでもよい。また、表示部１００１１はタッチ操作入力センサを設け、タッチパネルとして構成してもよい。

【００１４】

図１Ａの例では、人工知能応答出力装置１００１０が備える音声出力部１１４０はスピーカで構成されている。また、人工知能応答出力装置１００１０がマイク１１３９を備え、ユーザの声を収音できる。当該マイク１１３９からの音声入力や、後述する操作入力部を介したユーザの操作入力により、人工知能応答出力装置１００１０は、人工知能である大規模言語モデルへの指示文（プロンプト）の元となるユーザ入力を取得することができる。

【００１５】

人工知能応答出力装置１００１０は、人工知能応答出力装置１００１０自体にローカルの大規模言語モデルを備えてもよい。この場合、当該大規模言語モデルの応答を上記表示部１００１１の表示出力、および／または、音声出力部１１４０の音声出力として出力してもよい。

【００１６】

また、人工知能応答出力装置１００１０は、ローカルの大規模言語モデルを備えず、外部の大規模言語モデルサーバ１９０００１と通信し、大規模言語モデルサーバ１９０００１から受信する応答を上記表示部１００１１の表示出力、および／または、音声出力部１１４０の音声出力として出力してもよい。

【００１７】

または、人工知能応答出力装置１００１０は、ローカルの大規模言語モデルも備え、さらに、大規模言語モデルを有する外部の大規模言語モデルサーバ１９０００１またはマルチモーダルな大規模言語モデルを有する外部の大規模言語モデルサーバ２００００１と通信するように構成してもよい。この場合、当該ローカルの大規模言語モデルの応答と、大規模言語モデルサーバ１９０００１の大規模言語モデルまたはマルチモーダルな大規模言語モデルサーバ２００００１のマルチモーダルな大規模言語モデルから受信する応答とを、切り替えて、いずれか一方を、上記表示部１００１１の表示出力、および／または、音声出力部１１４０の音声出力として出力してもよい。または、当該ローカルの大規模言語モデルの応答と、大規模言語モデルサーバ１９０００１の大規模言語モデルまたはマルチモーダルな大規模言語モデルサーバ２００００１のマルチモーダルな大規模言語モデルから受信する応答の両者にもとづいて生成した応答を、上記表示部１００１１の表示出力、および／または、音声出力部１１４０の音声出力として出力してもよい。

【００１８】

人工知能応答出力装置１００１０が外部の大規模言語モデルサーバ１９０００１または大規模言語モデルサーバ２００００１と通信して連携する場合の構成は、以下のとおりである。人工知能応答出力装置１００１０は、通信部１１３２を介して、インターネット１９０００に接続された通信装置１９０１１と通信可能である。図１Ａの例では、通信部１１３２と通信装置１９０１１との通信は無線の例を示しているが、有線通信でも構わない。通信部１１３２と通信装置１９０１１までの通信経路において、有線の部分と無線の部分があってもよいし、ルータや中継器を経由してもよい。また、通信部１１３２からインターネット１９０００までの通信経路において、有線の部分と無線の部分があってもよいし、ルータや中継器を経由してもよい。人工知能応答出力装置１００１０は、通信装置１９０１１およびインターネット１９０００を介して、大規模言語モデルサーバ１９０００１と通信可能である。また、人工知能応答出力装置１００１０は、通信装置１９０１１およびインターネット１９０００を介して、大規模言語モデルサーバ１９０００１または大規模言語モデルサーバ２００００１、およびこれらのサーバと異なる第２のサーバ１９０００２と通信可能である。人工知能応答出力装置１００１０と大規模言語モデルサーバ１９０００１または大規模言語モデルサーバ２００００１とを含めた構成を一つのシステムとして考えてもよ

い。
【００１９】
以下の説明において、単に「大規模言語モデル」と説明し特に断りがない場合は、人工知能応答出力装置１００１０が備えるローカルの大規模言語モデルと、大規模言語モデルサーバ１９００１が備える大規模言語モデル、および大規模言語モデルサーバ２０００１が備えるマルチモーダルな大規模言語モデルを含めた概念と考えてよい。
【００２０】
図１Ａの例では、表示部１００１１が、ユーザから人工知能である大規模言語モデルへの指示文（プロンプト）を入力する指示文表示領域１００５１と、大規模言語モデルからの応答を表示する人工知能応答表示領域１００６１の２つの表示領域に各要素を表示する例を示している。図１Ａの例では、指示文表示領域１００５１には、ユーザを示すアイコン１００５２、指示文の構成要素としての自然言語やソフトウェアコードなどのテキスト１００５３、指示文の構成要素としての画像１００５４、指示文の構成要素としての動画１００５５、などを表示する例を示している。図１Ａの例では、人工知能応答表示領域１００６１には、人工知能または人工知能アシスタントを示すアイコン１００６２、人工知能からの応答の構成要素としての自然言語やソフトウェアコードなどのテキスト１００６３、人工知能からの応答の構成要素としての画像１００６４、人工知能からの応答の構成要素としての動画１００６５、などを表示する例を示している。なお、図１Ａに示す人工知能応答出力装置１００１０の表示部１００１１の表示例は、あくまで一例である。人工知能応答出力装置１００１０が用いられる実装例に応じて、図１Ａに示す例とは異なる表示を行えばよい。
【００２１】
ここで、大規模言語モデルについて説明する。大規模言語モデルはＬＬＭ（Ｌａｒｇｅ Ｌａｎｇｕａｇｅ Ｍｏｄｅｌ）とも表記される。具体的には、ＧＰＴ－１、ＧＰＴ－２、ＧＰＴ－３、ＩｎｓｔｒｕｃｔＧＰＴ、ＣｈａｔＧＰＴなど様々なモデルが公開されている。本実施例においてもこれらの技術を用いればよい。なお、これらの大規模言語モデルは人間界に存在する数多くの文書、テキストに含まれる自然言語を対象に、大規模な事前学習が行われて生成された人工知能モデルである。人工知能モデルのパラメータ数は億を超える。さらに、これに加えて、人間からのフィードバックにもとづく強化学習を施したモデルもある。ベースとなるモデルの一例はＴｒａｎｓｆｏｒｍｅｒと呼ばれるモデルなどである。これらのモデルの学習の一例として、例えば、参考文献１などが公開されている。
【００２２】
［参考文献１］
Long Ouyang, et. al. “Training language models to follow instructions with human feedback”, https://arxiv.org/pdf/2203.02155.pdf
【００２３】
これらの大規模言語モデルは、自然言語を対象とした翻訳、自然言語を対象とした文章校正、自然言語を対象とした文章要約、などが可能である。そのうち、高度なものでは、自然言語による質問回答（対話または会話ともよばれる）、自然言語による提案生成、プログラミングコードの生成などが可能である。これらの人工知能モデルのパラメータ数は非常に大きいため、学習には膨大なデータ、計算資源が必要である。よって、特定の用途に限って、このレベルの人工知能の学習を行うことは非常に資源効率が悪い。そこで、様々な用途に応用できる基盤モデル（Ｆｏｕｎｄａｔｉｏｎ Ｍｏｄｅｌ）として、大規模な事前学習を行ってモデルが生成されている。例えば、図１Ａに示す大規模言語モデルサーバ１９００１は、このような大規模言語モデルを備え、ＡＰＩ（Ａｐｐｌｉｃａｔｉｏｎ Ｐｒｏｇｒａｍｍｉｎｇ Ｉｎｔｅｒｆａｃｅ）を介して、様々な端末で利用できるように構成してもよい。また、図１Ａに示す人工知能応答出力装置１００１０がローカルの大規模言語モデルを備え、人工知能応答出力装置１００１０自身が利用するように構成してもよい。いずれの大規模言語モデルの学習自体は、別途、大規模な事前学習を行って

生成し、生成した大規模言語モデルを複製して、大規模言語モデルサーバ１９００１や人工知能応答出力装置１００１０などに備えればよい。このように、用途ごとや端末ごとに事前学習を行うのではなく、大規模な事前学習を行って生成した基盤モデルである大規模言語モデルを複製して個々のサーバや端末で利用すれば、学習に用いる資源消費を共有することができるため資源効率がよい。

【００２４】

なお、大規模な事前学習を行って生成した基盤モデルとしての大規模言語モデルであっても、個々のサーバや装置において、用途や目的に応じて転移学習などの追加学習を行うように構成してもよい。

【００２５】

また、大規模言語モデルは、自然言語を事前学習し、自然言語を対象とした入出力処理を行うことができる。さらに、自然言語のテキスト情報に加えて自然言語のテキスト情報以外の種類の情報も併せて処理が可能なマルチモーダルな大規模言語モデル人工知能も本発明の実施例に適用可能である。図１Ａにおいては、マルチモーダルな大規模言語モデルを有するサーバを大規模言語モデルサーバ２０００１を示す。例えば、マルチモーダルな大規模言語モデル人工知能の一例としては、具体的には、ＧＰＴ－４（参考文献２参照）、Ｇａｔｏ（参考文献３参照）などが公開されている。本実施例においてもこれらの技術を用いればよい。なお、これらのマルチモーダルな大規模言語モデルは人間界に存在する数多くの文書、テキストに含まれる自然言語および自然言語のテキスト情報以外の種類の情報（例えば、画像、動画、音声など）を対象に、大規模な事前学習が行われて生成された人工知能モデルである。さらに、これに加えて、人間からのフィードバックにもとづく強化学習を施したモデルもある。以下、画像、動画、音声などの自然言語のテキスト情報以外の種類の情報を非自然言語情報源と称してもよい。

【００２６】

［参考文献２］

Open AI “GPT-4 Technical Report”, https://cdn.openai.com/papers/gpt-4.pdf

［参考文献３］

Scott Reed, et. al. “A Generalist Agent”, https://arxiv.org/pdf/2205.06175.pdf

【００２７】

次に、図１Ｂを用いて、これらの大規模言語モデルなどの人工知能に対するユーザからの入力を受け付け、当該ユーザからの入力に対する大規模言語モデルなどの人工知能からの応答を出力する、人工知能応答出力装置１００１０の構成例について説明する。

【００２８】

人工知能応答出力装置１００１０は、表示部１００１１、制御部１１１０、メモリ１１０９、不揮発性メモリ１１０８、外部電源入力インタフェース１１１１、操作入力部１１０７、電源１１０６、二次電池１１１２、ストレージ部１１７０、映像制御部１１６０、姿勢センサ１１１３、通信部１１３２、音声出力部１１４０、マイク１１３９、映像信号入力部１１３１、音声信号入力部１１３３、撮像部１１８０、等を備えている。人工知能応答出力装置１００１０は、例えば、いわゆるモニタやテレビなどの大画面を有するものでもよい。

【００２９】

表示部１００１１は、平面ディスプレイでもよく、背面から映像を投影するスクリーンでもよく、光学像を空中に結像する空中浮遊映像を表示するものでもよい。表示部１００１１はが平面ディスプレイの場合は、液晶パネルとバックライトを有する液晶ディスプレイでもよい。また、表示部１００１１は、プラズマディスプレイでもよい。表示部１００１１は、画素が自発光する有機ＥＬディスプレイでもよい。表示部１００１１がパネルの場合は、表示パネルと称してもよい。表示部１００１１にタッチ操作入力センサを設け、ユーザ２３０の指によるタッチ操作入力を受け付けるように構成してもよい。この場合、

表示部１００１１はタッチパネルとして構成してもよい。当該タッチパネルを介したユーザの操作入力により、人工知能応答出力装置１００１０は、人工知能である大規模言語モデルへの指示文（プロンプト）の元となるユーザ入力を取得することができる。

【００３０】

通信部１１３２は、Ｗｉ－Ｆｉ方式の通信インタフェース、Ｂｌｕｅｔｏｏｔｈ（登録商標）方式の通信インタフェース、４Ｇ、５Ｇなどの移動体通信インタフェースなどで構成すればよい。これらの通信方式を用いて、人工知能応答出力装置１００１０の通信部１１３２は、インターネット１９０００に接続された通信装置１９０１１と通信可能である。なお、通信部１１３２と通信装置１９０１１までの通信経路において、有線の部分と無線の部分があってもよいし、ルータや中継器を経由してもよい。有線の場合は、通信部１１３２は、ハードウェアとしてイーサネットの接続インタフェースを有してＬＡＮ方式の通信方式を用いて通信を行ってもよい。これにより、人工知能応答出力装置１００１０はインターネット１９０００に接続された各種サーバと通信可能である。

【００３１】

人工知能応答出力装置１００１０にはＣＰＵなどの制御部１１１０およびメモリ１１０９が備えられており、当該制御部１１１０は、表示部１００１１や通信部１１３２などを制御する。

【００３２】

電源１１０６は、外部から外部電源入力インタフェース１１１１を介して入力されるＡＣ電流をＤＣ電流に変換し、人工知能応答出力装置１００１０の各部にそれぞれ必要なＤＣ電流を供給する。二次電池１１１２は、電源１１０６から供給される電力を蓄電する。また、二次電池１１１２は、外部電源入力インタフェース１１１１を介して、外部から電力が供給されない場合に、電力を必要とする各部に対して電力を供給する。

【００３３】

操作入力部１１０７は、例えば操作ボタンや、リモートコントローラ等の信号受信部または赤外光受光部であり、表示部１００１１のタッチ操作入力センサへのユーザによるタッチ操作とは異なる操作についての信号を入力する。表示部１００１１のタッチ操作入力センサをタッチ操作するユーザとは別に、操作入力部１１０７は、例えば管理者が人工知能応答出力装置１００１０を操作するために用いられてもよい。当該操作入力部１１０７を介したユーザの操作入力により、人工知能応答出力装置１００１０は、人工知能である大規模言語モデルへの指示文（プロンプト）の元となるユーザ入力を取得することができる。なお、表示部１００１１のタッチ操作入力センサも前記操作入力部１１０７の一部として含む構成とする変形例もあり得る。

【００３４】

映像信号入力部１１３１は、外部の映像出力装置を接続して映像データを入力する。映像信号入力部１１３１は、様々なデジタル映像入力インタフェースが考えられる。例えば、ＨＤＭＩ（登録商標）（High－Definition Multimedia Interface）規格の映像入力インタフェース、ＤＶＩ（Digital Visual Interface）規格の映像入力インタフェース、またはＤｉｓｐｌａｙＰｏｒｔ規格の映像入力インタフェースなどで構成すればよい。または、アナログＲＧＢや、コンポジットビデオなどのアナログ映像入力インタフェースを設けてもよい。映像信号入力部１１３１は、各種ＵＳＢインタフェースなどでもよい。

【００３５】

音声信号入力部１１３３は、外部の音声出力装置を接続して音声データを入力する。音声信号入力部１１３３は、ＨＤＭＩ規格の音声入力インタフェース、光デジタル端子インタフェース、または、同軸デジタル端子インタフェース、などで構成すればよい。音声信号入力部１１３３は、各種ＵＳＢインタフェースなどでもよい。ＨＤＭＩ規格のインタフェースの場合は、映像信号入力部１１３１と音声信号入力部１１３３とは、端子およびケーブルが一体化したインタフェースとして構成されてもよい。

【００３６】

音声出力部１１４０は、音声信号入力部１１３３に入力された音声データに基づいた音声を出力することが可能である。音声出力部１１４０は、ストレージ部１１７０に格納されている音声データに基づいた音声を出力することも可能である。音声出力部１１４０は、スピーカで構成してもよい。また、音声出力部１１４０は、内蔵の操作音やエラー警告音を出力してもよい。または、ＨＤＭＩ規格に規定されるＡｕｄｉｏ Ｒｅｔｕｒｎ Ｃｈａｎｎｅｌ機能のように、外部機器にデジタル信号として音声信号を出力する構成を音声出力部１１４０としてもよい。または、ヘッドホンなどの外部機器にアナログ信号として音声信号を出力する構成を音声出力部１１４０としてもよい。

【００３７】

マイク１０３９は、人工知能応答出力装置１００１０の周辺の音を収音し、信号に変換して音声信号を生成するマイクである。ユーザの声など人物の声をマイクが収録して、生成した音声信号を後述する制御部１１１０が音声認識処理を行って、当該音声信号から文字情報を取得するように構成してもよい。当該マイク１１３９からの音声入力により、人工知能応答出力装置１００１０は、人工知能である大規模言語モデルへの指示文（プロンプト）の元となるユーザ入力を取得することができる。

【００３８】

撮像部１１８０は、イメージセンサを有するカメラである。人工知能応答出力装置１００１０の表示部１００１１側の前面にカメラを設けてもよく、表示部１００１１側の背面にカメラを設けてもよい。前面のカメラと背面のカメラの両者を設けてもよい。本実施例では、撮像部１１８０は、前面のカメラと背面のカメラの両者を有するものとして説明する。

【００３９】

ストレージ部１１７０は、映像データ、画像データ、音声データ等の各種データなどの各種情報を記録する記憶装置である。ストレージ部１１７０は、ハードディスクドライブ（ＨＤＤ）などの磁気記録媒体記録装置や、ソリッドステートドライブ（ＳＳＤ）などの半導体素子メモリで構成してもよい。ストレージ部１１７０には、例えば、製品出荷時に予め映像データ、画像データ、音声データ等の各種データ等の各種情報が記録されていてもよい。また、ストレージ部１１７０は、通信部１１３２を介して外部機器や外部のサーバ等から取得した映像データ、画像データ、音声データ等の各種データ等の各種情報を記録してもよい。ストレージ部１１７０に記録された映像データ、画像データ等は、表示部１００１１に出力される。ストレージ部１１７０に記録された映像データ、画像データ等を、通信部１１３２を介して外部機器や外部のサーバ等に出力してもよい。

【００４０】

映像制御部１１６０は、表示部１００１１に入力する映像信号に関する各種制御を行う。映像制御部１１６０は、映像処理回路と称してもよく、例えば、ＡＳＩＣ、ＦＰＧＡ、映像用プロセッサなどのハードウェアで構成されてもよい。なお、映像制御部１１６０は、映像処理部、画像処理部と称してもよい。映像制御部１１６０は、例えば、メモリ１１０９に記憶させる映像信号と、映像信号入力部１１３１に入力された映像信号（映像データ）等のうち、どの映像信号を表示部１００１１に入力するかといった映像切り替えの制御等を行う。また、映像制御部１１６０は、映像信号入力部１１３１から入力された映像信号やメモリ１１０９に記憶させる映像信号等に対して画像処理を行う制御を行ってもよい。画像処理としては、例えば、画像の拡大、縮小、変形等を行うスケーリング処理、輝度を変更するブライト調整処理、画像のコントラストカーブを変更するコントラスト調整処理、画像を光の成分に分解して成分ごとの重みづけを変更するレティネックス処理等がある。

【００４１】

姿勢センサ１１１３は、重力センサまたは加速度センサ、またはこれらの組み合わせにより構成されるセンサであり、人工知能応答出力装置１００１０の姿勢を検出することができる。姿勢センサ１１１３の姿勢検出結果に基づいて、制御部１１１０が、接続される各部の動作を制御してもよい。

【００４２】
不揮発性メモリ１１０８は、人工知能応答出力装置１００１０で用いる各種データを格納する。不揮発性メモリ１１０８に格納されるデータには、例えば、人工知能応答出力装置１００１０の表示部１００１１に表示する各種操作用のデータ、表示アイコン、ユーザの操作が操作するためのオブジェクトのデータやレイアウト情報等が含まれる。メモリ１１０９は、表示部１００１１に表示する映像データや装置の制御用データ等を記憶する。制御部１１１０がストレージ部１１７０から各種ソフトウェアを読み出して、メモリ１１０９に展開して記憶してもよい。
【００４３】
ローカルＬＬＭ処理部１００２８は、大規模言語モデル（ＬＬＭ）を保持できるメモリを備え、制御部１１１０の制御にもとづいて、大規模言語モデルの推論を実行できる。ハードウェアとしてはいわゆるＧＰＵ（Graphics Processing Unit）などで構成すればよい。ローカルＬＬＭ処理部１００２８は、推論のみならず、学習を行ってもよい。なお、人工知能応答出力装置１００１０のローカル環境での大規模言語モデルの推論の実行が不要な場合などは必ずしも、ローカルＬＬＭ処理部１００２８を要しない。
【００４４】
制御部１１１０は、接続される各部の動作を制御する。また、制御部１１１０は、メモリ１１０９に記憶されるプログラムと協働して、人工知能応答出力装置１００１０内の各部から取得した情報に基づく演算処理を行ってもよい。制御部１１１０による制御状態には、例えば、ローカルＬＬＭ処理部１００２８の大規模言語モデルからの応答や、通信部１１３２を介して取得した、大規模言語モデルサーバ１９００１の大規模言語モデルまたはマルチモーダルな大規模言語モデルサーバ２０００１のマルチモーダルな大規模言語モデル大規模言語モデルからの応答を、表示部１００１１またはスピーカ等である音声出力部１１４０を介して出力する状態がある。
【００４５】
なお、上述のタッチパネル、マイク１１３９または操作入力部１１０７を介してユーザから入力があった場合に、当該入力にもとづいて指示文を生成し、人工知能応答出力装置１００１０が備えるローカルＬＬＭ処理部１００２８のローカルの大規模言語モデル、大規模言語モデルサーバ１９００１が備える大規模言語モデル、または大規模言語モデルサーバ２０００１が備えるマルチモーダルな大規模言語モデルへ送信し、これらの大規模言語モデルから応答を取得する制御は、いずれも制御部１１１０が行えばよい。
【００４６】
また、ストレージ部１１７０には、人工知能応答出力装置１００１０の指示文の応答として定型文を出力するための応答定型文データベース（応答定型文ＤＢと表記してもよい）を格納してもよい。制御部１１１０が当該応答定型文データベースに格納されるデータを用いて出力する応答を生成する制御を行えばよい。図１Ｃに応答定型文データベースの一例を示す。図１Ｃの例では、条件番号が付された各条件について、人工知能応答出力装置１００１０が出力する定型文の応答が格納されている。例えば、条件番号１のように、上述のタッチパネル、マイク１１３９または操作入力部１１０７を介してユーザから「おはようございます」が入力された場合、応答定型文として「おはようございます」または「今日は〇月〇日ですね。」を用いて応答を出力すればよい。「〇月〇日」など〇の部分は、人工知能応答出力装置１００１０が有するメモリ１１０９などに格納される情報を用いて生成すればよい。
【００４７】
また、図１Ｃに示すデータベースにおける応答定型文の例において、／で区切られた複数の応答定型文が格納されている場合は、制御部１１１０は、乱数などを用いてランダムに、いずれかの応答定型文を選択して応答が出力されるように制御すればよい。このようにすると、同一の条件における応答が単調になるという状況を解消して改善することができる。条件番号２、３、および４の例についても、条件番号１の例の説明と同様である。図１Ｃに示される各例の条件内容に対して、図１Ｃに示される各例の応答定型文を用いた

出力をするように、制御部１１１０が制御を行えばよい。
【００４８】
次に、図１Ｃに示される条件番号５の例について説明する。条件番号５は、タッチパネル、マイク１１３９または操作入力部１１０７を介して取得したユーザ入力について、制御部１１１０が自然言語として意味が理解できなかった場合またはユーザ入力に明らかに文法の誤りがある場合に、制御部１１１０が応答定型文として「ちょっと聞き取れませんでした」または「それについてはわからないかもしれません」を用いて応答を出力する制御を行う例である。このように応答することでユーザに再度の入力を促すことができ、修正されたユーザ入力を待つことができる。
【００４９】
次に、図１Ｃに示される条件番号６の例について説明する。条件番号６は、図１Ｂに示す人工知能応答出力装置１００１０を構成する各部のいずれかの部においてエラー（異常状態）であることを制御部１１１０が検出している状態で、タッチパネル、マイク１１３９または操作入力部１１０７を介してユーザ入力があった場合の例である。この場合、制御部１１１０は応答定型文として「調子が悪いみたいです」を用いて応答を出力する制御を行う。このように応答することでユーザに人工知能応答出力装置１００１０が不調であることを説明することができ、ユーザにエラー対応などを促すことができる。
【００５０】
人工知能応答出力装置１００１０は、人工知能応答出力装置１００１０が備えるローカルの大規模言語モデル、大規模言語モデルサーバ１９００１が備える大規模言語モデル、および大規模言語モデルサーバ２０００１が備えるマルチモーダルな大規模言語モデル、などの大規模言語モデルの応答に替えて、図１Ｃを用いて説明した応答定型文データベース（応答定型文ＤＢ）を用いた応答を出力してもよい。または、これらの大規模言語モデルの応答と応答定型文データベース（応答定型文ＤＢ）を用いた応答を組み合わせた応答を出力してもよい。
【００５１】
なお、以上説明した図１Ｃの応答定型文データベース（応答定型文ＤＢ）はストレージ部１１７０に格納され、人工知能応答出力装置１００１０の制御部１１１０がこれを用いればよい。しかしながら、図１Ｃに示す応答定型文データベース（応答定型文ＤＢ）を大規模言語モデルサーバ１９００１側または大規模言語モデルサーバ２０００１側に備えてもよい。この場合は、大規模言語モデルサーバ１９００１が有する制御部または大規模言語モデルサーバ２０００１が有する制御部が、当該応答定型文データベース（応答定型文ＤＢ）を用いた応答を生成すればよい。大規模言語モデルサーバ１９００１が有する制御部または大規模言語モデルサーバ２０００１が有する制御部は、それぞれのサーバに格納される大規模言語モデルにより生成する応答に替えて、応答定型文データベース（応答定型文ＤＢ）を用いて生成した応答を、人工知能応答出力装置１００１０へ送信すればよい。このようにすれば、人工知能応答出力装置１００１０に応答定型文データベース（応答定型文ＤＢ）が備えられていない場合でも、応答定型文データベース（応答定型文ＤＢ）を用いた応答の生成が可能となる。
【００５２】
なお、上記の説明では、人工知能応答出力装置１００１０は固定画素を用いた表示画面の表示パネルを有すると説明した。当該概念には、固定画素を用いた表示画面の表示パネルのあとに投射光学系を設けて、当該表示画面の表示パネルの映像の光学像をスクリーンや壁に投射する、投射型映像表示装置（プロジェクタ）を含んでもよい。
【００５３】
なお、図１Ａおよび図１Ｂの例では、人工知能応答出力装置１００１０が表示部１００１１を備える例を説明した。しかしながら、本発明の実施例にかかる人工知能応答出力装置１００１０は必ずしも表示部１００１１を備えていなくてもよい。例えば、表示部１００１１を備えなくとも、人工知能に対するユーザからの入力を音声信号入力部１１３３またはマイク１１３９を介して受け付け、当該ユーザからの入力に対する大規模言語モデル

などの人工知能からの応答を音声出力部１１４０を介して出力するように構成すればよい。

【００５４】

以上説明した本発明の実施例１に係る、人工知能応答出力装置および人工知能応答出力システムによれば、大規模言語モデルなどの人工知能に対するユーザからの入力を受け付け、ネットワーク上のサーバ装置が有する大規模言語モデルまたは人工知能応答出力装置自体が有するローカルの大規模言語モデルなどの人工知能の推論により生成された、当該ユーザからの入力に対する応答を出力することが可能となる。

【００５５】

＜実施例２＞

次に、本発明の実施例２として、実施例１で説明した人工知能応答出力装置１００１０をインターネットに接続し、インターネットを介して、大規模言語モデル人工知能を搭載したサーバと接続して動作を行う例について説明する。本実施例では、実施例１との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。

【００５６】

図２Ａを用いて、本発明の実施例２の人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１の接続状態の一例について説明する。実施例２に係る人工知能応答出力装置１００１０はキャラクター会話装置、と呼んでもよい。また、実施例２に係る人工知能応答出力装置１００１０および大規模言語モデルサーバ１９００１を含むシステムはキャラクター会話システム、と呼んでもよい。人工知能応答出力装置１００１０が表示する表示部１００１１にはキャラクター１９０５１の映像が表示されている。キャラクター１９０５１の映像は、仮想空間上のキャラクターの３Ｄモデルをレンダリングして生成された映像である。

【００５７】

また、本実施例におけるキャラクターは、人工知能である大規模言語モデルのサービスをユーザに提供し、ユーザの助力になることができる。よって、当該キャラクターはユーザにとって人工知能（ＡＩ）アシスタントとなることができる。この場合、本実施例におけるキャラクター会話装置や、キャラクター会話システムは、ＡＩアシスタント会話装置、ＡＩアシスタント表示装置、ＡＩアシスタント応答出力装置や、ＡＩアシスタント会話システム、ＡＩアシスタント表示システム、ＡＩアシスタント応答出力システムと称してもよい。

【００５８】

図２Ａの例では、人工知能応答出力装置１００１０が備える音声出力部１１４０はスピーカで構成されている。また、人工知能応答出力装置１００１０がマイク１１３９を備え、ユーザの声を収音できる。人工知能応答出力装置１００１０は、通信部１１３２を介して、インターネット１９０００に接続された通信装置１９０１１と通信可能である。図２Ａの例では、通信部１１３２と通信装置１９０１１との通信は無線の例を示しているが、有線通信でも構わない。通信部１１３２からインターネット１９０００までの通信経路において、有線の部分と無線の部分があっても構わない。人工知能応答出力装置１００１０は、通信装置１９０１１およびインターネット１９０００を介して、大規模言語モデルサーバ１９００１と通信可能である。また、人工知能応答出力装置１００１０は、通信装置１９０１１およびインターネット１９０００を介して、大規模言語モデルサーバ１９００１と異なる第２のサーバ１９００２と通信可能である。人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含めた構成を一つのシステムとして考えてもよい。

【００５９】

次に、図２Ｂを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。なお、図２Ｂでは、図２Ａに示したインターネット１９０００等の

通信経路の図示は省略した。図２Ｂでは、人工知能応答出力装置１００１０のユーザ２３０も図示されている。

【００６０】

ここで、人工知能応答出力装置１００１０の動作の一連の流れを説明する。なお、人工知能応答出力装置１００１０は、ストレージ部１１７０などに格納しているキャラクター動作プログラムをメモリ１１０９に展開し、当該キャラクター動作プログラムを制御部１１１０が実行することにより、以下に説明する各種処理が実現可能となる。

【００６１】

まず、人工知能応答出力装置１００１０はマイク１１３９を備えており、ユーザ２３０がキャラクター１９０５１に話しかけると、そのユーザの声（ユーザからの言葉）をマイク１１３９で収音し、音声信号に変換する。ここで、制御部１１１０により実行されるキャラクター動作プログラムは、当該音声信号からユーザ２３０が話しかけた言葉のテキストを抽出する。当該テキストは自然言語である。なお、ユーザ２３０が話しかけた言葉のテキストの抽出はすべての言葉を対象に継続して行ってもよいが、トリガーとなるキーワードの後の所定期間にユーザから言葉が発せられた場合に開始してもよい。例えば、トリガーとなるキーワードとは、ユーザから「ハロー」に続けてキャラクター名が発せられた場合などでよい。例えば、キャラクター１９０５１の名前が「Ｋｏｔｏ」だとすると、「ハロー、Ｋｏｔｏ！」をトリガーとなるキーワードとすればよい。

【００６２】

人工知能応答出力装置１００１０のキャラクター動作プログラムは、ユーザ２３０が話しかけた言葉のテキストにもとづいて、指示文（プロンプト）を作成し、ＡＰＩを用いて当該指示文を大規模言語モデルサーバ１９００１へ送信する。ここで、指示文は、マークアップ言語のマークアップ形式などのタグを用いた記法や、マークダウン形式などの所定の記号を用いた記法、またはＪＳＯＮなどの所定のスクリプトのオブジェクト記法、などにより記載された情報が格納されたメタデータなどでよい。当該指示文には、主たるメッセージとして自然言語のテキスト情報が格納されている。人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される指示文の種類としては、初期設定などの指示を格納する設定指示文と、ユーザからの指示を反映するユーザ指示文とがある。指示文が設定指示文であるかユーザ指示文であるかを識別する種類識別情報を、当該指示文の主たるメッセージ以外の部分に格納しておいてもよい。人工知能応答出力装置１００１０のキャラクター動作プログラムが、ユーザ２３０が話しかけた言葉のテキストにもとづいて指示文（プロンプト）を作成するときには、ユーザ指示文を作成し、大規模言語モデルサーバ１９００１に送信する。

【００６３】

次に、大規模言語モデルサーバ１９００１の人工知能の大規模言語モデルは、人工知能応答出力装置１００１０から送信された指示文にもとづいて、推論を実行し、その結果に基づいて、自然言語のテキスト情報を含む応答を生成する。大規模言語モデルサーバ１９００１は、ＡＰＩを用いて、当該応答を人工知能応答出力装置１００１０へ送信する。当該応答には、主たるメッセージとして自然言語のテキスト情報が格納されている。ここで、当該応答は、前述の指示文と同じフォーマットの記法（マークアップ言語のマークアップ形式などのタグを用いた記法や、マークダウン形式などの所定の記号を用いた記法、またはＪＳＯＮなどの所定のスクリプトのオブジェクト記法、など）により記載された情報が格納されたメタデータでもよい。当該応答において、前述の指示文と同じフォーマットを用いる場合には、前述の初期設定指示文と、ユーザ指示文とは異なる種類の情報であることを示すために種類識別情報を、主たるメッセージ以外の部分に格納してもよい。例えば、大規模言語モデルからの回答文であることを示す情報を格納するなどである。

【００６４】

次に、人工知能応答出力装置１００１０は、大規模言語モデルサーバ１９００１からの応答を受信し、当該応答に主たるメッセージとして格納されている自然言語のテキスト情報を抽出する。人工知能応答出力装置１００１０のキャラクター動作プログラムは、前述

の応答から抽出した自然言語のテキスト情報にもとづいて、音声合成技術を用いてユーザへの回答となる自然言語の音声を生成して、あたかもキャラクター１９０５１の声として聞こえるように、スピーカである音声出力部１１４０から出力する。この処理をキャラクターの「発話」と表現してもよい。

【００６５】

以上説明した、人工知能応答出力装置１００１０および大規模言語モデルサーバ１９００１の処理により、ユーザ２３０からの言葉に対するキャラクター１９０５１の応答音声の具体的な一例を、図２Ｃの会話例１～５に示す。このように、キャラクター１９０５１があたかも実在する人物かのように、ユーザ２３０は会話を行うことができる。

【００６６】

以上説明した、図２Ｂの人工知能応答出力装置１００１０、または人工知能応答出力装置１００１０を含むシステムによれば、学習が膨大なデータ、計算資源が必要な大規模言語モデル自体を、人工知能応答出力装置１００１０自体に搭載する必要はない。そのうえで、ＡＰＩを介して、大規模言語モデルの高度な自然言語処理の能力を利用することができ、ユーザがキャラクターに話しかけた場合に、ユーザへより好適な回答を行い、より好適な会話を行うことが可能となる。

【００６７】

次に、図２Ｄを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図２Ｄは、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１とユーザ２３０の会話の元となる、人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される指示文の主たるメッセージの自然言語テキストとその応答となるサーバ応答の主たるメッセージの自然言語テキストの一例である。

【００６８】

また、図２Ｄでは、表示の設定指示文、ユーザ指示文１巡目とその応答からユーザ指示文４巡目とその応答まで、時系列に指示文と応答のやり取りがなされているものとして示している。

【００６９】

図２Ｄに示すように、設定指示文により、大規模言語モデルサーバ１９００１の人工知能の大規模言語モデルに、大規模言語モデル自体の名前や、演じるべき役割、会話の特徴などを、初期設定の指示することができる。また、ユーザの名前も初期設定として理解させることができる。これにより、大規模言語モデルは、当該役割を守って１巡目以降の応答を生成する。すると、この１巡目以降の応答にもとづくキャラクター１９０５１の音声を聞いたユーザからすると、あたかもキャラクター１９０５１が設定指示文に記載された人物の設定や性格のように感じられる。また、本実施例に係る大規模言語モデルサーバ１９００１は、一連の会話が終了するまで会話の内容を記憶するメモリを備えており、一連のユーザ指示文とその応答を記憶したうえで、応答を生成するように構成されている。これにより、図２Ｄのような会話を実現できる。

【００７０】

次に、図２Ｅを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図２Ｅは、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１とユーザ２３０の会話の元となる、人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される指示文の主たるメッセージの自然言語テキストとその応答となるサーバ応答の主たるメッセージの自然言語テキストの一例である。

【００７１】

図２Ｅは、図２Ｄに示した一連の会話後に当該一連の会話の継続が終了した後に、再びユーザ２３０がキャラクター１９０５１に話しかけて新たに会話を行う場合の一例を示している。図２Ｅでは、ユーザ指示文１巡目とその応答からユーザ指示文３巡目とその応答まで、時系列に指示文と応答のやり取りがなされているものとして示している。

【００７２】

ここで、「一連の会話の継続」の「終了」とは、所定の条件を満たした場合に、大規模言語モデルサーバ１９００１が、一連の会話が継続しているときに保持していた会話の記憶を、大規模言語モデルサーバ１９００１から消去する処理である。所定の条件の一例は、例えば、人工知能応答出力装置１００１０から指示文により大規模言語モデルサーバ１９００１へ「一連の会話の継続」の「終了」を指示した場合が挙げられる。また、所定の条件の別の一例は、例えば、当該一連の会話についての人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１への指示文の送信が無くなって所定時間以上経った場合（タイムアウト）が挙げられる。また、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１との接続において、認証処理を行ったうえで、上記指示文と応答のやり取りを行っているときに通信切断や人工知能応答出力装置１００１０の電源オフ（ＯＦＦ）等の要因で認証処理が外れてしまった場合も挙げられる。

【００７３】

なお、「一連の会話の継続」が「終了」してしまうと、大規模言語モデルサーバ１９００１が、一連の会話が継続しているときに保持していた会話の記憶を、大規模言語モデルサーバ１９００１から消去してしまう。よって、図２Ｅに示される会話は、図２Ｄに示した一連の会話後であるにもかかわらず、ユーザ指示文に対するサーバ応答は、図２Ｄに示した設定指示文に含まれていた、大規模言語モデルに設定したキャラクターとしての名前や、演じるべき役割、会話の特徴、ユーザの名前などを全く記憶していない状態の内容の応答となっている。同様に、図２Ｅに示される会話は、図２Ｄに示した一連の会話の記憶も全く無い状態の内容の応答となっている。すなわち、図２Ｄに示す「一連の会話の継続」の「終了」により、図２Ｅの会話は、大規模言語モデルサーバ１９００１の人工知能の大規模言語モデルが初期化された状態から開始されていることになる。

【００７４】

これは、ユーザ２３０にとって、あたかもキャラクター１９０５１が自分との記憶を失ってしまったか、まるで別人のように感じる要因となる。ユーザ２３０からすると、このキャラクターの応答は違和感を強く感じるものであり、寂しく残念な気持ちになる体験になる。このような動作では、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１の名前、役割、または会話の特徴、性格などの設定や記憶の同一性が確保できない、という課題があった。

【００７５】

次に、図２Ｆを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図２Ｆは、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１とユーザ２３０の会話の元となる、人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される指示文の主たるメッセージの自然言語テキストとその応答となるサーバ応答の主たるメッセージの自然言語テキストの一例である。

【００７６】

図２Ｆは、図２Ｄに示した一連の会話後に当該一連の会話の継続が終了した後に、再びユーザ２３０がキャラクター１９０５１に話しかけて新たに会話を行う場合の一例を示している。図２Ｅの処理とは異なり、図２Ｆの処理では、新たに会話を開始する際に、人工知能応答出力装置１００１０は、大規模言語モデルサーバ１９００１への最初の指示文として設定指示文を送信する。当該設定指示文には、図２Ｄの初期設定の設定指示文と同じ自然言語テキストを格納する。これを再設定テキストと表現してもよい。当該設定指示文

には、続けて、過去の会話の履歴を説明する自然言語テキストを格納する。これを会話履歴テキストと表現してもよい。過去の会話の履歴は、図２Ｄで説明した会話の一連の会話の継続がなされている間に、人工知能応答出力装置１００１０が、その会話の履歴を自然言語テキスト情報としてストレージ部１１７０に会話した日時の情報と紐づけて記録しておけばよい。異なる日付の会話がある場合は、それぞれの会話ごとに日時の情報と紐づけて記録し、会話の履歴を蓄積しておけばよい。図２Ｆのような後日の会話の最初の指示文の設定指示文を生成する際に、ストレージ部１１７０に記録された会話の自然言語テキスト情報とその会話が行われた日時の情報を読み出して、当該設定指示文の生成に用いればよい。

【００７７】

なお、当該設定指示文の生成に過去の会話の履歴の自然言語テキスト情報を用いる場合、大規模言語モデルへ送信するデータであるため、フォーマットはある程度自由に決めて問題無いが、図２Ｆに示すように、「私は、〇月〇日に以下の話をしました。」、「あなたは、〇月〇日に以下の話をしました。」など、自然言語による接頭語や接尾語を用意して、記録されている会話の自然言語テキスト情報と融合させて、当該設定指示文の文面を生成する処理を行えばよい。また、ストレージ部１１７０から読み出した会話の日時の情報は、上記「〇月〇日」の部分などと融合させて当該設定指示文の文面の一部としてもよい。

【００７８】

一連の会話後に当該一連の会話の継続が終了した後に、再びユーザ２３０がキャラクター１９０５１に話しかけて新たに会話を行う場合であっても、以上説明した図２Ｆの設定指示文の生成処理、および送信処理を行えば、その後のユーザ指示文の応答が以前の会話の時点のキャラクターの役割、名前、会話の特徴、性格、および／または会話の特徴などの設定や会話履歴を反映したものとなる。これにより、ユーザからすると、以前の会話の時点の当該キャラクターの役割、名前、会話の特徴または性格などの設定や記憶の同一性がより確保できているように認識されるため、より好適である。

【００７９】

次に、図２Ｇを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図２Ｇは、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１とユーザ２３０の会話の元となる、人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される指示文の主たるメッセージの自然言語テキストとその応答となるサーバ応答の主たるメッセージの自然言語テキストの一例である。

【００８０】

図２Ｇは、図２Ｆに示した一連の会話における、最初の設定指示文以降に続くユーザ指示文１巡目とその応答からユーザ指示文３巡目とその応答までの一連の会話の一例を示している。図２Ｇでは、時系列に指示文と応答のやり取りがなされているものとして示している。設定指示文の内容は図２Ｆに示したとおりなので、繰り返しの記載は省略する。

【００８１】

図２Ｆの表のサーバ応答の自然言語テキストに示されるように、図２Ｆに示した設定指示文を用いることにより、大規模言語モデルサーバ１９００１の大規模言語モデル人工知能によるサーバ応答は、以前の会話の時点のキャラクターの役割、名前、会話の特徴、または性格などの設定や会話履歴を反映したものとなる。これにより、ユーザからすると、以前の会話の時点の当該キャラクターの役割、名前、会話の特徴、または性格などの設定や記憶の同一性がより確保できているように認識されるため、より好適である。なお、これは、ユーザから見てキャラクターが同一視できるというものであるので、ユーザから見たキャラクターの疑似同一性と称してもよい。

【００８２】

また、ユーザからすると、当該キャラクターと記憶を共有することができ、より楽しいキャラクター会話体験を得られる。

【００８３】

次に、図２Ｈを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０の動作の一例について説明する。これは、人工知能応答出力装置１００１０）と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図２Ｈは、人工知能応答出力装置１００１０の表示部１００１１に表示するキャラクターを複数のキャラクター候補の中から切り替えて表示する動作例を示している。人工知能応答出力装置１００１０の制御部１１１０により実行されるキャラクター動作プログラムが、例えば、操作入力部１１０７に入力される操作入力や表示部１００１１のタッチ操作入力センサが検出する操作にもとづいて、表示キャラクターの切り替えを行えばよい。

【００８４】

図２Ｈの例では、図２Ａ～図２Ｇの説明で用いたキャラクター１９０５１（名前は「Koto」）に加えて、キャラクター１９０５２（名前は「Tom」）とキャラクター１９０５３（名前は「Necco」）が示されている。キャラクター１９０５１（名前は「Koto」）とキャラクター１９０５２（名前は「Tom」）は人間のキャラクターであり、キャラクター１９０５３（名前は「Necco」）は猫のキャラクターである。表示部１００１１に表示するキャラクターの表示の切り替えは、キャラクターごとに異なる仮想３Ｄ空間上のキャラクターをレンダリングして生成した映像を表示部１００１１に切り替えて表示すればよい。それぞれのキャラクターの３Ｄモデルのレンダリング映像の表示を実現するための処理は、例えば、図１５Ａで説明した第１～第３の処理例のいずれかを行えばよい。また、キャラクターによっては、動的な２Ｄ画像を表示してもよい。

【００８５】

また、制御部１１１０により実行されるキャラクター動作プログラムは、表示部１００１１に表示するキャラクターの表示の切り替えを行うときには、キャラクターそれぞれの「発話」に用いる合成音声も変更することが、好適である。これは、予め、それぞれのキャラクターに対応付けられた声色の合成音声のデータをストレージ部１１７０に格納しておき、キャラクターの表示の切り替えのときに合成音声変更処理も行えばよい。

【００８６】

なお、図２Ｈの例では、ユーザ２３０がいずれのキャラクターとも会話することができるように、構成されている。図２Ｈの人工知能応答出力装置１００１０では、これらのキャラクターのそれぞれに、異なる役割、名前、会話の特徴、または性格などの設定を行う。また、会話の履歴に基づく各キャラクターの記憶もそれぞれのキャラクターごとに異なるものとして管理する。

【００８７】

そこで、人工知能応答出力装置１００１０は、ストレージ部１１７０に、図２Ｉに示すデータベースを構築し、当該データベースにより、キャラクターの設定とキャラクターの会話履歴を管理する。

【００８８】

次に、図２Ｉを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図２Ｉは、人工知能応答出力装置１００１０の表示部１００１１に表示する複数のキャラクターについての、キャラクターの設定とキャラクターの会話履歴を管理するためのデータベース１９２００の説明図である。

【００８９】

人工知能応答出力装置１００１０の制御部１１１０により実行されるキャラクター動作プログラムが、例えば、ストレージ部１１７０に当該データベース１９２００を構築する。キャラクターＩＤは、人工知能応答出力装置１００１０で表示可能な複数のキャラクタ

ーのそれぞれを識別する識別番号であり、自然数でもよいしアルファベット等を用いてもよい。名前は、人工知能応答出力装置１００１０で表示可能な複数のキャラクターのそれぞれの名前のデータである。

【００９０】

初期設定指示文は、人工知能応答出力装置１００１０で表示可能な複数のキャラクターのそれぞれのキャラクターの役割、名前、会話の特徴、または性格などの設定を説明する自然言語のテキスト情報である。当該初期設定指示文は、人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される設定指示文の主たるデータである自然言語のテキスト情報になるので、そのまま大規模言語モデルサーバ１９００１の人工知能の大規模言語モデルが読み取れる記載内容とすることが望ましい。

【００９１】

会話履歴１、２、…と続く会話履歴は、それぞれのキャラクターとユーザが会話した記録であり、キャラクターごとに切り分けて記録する。当該会話履歴は、人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される設定指示文の主たるデータである自然言語のテキスト情報に含められることになるので、そのまま大規模言語モデルサーバ１９００１の人工知能の大規模言語モデルが読み取れる記載内容とすることが望ましい。

【００９２】

人工知能応答出力装置１００１０の制御部１１１０により実行されるキャラクター動作プログラムは、人工知能応答出力装置１００１０の表示部１００１１に表示するキャラクターを切り替えた場合に、図２Ｉのデータベース１９２００を用いて、人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１へ送信される設定指示文の主たるデータである自然言語のテキスト情報に用いる初期設定指示文や会話履歴を、人工知能応答出力装置１００１０の表示部１００１１に表示するキャラクターに対応するように選択して切り替える。また、キャラクター動作プログラムは、ユーザ２３０とキャラクターの会話が行われるたびに、その会話の履歴を、図２Ｉのデータベース１９２００のうち、表示部１００１１に表示するキャラクターに対応する会話履歴の領域に記録していく。

【００９３】

人工知能応答出力装置１００１０の制御部１１１０により実行されるキャラクター動作プログラムがこのようにデータベース１９２００を用いることにより、同一の大規模言語モデルサーバ１９００１の同一の人工知能の大規模言語モデルの応答を利用したキャラクターの発話を用いて、ユーザ２３０とキャラクターの会話を成立させているにも関わらず、ユーザからすると、キャラクターのそれぞれの性格等の設定の独自性が守られ、かつキャラクターごとに異なる会話の記憶がキャラクターごとに継続していくように感じられる。ユーザからすると、それぞれのキャラクターにおいて、以前の会話の時点の当該キャラクターの役割、名前、会話の特徴、または性格などの設定や記憶の同一性がより確保できているように認識されるため、より好適である。これは、それぞれのキャラクターについて、ユーザから見たキャラクターの疑似同一性が確保できると表現してもよい。

【００９４】

したがって、人工知能応答出力装置１００１０が、表示部１００１１に表示するキャラクターを複数のキャラクター候補の中から切り替えて表示するように構成する場合も、以上説明したデータベース１９２００を用いた動作によれば、ユーザからすると、それぞれのキャラクターとの会話から感じる違和感が少なく、また、複数のキャラクターのそれぞれと記憶を共有することができ、より楽しいキャラクター会話体験を得られる。

【００９５】

なお、複数のキャラクターの初期設定指示文をユーザが編集できないようにすれば、それぞれのキャラクターの役割、名前、会話の特徴、または性格などの設定は、人工知能応答出力装置１００１０の提供者や、キャラクターのコンテンツの制作者の意図に近い状態で維持できる。これに対し、キャラクターの初期設定指示文を操作入力部１１０７などの入力に応じてユーザが編集できるようにしてもよい。この場合は、キャラクターの役割、

名前、会話の特徴、または性格などの設定を好みの設定にすることができ、ユーザが独自に設定したキャラクターと会話することが可能となる。この場合、これに合わせてキャラクターの３Ｄモデルやそのレンダリング映像、およびキャラクターの合成音声の種類を差し替えてもよい。

【００９６】

次に、図２Ｊを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、人工知能応答出力装置１００１０によるキャラクター会話装置や、人工知能応答出力装置１００１０および大規模言語モデルサーバ１９００１によるキャラクター会話システムによるキャラクター会話サービスをより安価に提供する方法について説明する。

【００９７】

図２Ｂにおいて説明したとおり、大規模言語モデルを特定の用途に限ってこのレベルの人工知能の学習を行うことは非常に資源効率が悪い。そのため、様々な用途に応用できる基盤モデル（Ｆｏｕｎｄａｔｉｏｎ　Ｍｏｄｅｌ）として、大規模な学習を行ってモデルを生成し、ＡＰＩ（Ａｐｐｌｉｃａｔｉｏｎ　Ｐｒｏｇｒａｍｍｉｎｇ　Ｉｎｔｅｒｆａｃｅ）を介して、様々な端末で利用するのが、資源効率がよい。すると、大規模言語モデルの提供元は、大規模言語モデルの学習に用いられたコストを端末のＡＰＩの利用料金として、端末の利用者から回収することが多い。そのとき、自然言語モデルでは、トークンと呼ばれる文章を区切った言葉の単位の処理数でＡＰＩの利用料金が請求される形式が多い。

【００９８】

そこで、本発明の実施例２の人工知能応答出力装置１００１０においても、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１の間を、ＡＰＩを用いて伝送される自然言語のテキスト情報内のトークン数の低減を図ることによって、人工知能応答出力装置１００１０によるキャラクター会話装置や、人工知能応答出力装置１００１０および大規模言語モデルサーバ１９００１によるキャラクター会話システムによるキャラクター会話サービスを、より安価にユーザに提供することができる。

【００９９】

例えば、図２Ｊの表に示すような例１～例３の処理や構成とすることにより、人工知能応答出力装置１００１０と大規模言語モデルサーバ１９００１の間を、ＡＰＩを用いて伝送される自然言語のテキスト情報内のトークン数を、技術的に低減することが可能である。

【０１００】

例１は、ＡＰＩの設定指示文に格納されて伝送される会話履歴テキストのトークン数を低減する方法のうち、文書要約処理を用いて会話履歴テキストを短くし、トークン数を低減する例である。例えば、ストレージ部１１７０に記録されるキャラクターとの会話履歴の自然言語自体を要約して記録する。文章要約は、次の会話の開始時に行ってもよいが、「一連の会話」の終了時に行った方が、時間的に余裕がある。

【０１０１】

また、文章要約処理は、大規模言語モデルサーバ１９００１の大規模言語モデル自体に要約を依頼してもよい。ただし、この場合、トークン数の節約効果は低い。よって、例えば、第２のサーバ１９００２において、ＡＰＩを介して、大規模言語モデルサーバ１９００１の大規模言語モデルよりも安価に自然言語の文章要約処理を提供している場合は、ＡＰＩを介して第２のサーバ１９００２に文章要約処理を依頼し、会話履歴の文章要約を大規模言語モデルサーバ１９００１への設定指示文に格納して伝送すればよい。

【０１０２】

また、文章要約処理だけであれば、端末側の処理でも可能であり、人工知能応答出力装置１００１０のメモリ１１０９に展開する文書要約プログラムを制御部１１１０が実行し

て文章要約を行ってもよい。この場合、トークン数の節約効果は高い。また、会話の履歴が長くなってきても、文章要約処理において要約後の文字数の上限を指定すれば、会話の履歴の文章長さの上限が決まるので、トークンの上限値を定めることができ、トークンの節約が可能となる。

【０１０３】

なお、キャラクターの役割、名前、会話の特徴、または性格などのキャラクター初期設定のテキスト情報は、会話の履歴ほど増加しないので、キャラクターの初期設定指示文のテキスト情報の記載は維持し、会話の履歴のテキスト情報のトークン数の低減を図るのが、効率的であり、好ましい。

【０１０４】

例１で説明した処理は、制御部１１１０により実行されるキャラクター動作プログラムが各部を制御して行えばよい。

【０１０５】

例２は、ＡＰＩの設定指示文に格納されて伝送される会話履歴テキストのトークン数を低減する他の例である。例えば、ストレージ部１１７０に記録されるキャラクターとの会話履歴のうち、古い方の履歴から消去してトークン数を低減する。会話履歴の文字数の上限を指定すれば、会話の履歴の文章長さの上限が決まるので、トークンの上限値を定めることができ、トークンの節約が可能となる。または、会話履歴の所定の期間を指定して、当該期間を越えた会話履歴を削除する方法でもよい。この場合も、トークンの節約が可能となる。なお、例２においても、キャラクターの役割、名前、会話の特徴、または性格などのキャラクター初期設定のテキスト情報は、会話の履歴ほど増加しないので、キャラクターの初期設定指示文のテキスト情報の記載は維持し、会話の履歴のテキスト情報のトークン数の低減を図るのが、効率的であり、好ましい。

【０１０６】

例２で説明した処理は、制御部１１１０により実行されるキャラクター動作プログラムが各部を制御して行えばよい。

【０１０７】

例３は、ＡＰＩを用いた設定指示文の送信頻度を減らしてトークン数を低減する方法である。具体的には、装置電源の立ち上げ後や、表示キャラクターの切り替え後に、表示されるキャラクターの映像設定および合成音声設定が完了した後も、設定指示文の事前送信を行わず、マイク１１３９で収音したユーザの発声に含まれる自然言語のテキスト情報が、人工知能の大規模言語モデルを用いるべきテキスト情報である、と制御部１１１０が判定した場合に、初めて設定指示文を大規模言語モデルサーバ１９００１に送信することで、設定指示文の大規模言語モデルサーバ１９００１への送信頻度を減らしてトークン数を低減する。

【０１０８】

具体的には、例えば、装置の電源オン（ＯＮ）後や、表示キャラクターの切り替えの操作入力後に、制御部１１１０により実行されるキャラクター動作プログラムの制御による表示部１００１１の表示処理により、図２Ｈのようにキャラクター１９０５１（名前は「Koto」）が表示部１００１１に表示される。このとき、例えば、キャラクター１９０５１に対応する登場時用の合成音声がストレージ部１１７０などに格納されて用意されている場合、「おはようございます。Kotoです。」、「こんにちは。Kotoです。」、「こんばんは。Kotoです。」などの登場時用の合成音声を音声出力部１１４０であるスピーカから出力してもよい。このとき、既に、表示部１００１１に表示されるキャラクターの映像としてキャラクター１９０５１の映像に設定し、音声出力部１１４０であるスピーカから出力される合成音声は、キャラクター１９０５１に対応する合成音声に設定する。

【０１０９】

ここで、既に説明した大規模言語モデルサーバ１９００１における人工知能の大規模言語モデルの推論処理も、指示文が長くなれば時間がかかる。特に、設定指示文に、過去の会話履歴に関するテキスト情報を含める場合は、指示文のトークン量が多くなるので、特

に推論処理時間が長くなる。設定指示文自体およびその応答はユーザ２３０に出力されない。設定指示文後のユーザ指示文の応答から、キャラクターの「発話」としての合成音声を音声出力部１１４０であるスピーカから出力する。すると、設定指示文は事前に人工知能応答出力装置１００１０から大規模言語モデルサーバ１９００１に送信して、設定指示文に対する大規模言語モデルの推論処理を事前に完了させておいた方が、ユーザ２３０がキャラクター１９０５１に話しかけた後のキャラクター１９０５１の「発話」の合成音声の出力の応答が早くなるので、一見好適に思える。

【０１１０】

しかしながら、ユーザ２３０が言葉を発する前に、大規模言語モデルサーバ１９００１に送信して、設定指示文に対する大規模言語モデルの推論処理を事前に完了させてしまった場合、例えば、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により人工知能応答出力装置１００１０の電源をオフ（ＯＦＦ）してしまう場合や、例えば、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により、キャラクター１９０５１から別のキャラクターへ表示キャラクターを切り替えてしまう場合もあり得る。この場合、事前に設定指示文を大規模言語モデルサーバ１９００１に送信し、大規模言語モデルの推論処理で処理したトークン数は、無駄に利用料金を浪費した処理トークン数となる。これは、人工知能応答出力装置１００１０によるキャラクター会話装置や、人工知能応答出力装置１００１０および大規模言語モデルサーバ１９００１によるキャラクター会話システムによるキャラクター会話サービスを、より安価にユーザに提供することの、妨げとなる。

【０１１１】

よって、人工知能応答出力装置１００１０は、装置の電源オン（ＯＮ）後や、表示キャラクターの切り替えの操作入力後に、制御部１１１０により実行されるキャラクター動作プログラムの制御により、表示部１００１１に表示されるキャラクターの映像としてキャラクター１９０５１の映像を設定し、音声出力部１１４０であるスピーカから出力される合成音声をキャラクター１９０５１に対応する合成音声に設定した後でも、ユーザ２３０がキャラクター１９０５１に話しかけることを認識する時点まで、設定指示文を大規模言語モデルサーバ１９００１に送信しない状態を継続することが望ましい。

【０１１２】

ここで、ユーザ２３０がキャラクター１９０５１に話しかけることを認識する時点とは、例えば、図２Ｂで説明したトリガーとなるキーワードを検出する時点まで、または、ユーザ２３０が話しかけた言葉のテキストの抽出が行われる時点などでよい。このようにすれば、無駄に利用料金を浪費する処理トークン数を低減することができ、人工知能応答出力装置１００１０によるキャラクター会話装置や、人工知能応答出力装置１００１０および大規模言語モデルサーバ１９００１によるキャラクター会話システムによるキャラクター会話サービスを、より安価にユーザに提供することができる。

【０１１３】

また、続いて、上述のユーザ２３０がキャラクター１９０５１に話しかけることを認識する時点を超えた後でも、例えば、マイク１１３９で収音したユーザ２３０の音声から抽出したテキスト情報が、大規模言語モデルの推論処理を必要としないプリセットのキーワードに対応するテキスト情報である場合も、設定指示文を大規模言語モデルサーバ１９００１に送信しない状態を継続することが望ましい。具体的には、プリセットのキーワードの例として、「ジャンプしてみて」、「ダンスしてみて」など、ユーザ２３０がキャラクター１９０５１に対して、キャラクター１９０５１が動くアニメーションや合成音声を発するなどのリアクションを依頼するキーワードであった場合などが挙げられる。この場合は、制御部１１１０により実行されるキャラクター動作プログラムが、ストレージ部に格納されるキャラクター１９０５１に対応するモーションデータ、アニメーション映像、および／または当該リアクションに対応する合成音声データを読み出して、これらのデータを用いて、表示部１００１１に表示する映像の生成処理や、音声出力部１１４０であるスピーカからの合成音声の出力処理を行えばよい。

【０１１４】
このような処理は、必ずしも、大規模言語モデルサーバ１９００１の大規模言語モデルの推論処理を要しない。当該処理の後に、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により人工知能応答出力装置１００１０の電源をオフ（ＯＦＦ）してしまう場合や、例えば、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により、キャラクター１９０５１から別のキャラクターへ表示キャラクターを切り替えてしまう場合も、先に設定指示文を大規模言語モデルサーバ１９００１に送信し、大規模言語モデルの推論処理で処理してしまうと、その処理のトークン数は無駄に利用料金を浪費した処理トークン数となってしまう。

【０１１５】
よって、上述のユーザ２３０がキャラクター１９０５１に話しかけることを認識する時点を超えた後でも、例えば、マイク１１３９で収音したユーザ２３０の音声から抽出したテキスト情報が、大規模言語モデルの推論処理を必要としないプリセットのキーワードに対応するテキスト情報であるか否かを判定する時点までは、設定指示文を大規模言語モデルサーバ１９００１に送信しない状態を継続することが望ましい。当該判定により、大規模言語モデルの推論処理を必要すると判定された場合に、はじめて、設定指示文を大規模言語モデルサーバ１９００１に送信し、大規模言語モデルの推論処理を進めることが望ましい。

【０１１６】
なお、例３で説明した処理は、制御部１１１０により実行されるキャラクター動作プログラムが各部を制御して行えばよい。

【０１１７】
以上説明した、図２Ｊの各例による大規模言語モデルの処理トークン数の低減（節約）方法によれば、人工知能応答出力装置１００１０によるキャラクター会話装置や、人工知能応答出力装置１００１０および大規模言語モデルサーバ１９００１によるキャラクター会話システムによるキャラクター会話サービスを、より安価にユーザに提供することができる。

【０１１８】
次に、図２Ｋを用いて、本発明の実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）の表示の一例について説明する。図２Ｋの例では、図２Ａ～図２Ｊの各図で説明したユーザからの指示文に対する大規模言語モデルからの応答を、キャラクター会話装置（人工知能応答出力装置１００１０）の表示部１００１１に表示する一例を示している。具体的には、大規模言語モデルからの応答であるテキスト１００６３を、キャラクター１９０５１の映像とともに、表示部１００１１に表示する例である。大規模言語モデルからの応答であるテキスト１００６３は、図２Ｋに示すようにキャラクター１９０５１の映像の手前に重畳して表示してもよい。また、大規模言語モデルからの応答であるテキスト１００６３は、キャラクター１９０５１の映像と重畳せずに、キャラクター１９０５１の映像ともに表示してもよい。

【０１１９】
図２Ｋの表示は一例であるが、例えば、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により、キャラクター会話装置（人工知能応答出力装置１００１０）の音声出力部１１４０の音声出力のボリュームを最小に調整するまたは音声出力をＯＦＦに設定した場合などは、ユーザ２３０は音声で大規模言語モデルからの応答を確認することができない。

【０１２０】
そこで、この場合、制御部１１１０は、大規模言語モデルからの応答であるテキスト１００６３を図２Ｋに示すように、キャラクター１９０５１の映像ともに表示する表示モードを開始するように制御してもよい。このようにすれば、音声出力を控えたい場合でも、ユーザ２３０がより好適に、キャラクター会話装置（人工知能応答出力装置１００１０）

を使用することができる。なお、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により、大規模言語モデルからの応答であるテキスト１００６３をキャラクター１９０５１の映像ともに表示する表示モードについてのＯＮ／ＯＦＦを手動で切り替え可能に構成してもよい。

【０１２１】

次に、図２Ｈおよび図２Ｉで説明した、複数のキャラクターを表示可能なキャラクター会話装置（人工知能応答出力装置１００１０）における、応答定型文データベース（応答定型文ＤＢ）の例について、図２Ｌを用いて説明する。図２Ｌの例では、条件番号および条件内容は、図１Ｃと同一である。これらの条件に対し、図２Ｌの例では、複数のキャラクターのそれぞれについて、個別の応答定型文が設定されている。例えば、図２Ｈおよび図２Ｉで説明した、キャラクター１：Ｋｏｔｏ、キャラクター２：Ｔｏｍ、キャラクター３：Ｎｅｃｃｏ、の３つのキャラクターのそれぞれについて、各条件に対する応答定型文が格納されている。応答定型文の出力制御については、図１Ｃと同様であるため、繰り返しの説明は省略する。

【０１２２】

図２Ｌの例では、制御部１１１０は、応答定型文データベース（応答定型文ＤＢ）から、キャラクター会話装置（人工知能応答出力装置１００１０）において表示されているキャラクターと、現在の条件とにもとづいて、対応する応答定型文を選択し、キャラクターが発する応答としての出力制御に用いればよい。例えば、図２Ｌの応答定型文データベース（応答定型文ＤＢ）の例では、同一の条件であっても、応答定型文はキャラクターの個性に対応した表現または内容にそれぞれ変えられている。これにより、キャラクター会話装置（人工知能応答出力装置１００１０）は表示されているキャラクターの個性に対応した会話をユーザに提供することができる。ユーザは、それぞれのキャラクターがより一貫性のある個性を持った存在のように感じることができる。これにより、複数のキャラクターに、より実在感を持たせるキャラクター会話装置（人工知能応答出力装置１００１０）を実現できる。

【０１２３】

なお、以上説明した図２Ｌの応答定型文データベース（応答定型文ＤＢ）はストレージ部１１７０に格納され、人工知能応答出力装置１００１０の制御部１１１０がこれを用いればよい。しかしながら、図２Ｌに示す応答定型文データベース（応答定型文ＤＢ）を大規模言語モデルサーバ１９００１側に備えてもよい。この場合は、大規模言語モデルサーバ１９００１が有する制御部が、当該応答定型文データベース（応答定型文ＤＢ）を用いた応答を生成すればよい。大規模言語モデルサーバ１９００１が有する制御部は、それぞれのサーバに格納される大規模言語モデルにより生成する応答に替えて、当該応答定型文データベース（応答定型文ＤＢ）を用いて生成した応答を人工知能応答出力装置１００１０へ送信すればよい。このようにすれば、人工知能応答出力装置１００１０に応答定型文データベース（応答定型文ＤＢ）が備えられていない場合でも、応答定型文データベース（応答定型文ＤＢ）を用いた応答の生成が可能となる。

【０１２４】

以上説明した、実施例２に係るキャラクター会話装置や、キャラクター会話システムによれば、人工知能応答出力装置１００１０に表示されるキャラクターとの会話からユーザが感じる違和感をより少なくすることができる。また、実施例２に係るキャラクター会話装置や、キャラクター会話システムによれば、キャラクター会話サービスをより安価にユーザに提供することができる。

【０１２５】

なお、実施例２の以上の説明では、大規模言語モデルとして、大規模言語モデルサーバ１９００１が有する大規模言語モデルを用いる例を説明した。これに対し、キャラクター会話装置（人工知能応答出力装置１００１０）が図１Ｂに示すローカルＬＬＭ処理部１００２８を備えるように構成し、大規模言語モデルサーバ１９００１が有する大規模言語モデルに変えて、ローカルＬＬＭ処理部１００２８が有する大規模言語モデルを用いてもよ

い。この場合、実施例２の以上の説明において、大規模言語モデルサーバ１９００１が有する大規模言語モデルをキャラクター会話装置（人工知能応答出力装置１００１０）のローカルＬＬＭ処理部１００２８が有する大規模言語モデルに読み替えればよい。

【０１２６】

この場合も、人工知能応答出力装置１００１０に表示されるキャラクターとの会話からユーザが感じる違和感をより少なくすることができる。なお、大規模言語モデルサーバ１９００１が有する大規模言語モデルに変えて、ローカルＬＬＭ処理部１００２８が有する大規模言語モデルを用いる場合は、処理トークン数に応じた利用料金を考慮する必要性は少なくなるが、ローカルＬＬＭ処理部１００２８が有する大規模言語モデルであっても処理トークン数の低減を図ることによって、推論にかかる電力などの消費リソースを低減できる。この場合、より消費電力の少ないキャラクター会話サービスをユーザに提供することができる。

【０１２７】

なお、実施例２の以上の説明では、キャラクターとの会話履歴をキャラクター会話装置（人工知能応答出力装置１００１０）のストレージ部１１７０に記録し保持する例を説明した。これに対し、キャラクターとの会話履歴は、インターネット１９０００に接続された第２のサーバ１９００２やその他のクラウドサーバに記録し保持してもよい。この場合、キャラクター会話装置（人工知能応答出力装置１００１０）は、新たにユーザとキャラクターが会話を始めるときに、当該第２のサーバ１９００２またはその他のクラウドサーバと通信し、当該キャラクターとユーザの過去の会話履歴を取得（ダウンロード）し、キャラクター会話装置（人工知能応答出力装置１００１０）のストレージ部１１７０やメモリ１１０９に保持して、大規模言語モデルへの指示文の作成に用いればよい。大規模言語モデルへの指示文を過去の会話履歴を用いる具体的な方法は実施例２の各図で説明した通りであるので、繰り返しの説明は省略する。

【０１２８】

また、キャラクター会話装置（人工知能応答出力装置１００１０）は、ユーザとキャラクターの会話が行われる度、または、ユーザとキャラクターの会話が終了した時点など、所定の時点に当該時点までのキャラクターの会話履歴を上述の第２のサーバ１９００２またはその他のクラウドサーバに対して送信（アップロード）すればよい。すなわち、キャラクター会話装置（人工知能応答出力装置１００１０）がキャラクターとの会話履歴を第２のサーバ１９００２またはその他のクラウドサーバに対して所定のタイミングでアップロードし、キャラクター会話装置（人工知能応答出力装置１００１０）が、ユーザがキャラクターと会話を開始するときに、第２のサーバ１９００２またはその他のクラウドサーバから最新の会話履歴をダウンロードし、大規模言語モデルへの指示文の生成に用いればよい。このようにすれば、ユーザが前日に使用したキャラクター会話装置（人工知能応答出力装置１００１０）と、ユーザが今から使用するキャラクター会話装置（人工知能応答出力装置１００１０）とが異なる個体の装置であって同一のキャラクターを表示可能であり、当該異なる個体の装置間の同一キャラクターとユーザが異なるタイミングで複数回にわたって会話を行う場合に、前回の会話から当該キャラクターの記憶が疑似的に引き継がれたかのような会話を実現することができ、ユーザにとってより好適である。

【０１２９】

以上説明した、キャラクター会話装置（人工知能応答出力装置１００１０）が、キャラクターとの会話履歴を第２のサーバ１９００２またはその他のクラウドサーバに対してアップロードおよびダウンロードして、キャラクターの記憶を疑似的に引き継ぐ処理は、図２Ｈや図２Ｉで説明した複数のキャラクターの会話履歴を含むデータベース１９２００を扱う場合でも有効である。すなわち、図２Ｉで説明したデータベース１９２００を第２のサーバ１９００２またはその他のクラウドサーバに対してアップロードおよびダウンロードするように構成すれば、１つのキャラクターのみならず、複数のキャラクターについて、異なる個体の装置間において、複数のキャラクターのそれぞれのキャラクターとユーザが異なるタイミングで複数回にわたって会話を行う場合に、前回の会話からそれぞれのキ

ャラクターの記憶が疑似的に引き継がれたかのような会話を実現することができ、ユーザにとってより好適である。

【０１３０】

＜実施例３＞

次に、本発明の実施例３は、実施例２の各図で説明したキャラクター会話装置（人工知能応答出力装置１００１０）およびキャラクター会話システムを改良したものである。本実施例では、実施例２との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。

【０１３１】

実施例２と同様、実施例３におけるキャラクターは、人工知能である大規模言語モデルのサービスをユーザに提供し、ユーザの助力になることができる。よって、当該キャラクターはユーザにとって人工知能（ＡＩ）アシスタントとなることができる。この場合、本実施例におけるキャラクター会話装置や、キャラクター会話システムは、ＡＩアシスタント会話装置、ＡＩアシスタント表示装置、ＡＩアシスタント応答出力装置や、ＡＩアシスタント会話システム、ＡＩアシスタント表示システム、ＡＩアシスタント応答出力システムと称してもよい。

【０１３２】

図３Ａを用いて、本発明の実施例３のキャラクター会話装置とキャラクター会話システムの一例について説明する。実施例３のキャラクター会話システムでは、図２Ａの大規模言語モデルサーバ１９００１に替えて大規模言語モデルサーバ２００００１が備えられ、インターネット１９０００に接続されている。

【０１３３】

ここで、大規模言語モデルサーバ２００００１は、大規模言語モデル人工知能を搭載したサーバであるが、大規模言語モデルサーバ１９００１で処理可能であった自然言語のテキスト情報に加えて自然言語のテキスト情報以外の種類の情報も併せて処理が可能なマルチモーダルな大規模言語モデル人工知能である。

【０１３４】

また、キャラクター会話装置である人工知能応答出力装置１００１０は、一例として、実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）と同等の構成を有するものとして説明する。

【０１３５】

実施例３においても、キャラクター会話装置である人工知能応答出力装置１００１０は、ＡＰＩを用いて、インターネット１９０００を介して大規模言語モデルサーバ２０００１の大規模言語モデルと通信可能である。

【０１３６】

実施例３のキャラクター会話システムには、ユーザ２３０が使用するモバイル情報処理端末２００１０が存在する。モバイル情報処理端末２００１０はいわゆるスマートフォンやタブレット情報処理端末である。

【０１３７】

ここで、図３Ｂを用いて、モバイル情報処理端末２００１０の一例について説明する。モバイル情報処理端末２００１０は、タッチ操作入力パネルである表示パネル２００１１、制御部２００１２、外部電源入力インタフェース２００１３、電源２００１４、二次電池２００１５、ストレージ部２００１６、映像制御部２００１７、姿勢センサ２００１８、通信部２００２０、音声出力部２００２１、マイク２００２２、映像信号入力部２００２３、音声信号入力部２００２４、撮像部２００２５、等を備えている。

【０１３８】

表示パネル２００１１は、タッチ操作入力センサが備えられており、ユーザ２３０の指によるタッチ操作入力を受け付けることができる。表示パネル２００１１は、液晶パネルや有機ＥＬパネルで表示を行うものであり、映像を表示することができる。表示パネル２００１１は表示部と称してもよい。

【０１３９】
通信部２００２０は、Ｗｉ－Ｆｉ方式の通信インタフェース、Ｂｌｕｅｔｏｏｔｈ方式の通信インタフェース、４Ｇ、５Ｇなどの移動体通信インタフェースなどで構成すればよい。これらの通信方式を用いて、モバイル情報処理端末２００１０の通信部２００２０は、キャラクター会話装置（人工知能応答出力装置１００１０）の通信部１１３２と通信可能である。モバイル情報処理端末２００１０にはＣＰＵなどの制御部およびメモリが備えられており、当該制御部は、表示パネル２００１１や通信部２００２０などを制御する。また、通信部２００２０のいずれかの通信方式により、通信部２００２０はインターネット１９０００に接続された通信装置１９０１１と通信可能である。これにより、モバイル情報処理端末２００１０はインターネット１９０００に接続された各種サーバと通信可能である。

【０１４０】
電源２００１４は、外部から外部電源入力インタフェース２００１３を介して入力されるＡＣ電流をＤＣ電流に変換し、モバイル情報処理端末２００１０の各部にそれぞれ必要なＤＣ電流を供給する。二次電池２００１５は、電源２００１４から供給される電力を蓄電する。また、二次電池２００１５は、外部電源入力インタフェース２００１３を介して、外部から電力が供給されない場合に、電力を必要とする各部に対して電力を供給する。

【０１４１】
映像信号入力部２００２３は、外部の映像出力装置を接続して映像データを入力する。映像信号入力部２００２３は、様々なデジタル映像入力インタフェースが考えられる。例えば、ＨＤＭＩ（登録商標）（High－Definition Multimedia Interface）規格の映像入力インタフェース、ＤＶＩ（Digital Visual Interface）規格の映像入力インタフェース、またはＤｉｓｐｌａｙＰｏｒｔ規格の映像入力インタフェースなどで構成すればよい。または、アナログＲＧＢや、コンポジットビデオなどのアナログ映像入力インタフェースを設けてもよい。映像信号入力部２００２３は、各種ＵＳＢインタフェースなどでもよい。

【０１４２】
音声信号入力部２００２４は、外部の音声出力装置を接続して音声データを入力する。音声信号入力部２００２４は、ＨＤＭＩ規格の音声入力インタフェース、光デジタル端子インタフェース、または、同軸デジタル端子インタフェース、などで構成すればよい。音声信号入力部２００２４は、各種ＵＳＢインタフェースなどでもよい。ＨＤＭＩ規格のインタフェースの場合は、映像信号入力部２００２３と音声信号入力部２００２４とは、端子およびケーブルが一体化したインタフェースとして構成されてもよい。

【０１４３】
音声出力部２００２１は、音声信号入力部２００２４に入力された音声データに基づいた音声を出力することが可能である。音声出力部２００２１は、ストレージ部２００１６に格納されている音声データに基づいた音声を出力することも可能である。音声出力部２００２１は、スピーカで構成してもよい。また、音声出力部２００２１は、内蔵の操作音やエラー警告音を出力してもよい。または、ＨＤＭＩ規格に規定されるＡｕｄｉｏ Ｒｅｔｕｒｎ Ｃｈａｎｎｅｌ機能のように、外部機器にデジタル信号として出力する構成を音声出力部２００２１としてもよい。

【０１４４】
マイク２００２２は、モバイル情報処理端末２００１０の周辺の音を収音し、信号に変換して音声信号を生成するマイクである。ユーザの声など人物の声をマイクが収録して、生成した音声信号を後述する制御部２００１２が音声認識処理を行って、当該音声信号から文字情報を取得するように構成してもよい。

【０１４５】
撮像部２００２５は、イメージセンサを有するカメラである。モバイル情報処理端末２００１０の表示パネル２００１１側の前面にカメラを設けてもよく、表示パネル２００１１側の背面にカメラを設けてもよい。前面のカメラと背面のカメラの両者を設けてもよい

。本実施例では、撮像部２００２５は、前面のカメラと背面のカメラの両者を有するものとして説明する。

【０１４６】

ストレージ部２００１６は、映像データ、画像データ、音声データ等の各種データなどの各種情報を記録する記憶装置である。ストレージ部２００１６は、ハードディスクドライブ（ＨＤＤ）などの磁気記録媒体記録装置や、ソリッドステートドライブ（ＳＳＤ）などの半導体素子メモリで構成してもよい。ストレージ部２００１６には、例えば、製品出荷時に予め映像データ、画像データ、音声データ等の各種データ等の各種情報が記録されていてもよい。また、ストレージ部２００１６は、通信部２００２０を介して外部機器や外部のサーバ等から取得した映像データ、画像データ、音声データ等の各種データ等の各種情報を記録してもよい。ストレージ部２００１６に記録された映像データ、画像データ等は、表示パネル２００１１に出力される。ストレージ部２００１６に記録された映像データ、画像データ等を、通信部２００２０を介して外部機器や外部のサーバ等に出力してもよい。

【０１４７】

映像制御部２００１７は、表示パネル２００１１に入力する映像信号に関する各種制御を行う。映像制御部２００１７は、映像処理回路と称してもよく、例えば、ＡＳＩＣ、ＦＰＧＡ、映像用プロセッサなどのハードウェアで構成されてもよい。なお、映像制御部２００１７は、映像処理部、画像処理部と称してもよい。映像制御部２００１７は、例えば、メモリ２００２６に記憶させる映像信号と、映像信号入力部２００２３に入力された映像信号（映像データ）等のうち、どの映像信号を表示パネル２００１１に入力するかといった映像切り替えの制御等を行う。また、映像制御部２００１７は、映像信号入力部２００２３から入力された映像信号やメモリ２００２６に記憶させる映像信号等に対して画像処理を行う制御を行ってもよい。画像処理としては、例えば、画像の拡大、縮小、変形等を行うスケーリング処理、輝度を変更するブライト調整処理、画像のコントラストカーブを変更するコントラスト調整処理、画像を光の成分に分解して成分ごとの重みづけを変更するレティネックス処理等がある。

【０１４８】

姿勢センサ２００１８は、重力センサまたは加速度センサ、またはこれらの組み合わせにより構成されるセンサであり、モバイル情報処理端末２００１０の姿勢を検出することができる。姿勢センサ２００１８の姿勢検出結果に基づいて、制御部２００１２が、接続される各部の動作を制御してもよい。

【０１４９】

不揮発性メモリ２００２７は、モバイル情報処理端末２００１０で用いる各種データを格納する。不揮発性メモリ２００２７に格納されるデータには、例えば、モバイル情報処理端末２００１０の表示パネル２００１１に表示する各種操作用のデータ、表示アイコン、ユーザの操作が操作するためのオブジェクトのデータやレイアウト情報等が含まれる。メモリ２００２６は、表示パネル２００１１に表示する映像データや装置の制御用データ等を記憶する。制御部２００１２がストレージ部２００１６から各種ソフトウェアを読み出して、メモリ２００２６に展開して記憶してもよい。

【０１５０】

制御部２００１２は、接続される各部の動作を制御する。また、制御部２００１２は、メモリ２００２６に記憶されるプログラムと協働して、モバイル情報処理端末２００１０内の各部から取得した情報に基づく演算処理を行ってもよい。

【０１５１】

次に、図３Ｃを用いて、本発明の実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ２０００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。実施例３においても、キャラクター会話装置（人工知能応答出力装置１００１０）は、ストレージ部１１７０などに格納しているキャラクター動作プログラ

ムをメモリ１１０９に展開し、当該キャラクター動作プログラムを制御部１１１０が実行することにより、以下に説明する各種処理が実現可能となる。

【０１５２】

実施例２では、キャラクター会話装置（人工知能応答出力装置１００１０）に対して、ユーザ２３０が行うアクションは主にユーザ２３０の声による呼びかけであった。実施例２のキャラクター会話装置（人工知能応答出力装置１００１０）では、ユーザ２３０の声をマイクで収音する処理から一連の動作を行っていた。これに対し、実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）では、実施例２において説明した、キャラクター会話装置（人工知能応答出力装置１００１０）がユーザ２３０の声をマイクで収音する処理から行う一連の動作も実行できるものとする。これに加えて、実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）では、ユーザ２３０が図１Ｂの操作入力部１１０７を介したユーザ操作により、キャラクター会話装置（人工知能応答出力装置１００１０）に対して、ユーザ２３０がアクションを行うことができる。ここで、図１Ｂの操作入力部１１０７の一例としては、マウス、キーボード、タッチパネルなどが挙げられる。

【０１５３】

また、実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）では、ユーザ２３０が図１Ｂの表示部１００１１のタッチ操作入力センサにより検出可能なユーザのタッチ操作により、キャラクター会話装置（人工知能応答出力装置１００１０）に対して、ユーザ２３０がアクションを行うことができる。

【０１５４】

また、ユーザ２３０が、モバイル情報処理端末２００１０を操作して、キャラクター会話装置（人工知能応答出力装置１００１０）に対してモバイル情報処理端末２００１０から通信を行うことにより、ユーザ２３０の操作入力をキャラクター会話装置（人工知能応答出力装置１００１０）に入力することも可能である。

【０１５５】

また、モバイル情報処理端末２００１０の表示パネル２００１１に、ユーザがキャラクター会話装置（人工知能応答出力装置１００１０）に伝達したい情報を格納した二次元コードなどの情報格納画像を表示させて、当該表示をキャラクター会話装置（人工知能応答出力装置１００１０）が有する図１Ｂの撮像部１１８０が撮像するようにしてもよい。キャラクター会話装置（人工知能応答出力装置１００１０）の制御部１１１０が、撮像部１１８０が撮像した二次元コードなどの情報格納画像から情報を抽出し、当該情報を得てもよい。また、モバイル情報処理端末２００１０の表示パネル２００１１に、ユーザがキャラクター会話装置（人工知能応答出力装置１００１０）に伝達したい画像を表示させて、当該表示をキャラクター会話装置（人工知能応答出力装置１００１０）が有する図１Ｂの撮像部１１８０が撮像するようにしてもよい。キャラクター会話装置（人工知能応答出力装置１００１０）の制御部１１１０が、撮像部１１８０が撮像した画像に対して画像認識処理を行って、当該画像認識処理の結果を取得してもよい。

【０１５６】

このように、実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）では、実施例２において説明した、キャラクター会話装置（人工知能応答出力装置１００１０）よりも、ユーザ２３０からキャラクター会話装置（人工知能応答出力装置１００１０）に対して、行い得るアクションの種類が増えている。これにより、実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）は、ユーザの声以外のユーザ２３０が行うアクションの結果を取得し、それにもとづいて大規模言語モデルサーバ２０００１へ送信する指示文（プロンプト）を生成することができる。これにより、大規模言語モデルサーバ２０００１へ送信する指示文にユーザの声から抽出した自然言語のテキスト情報以外の種類の情報をより好適に含めることができる。ユーザの声から抽出した自然言語のテキスト情報以外の種類の情報とは、例えば、画像、動画、音声などである。

【０１５７】

次に、本実施例のキャラクター会話装置（人工知能応答出力装置１００１０）はＡＰＩを用いて指示文を大規模言語モデルサーバ２０００１へ送信する。本実施例においても、指示文はマークアップ言語のマークアップ形式などのタグを用いた記法や、マークダウン形式などの所定の記号を用いた記法、またはＪＳＯＮなどの所定のスクリプトのオブジェクト記法、などにより記載された情報が格納されたメタデータなどでよい。本実施例においても、指示文の種類としては、初期設定などの指示を格納する設定指示文と、ユーザからの指示を反映するユーザ指示文とがある。指示文が設定指示文であるかユーザ指示文であるかを識別する種類識別情報を、当該指示文の主たるメッセージ以外の部分に格納しておいてもよい。このとき、指示文には、主たるメッセージとして自然言語のテキスト情報が含められている。さらに、本実施例では、指示文の主たるメッセージに、自然言語のテキスト情報以外に、自然言語のテキスト情報以外の種類の情報として、例えば、画像、動画、または音声などの非自然言語情報源を含めることができる。指示文に非自然言語情報源を含める具体的な方法については後述する。

【０１５８】

本実施例の大規模言語モデルサーバ２０００１は、自然言語のテキスト情報と併せて非自然言語情報源を処理できるマルチモーダルな大規模言語モデルを有する。大規模言語モデルサーバ２０００１は、キャラクター会話装置（人工知能応答出力装置１００１０）から指示文を受信する。当該指示文にもとづいて、マルチモーダルな大規模言語モデルが推論を実行し、推論の結果である自然言語のテキスト情報を含む応答を生成する。ここで、当該大規模言語モデルサーバ２０００１の人工知能はマルチモーダルな大規模言語モデルであるので、当該応答に、自然言語のテキスト情報に加えて、画像、動画、または音声などの非自然言語情報源を含めることができる。

【０１５９】

キャラクター会話装置（人工知能応答出力装置１００１０）は、大規模言語モデルサーバ２０００１からの応答を受信し、当該応答に主たるメッセージとして格納されている自然言語のテキスト情報および画像、動画、または音声などの非自然言語情報源を抽出する。キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムは、前述の応答から抽出した自然言語のテキスト情報にもとづいて、音声合成技術を用いてユーザへの回答となる自然言語の音声を生成して、あたかも表示画面に表示しているキャラクター１９０５１の声として聞こえるように、スピーカである音声出力部１１４０から出力してもよい。

【０１６０】

また、キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムは、前述の応答から抽出した自然言語のテキスト情報にもとづいて、ユーザへの回答となる自然言語の文字を、キャラクター会話装置（人工知能応答出力装置１００１０）の表示画面に表示してもよい。このとき、当該文字は、キャラクター１９０５１とともに表示してもよく、キャラクター１９０５１の映像に重畳して表示してもよく、キャラクター１９０５１の映像に替えて表示してもよい。これらの具体的な処理は、映像制御部１１６０が実行すればよい。

【０１６１】

また、キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムは、前述の応答から抽出した非自然言語情報源の画像の情報にもとづいて、ユーザへ提示するために、当該画像をキャラクター会話装置（人工知能応答出力装置１００１０）の表示画面に表示してもよい。このとき、当該画像は、キャラクター１９０５１とともに表示してもよく、キャラクター１９０５１の映像に重畳して表示してもよく、キャラクター１９０５１の映像に替えて表示してもよい。これらの具体的な処理は、映像制御部１１６０が実行すればよい。

【０１６２】

また、キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムは、前述の応答から抽出した非自然言語情報源の動画の情報にもとづいて、ユ

ーザへ提示するために、当該動画をキャラクター会話装置（人工知能応答出力装置１００１０）の表示画面に表示してもよい。このとき、当該動画は、キャラクター１９０５１とともに表示してもよく、キャラクター１９０５１の映像に重畳して表示してもよく、キャラクター１９０５１の映像に替えて表示してもよい。これらの具体的な処理は、映像制御部１１６０が実行すればよい。

【０１６３】

また、キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムは、前述の応答から抽出した非自然言語情報源の音声の情報にもとづいて生成した音声をスピーカである音声出力部１１４０から出力してもよい。

【０１６４】

以上説明した、図３Ｃのキャラクター会話装置（人工知能応答出力装置１００１０）、またはキャラクター会話装置（人工知能応答出力装置１００１０）と大規模言語モデルサーバ２０００１とを含むキャラクター会話システムによれば、学習が膨大なデータ、計算資源が必要な大規模言語モデル自体を、キャラクター会話装置（人工知能応答出力装置１００１０）自体に搭載する必要はない。そのうえで、ＡＰＩを介して、マルチモーダルな大規模言語モデルの高度な自然言語処理および非自然言語情報処理の能力を利用することができる。ユーザからのキャラクターに対するアクションに対して、自然言語のテキストに基づく回答以外に、非自然言語情報源に基づく回答を行うことができ、より好適な会話を行うことが可能となる。

【０１６５】

次に、図３Ｄを用いて、本発明の実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ２０００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図３Ｄは、キャラクター会話装置（人工知能応答出力装置１００１０）から大規模言語モデルサーバ２０００１へ送信される指示文の主たるメッセージの自然言語テキストと画像などの非自然言語情報源の例と、その応答となるサーバ応答の主たるメッセージの自然言語テキストと画像などの非自然言語情報源の一例である。本実施例において、非自然言語情報源は、画像、動画、音声などを用いることが可能であるが、図３Ｄでは、非自然言語情報源として画像の例を示している。

【０１６６】

また、図３Ｄでは、設定指示文、ユーザ指示文１巡目とその応答からユーザ指示文２巡目とその応答まで、時系列に指示文と応答のやり取りがなされているものとして示している。ここで、図３Ｄに示される指示文と応答には、実施例２の図２Ｄには示されていなかった、非自然言語情報源２００６１と非自然言語情報源２００６２が含まれている。図３Ｄの例では、非自然言語情報源２００６１と非自然言語情報源２００６２はいずれも画像である。

【０１６７】

ここで、図３Ｄでは、説明を簡単にするため、指示文の中に、非自然言語情報源２００６１の画像が貼り付けられた状態で示されている。しかしながら、キャラクター会話装置（人工知能応答出力装置１００１０）から大規模言語モデルサーバ２０００１へ送信する指示文における、非自然言語情報源２００６１のデータの伝送またはデータの指定の方法には、複数の方法がある。キャラクター会話装置（人工知能応答出力装置１００１０）はこれらの複数の方法のいずれか一つを用いる、または切り替えて用いればよい。以下それぞれの方法の一例について、説明する。

【０１６８】

指示文において非自然言語情報源データの伝送または指定する第１の方法は、例えば、指定したい非自然言語情報源がインターネット等のネットワークに接続されるサーバ等の場所に存在する非自然言語情報源である場合などに用いる。第１の方法の具体的な方法としては、指示文内のタグや記号などの情報を用いてインターネット等のネットワーク上に存在する非自然言語情報源ファイルを、インターネット等のネットワークのロケーション

情報（いわゆるＵＲＬなど）とファイル名で指定する方法である。
【０１６９】
例えば、マークアップ言語などで画像を指定するタグである〈img src=“****”〉を用いて、****の部分に、画像ファイルのロケーション情報およびファイル名情報を記載することによって、インターネット等のネットワーク上に存在する画像を指定してもよい。また、マークアップ言語などで動画を指定するタグである〈video src=“****”〉を用いて、****の部分に、動画ファイルのロケーション情報およびファイル名情報を記載することによって、インターネット等のネットワーク上に存在する動画を指定してもよい。また、マークアップ言語などで音声を指定するタグである〈audio src=“****”〉を用いて、****の部分に、音声ファイルのロケーション情報およびファイル名情報を記載することによって、インターネット等のネットワーク上に存在する音声を指定してもよい。また、ＪＳＯＮ形式の記法であれば、img_srcなどのキーを用意し、値に画像ファイルのロケーション情報およびファイル名情報を記載することによって、インターネット等のネットワーク上に存在する画像を指定してもよい。動画ファイルや音声ファイルの場合も、それぞれのキーと値を用意すればよい。当該フォーマットの具体例は一例であり、他の独自のフォーマットを用いてもよい。いずれの場合も、非自然言語情報源ファイルのロケーション情報およびファイル名情報を指定する情報を指示文に格納すればよい。
【０１７０】
第１の方法のように、非自然言語情報源ファイルのロケーション情報およびファイル名情報を指定する情報を、指示文に格納する場合は、指示文自体には非自然言語情報源ファイルのデータ自体を格納する必要はない。よって、指示文のデータ量を削減することができる。第１の方法で、非自然言語情報源データが指定された指示文を受信した大規模言語モデルサーバ２０００１は、当該指示文に格納された非自然言語情報源ファイルのロケーション情報およびファイル名情報を用いて、インターネット等のネットワークに接続されるサーバ等の場所にある非自然言語情報源ファイルを取得すればよい。
【０１７１】
ここで、キャラクター会話装置（人工知能応答出力装置１００１０）が、第１の方法で、指示文において非自然言語情報源データの指定を行う場合、ロケーション情報およびファイル名情報の入力をどのように行うかについて、説明する。図３Ｃにおいて、本実施例では、ユーザ２３０からキャラクター会話装置（人工知能応答出力装置１００１０）に対して、行い得るアクションの種類が、実施例２に比べて、ユーザ２３０の声以外にも増えていることについて説明した。よって、例えば、ユーザ２３０が図１Ｂの操作入力部１１０７を介したユーザ操作（例えば、マウス、キーボード、タッチパネル）により、非自然言語情報源データの指定のためのＵＲＬなどのロケーション情報や、ファイル名情報などを入力してもよい。
【０１７２】
また、キャラクター会話装置（人工知能応答出力装置１００１０）において、制御部１１１０がメモリ１１０９と協働して、ＷＥＢブラウザプログラムを実行し、キャラクター会話装置（人工知能応答出力装置１００１０）の表示画面に当該ＷＥＢブラウザプログラムのＧＵＩを表示してもよい。当該ＷＥＢブラウザプログラムのＧＵＩに対するユーザ操作を、操作入力部１１０７を介したユーザ操作（例えば、マウス、キーボード、タッチパネル）または表示部１００１１のタッチ操作入力センサにより検出可能なユーザのタッチ操作により受け付けて、ＷＥＢブラウザプログラムのブラウザ画面で選択した、画像、動画、音声などの非自然言語情報源データを、指示文における指定対象のデータとしてもよい。この場合、ＷＥＢブラウザプログラムが、当該非自然言語情報源データのロケーション情報およびファイル名情報を取得し、キャラクター動作プログラムに引き渡せばよい。
【０１７３】
また、ユーザ２３０が、モバイル情報処理端末２００１０を操作して、キャラクター会話装置（人工知能応答出力装置１００１０）に対してモバイル情報処理端末２００１０から通信を行うことにより、非自然言語情報源データの指定のためのＵＲＬなどのロケーシ

ョン情報をキャラクター会話装置（人工知能応答出力装置１００１０）に入力してもよい。また、図３Ｃで説明した、モバイル情報処理端末２００１０の表示パネル２００１１に二次元コードなどの情報格納画像を表示して、キャラクター会話装置（人工知能応答出力装置１００１０）の撮像部１１８０が撮像した画像に対して画像認識処理を行って、当該画像認識処理の結果を取得する方式で、非自然言語情報源データの指定のためのＵＲＬなどのロケーション情報や、ファイル名情報などを入力してもよい。

【０１７４】

なお、指示文において非自然言語情報源データの伝送または指定する第１の方法の使用は、非自然言語情報源ファイルが予めインターネット等のネットワークに接続されるサーバ等の場所に存在する場合に限られない。例えば、キャラクター会話装置（人工知能応答出力装置１００１０）のストレージ部１１７０に格納されている画像、動画、音声などの非自然言語情報源データを指示文に含めたい場合、キャラクター会話装置（人工知能応答出力装置１００１０）は、インターネット１９０００を介して第２のサーバ１９００２に当該非自然言語情報源データをアップロードし、アップロードした第２のサーバ１９００２の当該非自然言語情報源データのインターネット上でのロケーション情報（いわゆるＵＲＬなど）とファイル名を指示文に含めてもよい。この場合、第２のサーバ１９００２はいわゆる中間サーバとして機能する。

【０１７５】

同様に、モバイル情報処理端末２００１０のストレージ部２００１６に格納されている画像、動画、音声などの非自然言語情報源データを指示文に含めたい場合、モバイル情報処理端末２００１０が、インターネット１９０００を介して第２のサーバ１９００２に当該非自然言語情報源データをアップロードしてもよい。モバイル情報処理端末２００１０または第２のサーバ１９００２から第２のサーバ１９００２の当該非自然言語情報源データのインターネット上でのロケーション情報（いわゆるＵＲＬなど）とファイル名を、キャラクター会話装置（人工知能応答出力装置１００１０）に送信し、キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムが、取得した、第２のサーバ１９００２にアップロードされた当該非自然言語情報源データのインターネット上でのロケーション情報（いわゆるＵＲＬなど）とファイル名を指示文に含めてもよい。

【０１７６】

さらに、キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムが、メモリ１１０９、ストレージ部１１７０と協働してインターネット１９０００を介して他のサーバからアクセスできるメディアサーバをキャラクター会話装置（人工知能応答出力装置１００１０）内に構築してもよい。この場合、キャラクター会話装置（人工知能応答出力装置１００１０）は、第１の方法で、指示文において非自然言語情報源データの指定を行う場合に、キャラクター会話装置（人工知能応答出力装置１００１０）自身の内部に構築したメディアサーバ内を示すインターネット上でのロケーション情報（いわゆるＵＲＬなど）と該当する非自然言語情報源データのファイル名を、前記指示文に格納すればよい。

【０１７７】

次に、指示文において非自然言語情報源データの伝送または指定を指定する第２の方法は、例えば、単純に指示文（プロンプト）に、非自然言語情報源データそのものを格納（添付）して送信する方法である。一般に、画像、動画、音声等の非自然言語情報源データは、自然言語であるテキスト情報に比べてデータ量が大きい。よって、この場合、指示文（プロンプト）自体のデータ量は第１の方法よりも大きくなる。キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムは、指示文（プロンプト）に格納（添付）したい非自然言語情報源データを、一旦メモリ１１０９に格納して、指示文（プロンプト）を送信する際に、メモリ１１０９から通信部１１３２を介して、指示文（プロンプト）に格納（添付）して大規模言語モデルサーバ２０００１へ出力すればよい。キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プ

ログラムがメモリ１１０９に格納する非自然言語情報源データ自体は、インターネット１９０００を介して通信部１１３２が取得してもよく、モバイル情報処理端末２００１０から通信部１１３２が取得してもよく、ストレージ部１１７０から読みだしてメモリ１１０９に格納してもよい。

【０１７８】

以上説明した方法により、キャラクター会話装置（人工知能応答出力装置１００１０）は、指示文による非自然言語情報源データの伝送または指定を行うことが可能である。

【０１７９】

大規模言語モデルサーバ２０００１は、自然言語のテキスト情報と併せて非自然言語情報源を処理できるマルチモーダルな大規模言語モデルであるので、図３Ｄの例に示す、ユーザ指示文１巡目により、非自然言語情報源２００６１であるスイミングプールとプールサイドの画像と、自然言語であるテキスト情報を取得して、その推論結果として、ユーザ指示文１巡目への応答として、図に示すような、自然言語であるテキスト情報を出力することができる。

【０１８０】

また、大規模言語モデルサーバ２０００１は、自然言語のテキスト情報と併せて非自然言語情報源を処理できるマルチモーダルな大規模言語モデルであるので、図３Ｄの例に示す、ユーザ指示文２巡目の応答に示すように、大規模言語モデルサーバ２０００１はマルチモーダルな大規模言語モデルの推論により生成した非自然言語情報源２００６２を応答に含めてキャラクター会話装置（人工知能応答出力装置１００１０）へ送信することができる。図３Ｄでは、非自然言語情報源２００６２は、非自然言語情報源２００６１であるスイミングプールとプールサイドの画像に丸の画像を付した画像の例を示している。なお、応答に格納する非自然言語情報源２００６２は図３Ｄに示す画像に限られず、動画でも音声でもよい。

【０１８１】

大規模言語モデルサーバ２０００１からの応答に、自然言語のテキスト情報以外の非自然言語情報源を含める場合の方法も、上述のキャラクター会話装置（人工知能応答出力装置１００１０）が指示文において非自然言語情報源データの伝送または指定を行う第１の方法にまたは第２の方法に準ずる方法を用いればよい。

【０１８２】

具体的には、上述の第１の方法に準ずる方法として、大規模言語モデルサーバ２０００１は、応答に、非自然言語情報源ファイルのロケーション情報およびファイル名情報を指定する情報を指示文に格納すればよい。画像、動画、音声などの非自然言語情報源２００６２自体は、大規模言語モデルサーバ２０００１に保持しておいてもよいし、中間サーバとして機能させる第２のサーバ１９００２に当該非自然言語情報源２００６２を移送して、保持させておいてもよい。いずれの場合も、大規模言語モデルサーバ２０００１は、応答に、非自然言語情報源ファイルのロケーション情報およびファイル名情報を指定する情報を指示文に格納すればよい。応答を取得したキャラクター会話装置（人工知能応答出力装置１００１０）は、指示文に記載される非自然言語情報源ファイルのロケーション情報およびファイル名情報を用いて、大規模言語モデルサーバ２０００１や第２のサーバ１９００２にアクセスして、非自然言語情報源２００６２を取得すればよい。

【０１８３】

また、具体的には、上述の第２の方法に準ずる方法として、大規模言語モデルサーバ２０００１は、応答に、非自然言語情報源２００６２のファイルデータそのものを格納（添付）してキャラクター会話装置（人工知能応答出力装置１００１０）に送信してもよい。キャラクター会話装置（人工知能応答出力装置１００１０）は、指示文に格納（添付）された非自然言語情報源２００６２のデータを取得して、ユーザ２３０に対する各種出力に用いることできる。

【０１８４】

以上図３Ｄを用いて説明した、実施例３のキャラクター会話装置（人工知能応答出力装

置１００１０）およびキャラクター会話システムの動作によれば、キャラクター会話装置（人工知能応答出力装置１００１０）に表示されるキャラクターとユーザ２３０の間において、非自然言語情報である画像、動画、音声を用いた会話を実現するための指示文と応答の送受信を行う。これにより、図３Ｄの各メッセージに示すようなより高度で自然な会話を実現することが可能となる。

【０１８５】

次に、図３Ｅを用いて、本発明の実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ２００００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図３Ｅは、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１とユーザ２３０の会話の元となる、人工知能応答出力装置１００１０から大規模言語モデルサーバ２００００１へ送信される指示文の主たるメッセージとその応答となるサーバ応答の主たるメッセージの一例である。

【０１８６】

図３Ｅは、図３Ｄに示した一連の会話後に当該一連の会話の継続が終了した後に、再びユーザ２３０がキャラクター１９０５１に話しかけて新たに会話を行う場合の一例を示している。図３Ｅの例では、実施例２の図２Ｆ、図２Ｇ、図２Ｉで説明したような、会話履歴を用いた処理を行っていない。よって、図３Ｅは、実施例２の図２Ｅと同様に、設定指示文に含まれていた大規模言語モデル自体の名前や、演じるべき役割、会話の特徴、ユーザの名前、会話の履歴などを全く記憶していない状態の内容の応答となっている。

【０１８７】

次に、図３Ｆを用いて、本発明の実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ２００００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図３Ｆは、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１とユーザ２３０の会話の元となる、人工知能応答出力装置１００１０から大規模言語モデルサーバ２００００１へ送信される指示文の主たるメッセージとその応答となるサーバ応答の主たるメッセージの一例である。

【０１８８】

図３Ｆは、図３Ｄに示した一連の会話後に当該一連の会話の継続が終了した後に、再びユーザ２３０がキャラクター１９０５１に話しかけて新たに会話を行う場合の一例を示している。ここで、図３Ｆでは、実施例２の図２Ｆで説明した、設定指示文に過去の会話の履歴を説明するメッセージを格納する方法を、実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）にも適用したものである。具体的には、図３Ｄの設定指示文の内容となるメッセージを、図３Ｆでは、再設定メッセージとして格納し、再設定メッセージに続いて、過去の会話の履歴を説明するメッセージを会話履歴メッセージとして格納する。

【０１８９】

実施例３の大規模言語モデルサーバ２００００１は、自然言語のテキスト情報と併せて非自然言語情報源を処理できるマルチモーダルな大規模言語モデルであるので、過去の指示文および応答において非自然言語情報源データの伝送または指定がなされている場合がある。よって、図３Ｆの例では、当該会話履歴メッセージに、過去の指示文および応答における自然言語のテキスト情報のみならず、過去の指示文および応答における非自然言語情報源データの伝送または指定を反映する。図３Ｆの指示文における、非自然言語情報源データの伝送または指定の具体的な方法は、図３Ｄで説明したように非自然言語情報源データの伝送または指定と同様であるので、繰り返しの説明は省略する。

【０１９０】

図３Ｄの例では、非自然言語情報源データの伝送または指定の方法には、指示文に非自然言語情報源データそのものを格納（添付）する場合と、指示文に非自然言語情報源データを格納（添付）しない場合がある。この点について、図３Ｆの指示文でも同様である。

【０１９１】
次に、図３Ｇを用いて、本発明の実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ２００００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図３Ｇは、人工知能応答出力装置１００１０に表示されるキャラクター１９０５１とユーザ２３０の会話の元となる、人工知能応答出力装置１００１０から大規模言語モデルサーバ２００００１へ送信される指示文の主たるメッセージとその応答となるサーバ応答の主たるメッセージの一例である。
【０１９２】
図３Ｇは、図３Ｆに示した一連の会話における、最初の設定指示文以降に続くユーザ指示文１巡目とその応答からユーザ指示文３巡目とその応答までの一連の会話の一例を示している。図３Ｇでは、時系列に指示文と応答のやり取りがなされているものとして示している。設定指示文の内容は図３Ｆに示したとおりなので、繰り返しの記載は省略する。
【０１９３】
以上説明したように、実施例３の自然言語のテキスト情報と併せて非自然言語情報源を処理できるマルチモーダルな大規模言語モデルを有する大規模言語モデルサーバ２０００１を用いる場合でも、一連の会話後に当該一連の会話の継続が終了した後に、再びユーザ２３０がキャラクター１９０５１に話しかけて新たに会話を行う場合であっても、図３Ｆの設定指示文の生成処理、および送信処理を行えば、その後のユーザ指示文の応答は、図３Ｇに示すように、以前の会話の時点のキャラクターの役割、名前、会話の特徴、性格、および／または会話の特徴などの設定や会話履歴を反映したものとなる。これにより、ユーザからすると、以前の会話の時点の当該キャラクターの役割、名前、会話の特徴または性格などの設定や記憶の同一性がより確保できているように認識されるため、より好適である。
【０１９４】
次に、図３Ｈを用いて、本発明の実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）の動作の一例について説明する。これは、人工知能応答出力装置１００１０と大規模言語モデルサーバ２００００１とを含むキャラクター会話システムの動作の一例の説明ともいえる。具体的には、図３Ｈは、キャラクター会話装置（人工知能応答出力装置１００１０）の表示部１００１１に表示する複数のキャラクターについての、キャラクターの設定とキャラクターの会話履歴を管理するためのデータベース２０２００の説明図である。ここで、図３Ｈは、キャラクター会話装置（人工知能応答出力装置１００１０）の表示部１００１１に表示する複数のキャラクターの設定等は、実施例２の図２Ｈで説明した例を用いる。よって複数のキャラクターの設定等についての繰り返しの説明は省略する。
【０１９５】
また、図３Ｈに示す、キャラクターの設定とキャラクターの会話履歴を管理するためのデータベース２０２００は、実施例２の図２Ｉに示すデータベース１９２００と同様のフォーマットを有するものであり、図３Ｈにおいては、図２Ｉに示すデータベース１９２００との差分についてのみ説明する。また、データベース内のキャラクター「Ｋｏｔｏ」についての内容を説明し、他のキャラクターの内容については省略する。
【０１９６】
ここで、上述のとおり、実施例３の大規模言語モデルサーバ２００００１は、自然言語のテキスト情報と併せて非自然言語情報源を処理できるマルチモーダルな大規模言語モデルであるので、キャラクター会話装置（人工知能応答出力装置１００１０）からの指示文にも、大規模言語モデルサーバ２００００１からの応答にも、自然言語のテキスト情報のみならず、非自然言語情報源データの伝送または指定が含まれる。よって、図３Ｈに示す、データベース２０２００では、会話履歴のデータにおいて、これらの指示文や応答に含まれる、自然言語のテキスト情報のみならず、非自然言語情報源データの伝送または指定の情報も記録する。当該会話履歴の記録における、非自然言語情報源データの伝送または指定

の具体的な方法は、図３Ｄで説明した非自然言語情報源データの伝送または指定を指定と同様であるので、繰り返しの説明は省略する。

【０１９７】

図３Ｄの例では、非自然言語情報源データの伝送または指定の方法には、指示文に非自然言語情報源データそのものを格納（添付）する場合と、指示文に非自然言語情報源データを格納（添付）しない場合がある。この点について、図３Ｈの会話履歴でも同様である。ただし、図３Ｈの会話履歴において、非自然言語情報源データの指定方法として、インターネット等のネットワーク上に存在するサーバ（中間サーバとして機能させる第２のサーバ１９００２やその他のクラウドサーバ）の非自然言語情報源ファイルのロケーション情報およびファイル名情報を指定する場合、会話履歴の期間が長くなると、当該サーバ上の当該非自然言語情報源ファイルが削除されてしまう可能性もある。すると、当該ロケーション情報およびファイル名情報を用いても当該非自然言語情報源ファイルを後日取得することができなくなり会話記録の情報が欠落する可能性がある。

【０１９８】

これを防ぐためには、キャラクター会話装置（人工知能応答出力装置１００１０）は、指示文および応答のメッセージを会話履歴に変換して記録する際に、指示文および応答において指定される非自然言語情報源ファイル自体を、当該ロケーション情報およびファイル名情報を用いて、ネットワーク上のサーバ等から取得し、ストレージ部１１７０に格納すればよい。さらに、当該非自然言語情報源ファイルのロケーション情報およびファイル名を、キャラクター会話装置（人工知能応答出力装置１００１０）のキャラクター動作プログラムが、キャラクター会話装置（人工知能応答出力装置１００１０）内に構築するメディアサーバのメディアサーバ内を示すインターネット上でのロケーション情報（いわゆるＵＲＬなど）に書き換えてから、会話記録に記録すればよい。このようにすれば、キャラクター会話装置（人工知能応答出力装置１００１０）自身がストレージ部１１７０から当該非自然言語情報源ファイルを消去しなければ、当該非自然言語情報源は会話記録の情報から欠落しないため、会話記録の保全としてより好適である。

【０１９９】

以上説明した図３Ｈのデータベースを用いれば、キャラクター会話装置（人工知能応答出力装置１００１０）が、表示部１００１１に表示するキャラクターを複数のキャラクター候補の中から切り替えて表示するように構成する場合も、ユーザからすると、それぞれのキャラクターとの会話から感じる違和感が少なく、また、複数のキャラクターのそれぞれと記憶を共有することができ、より楽しいキャラクター会話体験を得られる、という実施例２の図２Ｉの効果を得ることができ、かつ、大規模言語モデルサーバ２０００１が自然言語のテキスト情報と併せて非自然言語情報源を処理できるマルチモーダルな大規模言語モデルである場合にも、この効果を発揮することができる。

【０２００】

なお、実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）またはキャラクター会話システムにおいては、大規模言語モデルサーバ２０００１において、自然言語のテキスト情報に加えて自然言語のテキスト情報以外の非自然言語情報も併せて処理が可能なマルチモーダルな大規模言語モデル人工知能を用いる。

【０２０１】

ここで、キャラクター会話装置（人工知能応答出力装置１００１０）と大規模言語モデルサーバ２０００１との間はＡＰＩを用いて通信される。マルチモーダルな大規模言語モデルでは、自然言語のテキスト情報をトークンと呼ばれる文章を区切った言葉の単位の処理数に加えて、非自然言語情報源のデータ量に応じて、ＡＰＩの利用料金が請求される形式もあり得る。

【０２０２】

そこで、本実施例に係るキャラクター会話システムによるキャラクター会話サービスを、より安価にユーザに提供するために、以下のような変形例を用いればよい。

【０２０３】

第１の変形例としては、図３Ｈのデータベースの会話履歴の記録において、非自然言語情報源データの伝送または指定の情報も記録されている。しかしながら、キャラクターとユーザは、当該自然言語情報源データについての会話を自然言語のテキスト情報でやり取りしており、その内容は自然言語のテキスト情報で記録されている。すると、図３Ｈのデータベースの会話履歴の記録において、当該自然言語情報源データの伝送または指定の情報の記録を省略しても、当該自然言語情報源データについての会話自体はある程度自然言語のテキスト情報として記録されることとなる。よって、ある程度の情報の減少を許容すれば、図３Ｈのデータベースの会話履歴の記録において、当該自然言語情報源データの伝送または指定の情報の記録を省略してもよい。この場合、図３Ｆの設定指示文の会話履歴メッセージからも当該自然言語情報源データの伝送または指定の情報が省略される。これにより、ＡＰＩを用いて通信される非自然言語情報源のデータ量を低減することができる。

【０２０４】

次に、第２の変形例としては、図３Ｈのデータベースの会話履歴の記録において、非自然言語情報源データの伝送または指定の情報の記録の代わりに、非自然言語情報源データの内容を説明する自然言語のテキスト情報を記録する例である。非自然言語情報源データの内容を説明する自然言語のテキスト情報は、例えば、キャラクターとしての会話とは別に、大規模言語モデルサーバ２０００１の大規模言語モデルとキャラクター会話装置（人工知能応答出力装置１００１０）との会話を立上げ、大規模言語モデルサーバ２０００１に非自然言語情報源データの内容を所定の文字数制限を指定して説明させて取得してもよい。また、大規模言語モデルサーバ２０００１の大規模言語モデルよりも安価に利用できる、他のサーバの他の大規模言語モデルとの会話により、非自然言語情報源データの内容を所定の文字数制限を指定して説明させて取得してもよい。また、非自然言語情報源データの取得時点から、代替テキストデータが準備されている場合は、当該代替テキストデータを非自然言語情報源データの内容を説明する自然言語のテキスト情報としてもよい。非自然言語情報源データの代替テキストデータの具体例としては、マークアップ言語のタグの＜img src =“ ”alt="****"＞、＜video src =“ ”alt="****"＞、＜audio src =“ ”alt="****"＞などの****の部分に記載されるテキスト情報である。

【０２０５】

また、ＪＳＯＮ形式の記法であれば、非自然言語情報源データのロケーション情報を示すキーと値である非自然言語情報源データのロケーション情報およびファイル名情報と対応づけられて格納されているオブジェクトにおいて、さらに、代替テキストに対応するキーと、代替テキストデータ自体である値を紐づけて格納すればよい。

【０２０６】

この場合も、図３Ｈのデータベースの会話履歴の記録において、当該自然言語情報源データの伝送または指定の情報の記録を省略でき、図３Ｆの設定指示文の会話履歴メッセージからも当該自然言語情報源データの伝送または指定の情報が省略される。これにより、ＡＰＩを用いて通信される非自然言語情報源のデータ量を低減することができる。

【０２０７】

次に、第３の変形例としては、図３Ｄのユーザ指示文1巡目の時点で、非自然言語情報源データの伝送または指定の情報をユーザ指示文に格納せず、非自然言語情報源データの内容を説明する自然言語のテキスト情報に差し替えてしまう例である。例えば、図３Ｄのユーザ指示文1巡目において、非自然言語情報源データ２００６１の伝送または指定の情報をユーザ指示文に替えて、「この画像とは、スイミングプールとプールサイドにあるシートとパラソルの画像です。スイミングプールには、水があります。シートの脇のテーブルにはドリンクがあります」との説明文が自然言語のテキスト情報として格納されるように構成すればよい。このとき、当該説明文は、大規模言語モデルサーバ２０００１の大規模言語モデルよりも安価に利用できる、他のサーバの他の大規模言語モデルとの会話により、非自然言語情報源データの内容を所定の文字数制限を指定して説明させて取得してもよい。また、当該説明文は、画像、動画、音声などの非自然言語情報源データの概要や内

容の説明を取得できるその他各種サービスのサーバから取得してもよい。また、非自然言語情報源データの取得時点から、代替テキストデータが準備されている場合は、当該代替テキストデータを非自然言語情報源データの内容を説明する自然言語のテキスト情報としてもよい。

【０２０８】

次に、図３Ｉを用いて、本発明の実施例３のキャラクター会話装置（人工知能応答出力装置１００１０）の表示例の一例について説明する。図３Ｉの例では、図３Ａ～図３Ｈの各図で説明したユーザからの指示文に対する大規模言語モデルからの応答を、キャラクター会話装置（人工知能応答出力装置１００１０）の表示部１００１１に表示する例の一例を示している。具体的には、大規模言語モデルからの応答である自然言語情報源データのテキスト１００６３、非自然言語情報源データの画像１００６４、および／または非自然言語情報源データの動画１００６５を、キャラクター１９０５１の映像とともに、表示部１００１１に表示する例である。大規模言語モデルからの応答であるテキスト１００６３、画像１００６４、および／または動画１００６５は、図３Ｉに示すようにキャラクター１９０５１の映像の手前に重畳して表示してもよい。

【０２０９】

また、大規模言語モデルからの応答であるテキスト１００６３、画像１００６４、および／または動画１００６５は、キャラクター１９０５１の映像と重畳せずに、キャラクター１９０５１の映像ともに表示してもよい。図３Ｉの表示は一例であるが、例えば、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により、キャラクター会話装置（人工知能応答出力装置１００１０）の音声出力部１１４０の音声出力のボリュームを最小に調整するまたは音声出力をＯＦＦに設定した場合などは、ユーザ２３０は音声で大規模言語モデルからの応答を確認することができない。そこで、この場合、制御部１１１０は、大規模言語モデルからの応答であるテキスト１００６３、画像１００６４、および／または動画１００６５を図３Ｉに示すように、キャラクター１９０５１の映像ともに表示する表示モードを開始するように制御してもよい。

【０２１０】

このようにすれば、音声出力を控えたい場合でも、ユーザ２３０がより好適に、キャラクター会話装置（人工知能応答出力装置１００１０）を使用することができる。なお、ユーザ２３０が操作入力部１１０７や表示部１００１１のタッチ操作入力センサを介した操作により、大規模言語モデルからの応答であるテキスト１００６３、画像１００６４、および／または動画１００６５をキャラクター１９０５１の映像ともに表示する表示モードについてのＯＮ／ＯＦＦを手動で切り替え可能に構成してもよい。図３Ｉの表示例によれば、マルチモーダルに対応したキャラクター会話装置（人工知能応答出力装置１００１０）において、より好適に大規模言語モデルからの応答を出力することが可能となる。

【０２１１】

以上説明した、実施例３に係るキャラクター会話装置や、キャラクター会話システムによれば、実施例２に係るキャラクター会話装置や、キャラクター会話システムにおける効果に加えて、マルチモーダルな大規模言語モデルを用いて、自然言語の情報以外に非自然言語の情報を含めたより高度な会話体験をユーザに提供することができる。また、実施例３に係るキャラクター会話装置や、キャラクター会話システムによれば、キャラクター会話サービスをより安価にユーザに提供することができる。

【０２１２】

なお、実施例３の以上の説明では、大規模言語モデルとして、大規模言語モデルサーバ２０００１が有する大規模言語モデルを用いる例を説明した。これに対し、キャラクター会話装置（人工知能応答出力装置１００１０）が、図１Ｂに示すローカルＬＬＭ処理部１００２８を備え、当該ローカルＬＬＭ処理部１００２８が有するマルチモーダルな大規模言語モデルを用いてもよい。この場合、大規模言語モデルサーバ２０００１が有するマルチモーダルな大規模言語モデルに変えて、ローカルＬＬＭ処理部１００２８が有するマルチモーダルな大規模言語モデルを用いてもよい。

【０２１３】
この場合、実施例３の以上の説明において、大規模言語モデルサーバ２０００１が有するマルチモーダルな大規模言語モデルをキャラクター会話装置（人工知能応答出力装置１００１０）のローカルＬＬＭ処理部１００２８が有するマルチモーダルな大規模言語モデルに読み替えればよい。この場合も、マルチモーダルな大規模言語モデルを用いて、自然言語の情報以外に非自然言語の情報を含めたより高度な会話体験をユーザに提供することができる。なお、大規模言語モデルサーバ２０００１が有するマルチモーダルな大規模言語モデルに変えて、ローカルＬＬＭ処理部１００２８が有するマルチモーダルな大規模言語モデルを用いる場合は、処理トークン数や非自然言語情報源のデータ量に応じた利用料金を考慮する必要性は少なくなるが、ローカルＬＬＭ処理部１００２８が有するマルチモーダルな大規模言語モデルであっても処理トークン数や非自然言語情報源のデータ量の低減を図ることによって、推論にかかる電力などの消費リソースを低減できる。この場合、より消費電力の少ないキャラクター会話サービスをユーザに提供することができる。
【０２１４】
なお、実施例２で説明した、キャラクターとの会話履歴や、キャラクターとの会話履歴を含むデータベースのデータを第２のサーバ１９００２またはその他のクラウドサーバに対してアップロードおよびダウンロードする構成は、実施例３で説明した、マルチモーダルな大規模言語モデルを用いる例においても利用できる。この場合も、１つのキャラクターまたは複数のキャラクターについて、異なる個体の装置間において、複数のキャラクターのそれぞれのキャラクターとユーザが異なるタイミングで複数回にわたって会話を行う場合に、前回の会話からそれぞれのキャラクターの記憶が疑似的に引き継がれたかのような会話を実現することができ、ユーザにとってより好適である。
【０２１５】
＜実施例４＞
次に、本発明の実施例４は、実施例２、または実施例３の各図で説明した人工知能応答出力装置１００１０、キャラクター会話装置、またはこれらのシステムを改良したものである。本実施例では、実施例２、または実施例３との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。
【０２１６】
上述の実施例同様、人工知能応答出力装置１００１０は、人工知能応答出力装置、ＡＩアシスタント装置、ＡＩアシスタント表示装置、人工知能インタフェース装置と称してもよい。人工知能応答出力装置１００１０と大規模言語モデルサーバを含むシステムは、人工知能応答出力システム、ＡＩアシスタントシステム、ＡＩアシスタント表示システム、人工知能インタフェースシステムと称してもよい。
【０２１７】
図４Ａを用いて、本発明の実施例４のキャラクター会話装置（人工知能応答出力装置１００１０）におけるデータベースを用いた動作の一例について説明する。図４Ａに示す実施例４に係るデータベースは、図２Ｉまたは図３Ｉで説明したデータベースを拡張したものである。図４Ａに示すデータベースは、具体的には、複数の異なるユーザが、同一のキャラクター会話装置（人工知能応答出力装置１００１０）または同一のキャラクター会話システムを利用する場合を想定し、それぞれのユーザとキャラクタに対応する初期設定指示文および会話履歴をデータベースに格納する。
【０２１８】
図４Ａの例では、ユーザＩＤが1であるユーザ１について、キャラクターＩＤが１であるキャラクターＫｏｔｏ、キャラクターＩＤが２であるキャラクターＴｏｍ、およびキャラクターＩＤが３であるキャラクターＮｅｃｃｏ、のそれぞれのキャラクターの初期設定指示文および会話履歴を格納している。これに加えて、ユーザＩＤが２であるユーザ２、ユーザＩＤが３であるユーザ３のそれぞれについても、キャラクターＩＤが１であるキャラクターＫｏｔｏ、キャラクターＩＤが２であるキャラクターＴｏｍ、およびキャラクターＩＤが３であるキャラクターＮｅｃｃｏ、のそれぞれのキャラクターの初期設定指示文

および会話履歴を格納している。

【０２１９】

これらの初期設定指示文および会話履歴のデータは、ユーザとキャラクターの組み合わせごと、に異なる領域に別のデータとして格納されている。図４Ａでは説明のため、それぞれの領域に格納しているデータをデータ１１、１２、１３、２１、２２、２３、３１、３２、３３と表記して示している。キャラクター会話装置（人工知能応答出力装置１００１０）の制御部１１１０は、現在キャラクター会話装置（人工知能応答出力装置１００１０）またはそのシステムを使用（ログイン）しているユーザに基づき、ユーザとキャラクターの組み合わせごとに異なる領域に格納された初期設定指示文および会話履歴を用いることで、キャラクターの個性の一貫性や記憶の継続性を、異なるユーザのそれぞれに対してより好適に維持することが可能となる。

【０２２０】

具体的には、キャラクター会話装置（人工知能応答出力装置１００１０）を用いて先にユーザ１がキャラクターＴｏｍと会話しており、その会話をユーザ２が知らなかった場合に、その後にユーザ２がキャラクターＴｏｍと会話する状況について考える。このとき、人工知能応答出力装置１００１０がユーザの識別をしない初期設定指示文または会話履歴のデータベースを用いている場合、人工知能応答出力装置１００１０から出力される応答がユーザ２の記憶にない会話の履歴に基づくものとなり、ユーザ２と人工知能応答出力装置１００１０のキャラクターとの会話が整合しなくなる可能性がある。

【０２２１】

これに対し、同様の状況であっても、図４Ａに示すデータベースを用いれば、キャラクター会話装置（人工知能応答出力装置１００１０）の制御部１１１０はユーザをＩＤにより識別して、ユーザごとに異なる領域に初期設定指示文および会話履歴を格納し、ユーザごとに異なる領域に格納された初期設定指示文および会話履歴を人工知能応答の生成に使用する。これにより、それぞれのユーザに対する人工知能応答の生成に使用する初期設定指示文および会話履歴は当該ユーザの操作または会話の経緯に基づくものとなり、他のユーザの操作または会話の経緯とは区別して管理される。これにより、それぞれのユーザと人工知能応答出力装置１００１０のそれぞれのキャラクターとの会話の履歴の整合性をより好適にすることができる。

【０２２２】

なお、図４Ａで説明した、初期設定指示文および／または会話履歴のデータベースは、人工知能応答出力装置１００１０のストレージ部１１７０に格納して、制御部１１１０が用いればよい。また、これに限られず、初期設定指示文および／または会話履歴のデータベースは、ネットワーク上のサーバに格納してもよい。例えば、人工知能応答出力装置１００１０が、人工知能応答の生成において、大規模言語モデルサーバ１９００１の大規模言語モデルまたは大規模言語モデルサーバ２０００１のマルチモーダルな大規模言語モデルを用いる場合はこれらのサーバ自体に、図４Ａで説明した、初期設定指示文および／または会話履歴のデータベースを格納すればよい。このようにすれば、人工知能応答出力装置１００１０からこれらのサーバに対して、初期設定指示文および会話履歴を再度指示文に含めて送信する処理を省略することができ、大規模言語モデルの利用についての送信トークン数を節約することができる。

【０２２３】

図４Ａで説明した、初期設定指示文および／または会話履歴のデータベースを格納する場合は、人工知能応答出力装置１００１０からこれらのサーバへ、ユーザＩＤ、キャラクターＩＤ、および、その後の会話のためのユーザ指示文を送信すればよい。これらのサーバ上の大規模言語モデルは、人工知能応答出力装置１００１０から取得したユーザＩＤ、キャラクターＩＤを用いて、図４Ａの初期設定指示文および／または会話履歴のデータベースから、該当する初期設定指示文および会話履歴を取得する。これらのサーバ上の大規模言語モデルは、当該初期設定指示文および会話履歴と、人工知能応答出力装置１００１０から送信されたその後の会話のためのユーザ指示文を用いて、推論を実行し、人工知能

応答を生成して、人工知能応答出力装置１００１０へ送信すればよい。このようにすれば、キャラクターの個性の一貫性や記憶の継続性を、異なるユーザのそれぞれに対してより好適に維持する効果を、大規模言語モデルの利用についての送信トークン数を節約しながら得ることができる。

【０２２４】

次に、図４Ｂを用いて、本発明の実施例４のキャラクター会話装置（人工知能応答出力装置１００１０）におけるデータベースを用いた動作の一例について説明する。図４Ｂに示す実施例４に係るデータベースは、図１Ｃまたは図２Ｌで説明したデータベースを拡張したものである。図４Ｂに示すデータベースは、具体的には、複数の異なるユーザが、同一のキャラクター会話装置（人工知能応答出力装置１００１０）または同一のキャラクター会話システムを利用する場合を想定し、それぞれのユーザとキャラクターに対応する応答定型文のデータをデータベースに格納する。

【０２２５】

図４Ｂの例では、ユーザＩＤが1であるユーザ１について、キャラクターＩＤが１であるキャラクターＫｏｔｏ、キャラクターＩＤが２であるキャラクターＴｏｍ、およびキャラクターＩＤが３であるキャラクターＮｅｃｃｏ、のそれぞれのキャラクターの応答定型文データを格納している。これに加えて、ユーザＩＤが２であるユーザ２、ユーザＩＤが３であるユーザ３のそれぞれについても、キャラクターＩＤが１であるキャラクターＫｏｔｏ、キャラクターＩＤが２であるキャラクターＴｏｍ、およびキャラクターＩＤが３であるキャラクターＮｅｃｃｏ、のそれぞれのキャラクターの応答定型文データを格納している。

【０２２６】

これらの応答定型文のデータは、ユーザとキャラクターの組み合わせごと、に異なる領域に別のデータとして格納されている。図４Ｂでは説明のため、それぞれの領域に格納しているデータを応答定型文データ１０１、１０２、１０３、２０１、２０２、２０３、３０１、３０２、３０３と表記している。例えば、応答定型文データ１０１には、図２Ｌに示すキャラクター１：Ｋｏｔｏの条件番号１～７に対応する応答定型文に対応するテーブルなどのデータベースとして格納されている。図４Ｂのデータ２０１には、図２Ｌに示すキャラクター２：Ｔｏｍの条件番号１～７に対応する応答定型文に対応するテーブルなどのデータベースとして格納されている。

【０２２７】

図４Ｂのデータ３０１には、図２Ｌに示すキャラクター３：Ｎｅｃｃｏの条件番号１～７に対応する応答定型文に対応するテーブルなどのデータベースとして格納されている。図４Ｂのデータ１０２、２０２、３０２には、同様のフォーマットで、ユーザ２向けに応答定型文が変更されたものが格納されている。図４Ｂのデータ１０３、２０３、３０３には、同様のフォーマットで、ユーザ３向けに応答定型文が変更されたものが格納されている。キャラクター会話装置（人工知能応答出力装置１００１０）の制御部１１１０は、現在キャラクター会話装置（人工知能応答出力装置１００１０）またはそのシステムを使用（ログイン）しているユーザに基づき、ユーザとキャラクターの組み合わせごとに異なる領域に格納された応答定型文データを用いる。

【０２２８】

このようにすれば、同一のキャラクターであっても、ユーザごとに異なる応答定型文による応答を行うことが可能となる。すなわち、同一のキャラクターであっても、キャラクターとユーザの関係性によっては、応答定型文の内容を変化させた方が好適であることもあり得る。例えば、キャラクターの設定の年齢と、人工知能応答出力装置１００１０またはシステムに登録されたユーザの年齢との関係により、ユーザがキャラクターに対して年上の場合もあれば、同い年の場合もあれば、年下の場合もある。このとき、年上のユーザに対するキャラクターの応答定型文と、同い年のユーザに対するキャラクターの応答定型文と、年下のユーザに対するキャラクターの応答定型文とで、それぞれ内容を変えた方が、ユーザとキャラクターの会話がより好適または、より自然となる。すなわち、図４Ｂの

データベースを用いた動作を行うことにより、キャラクターとユーザの関係性ごとに、応答定型文の内容を異ならせることにより、より好適または、より自然な会話を演出することが可能となる。

【０２２９】

なお、以上説明した、図４Ｂの応答定型文データベース（応答定型文ＤＢ）はストレージ部１１７０に格納され、人工知能応答出力装置１００１０の制御部１１１０がこれを用いればよい。しかしながら、図４Ｂに示す応答定型文データベース（応答定型文ＤＢ）を大規模言語モデルサーバ１９００１側または大規模言語モデルサーバ２０００１側に備えてもよい。この場合は、大規模言語モデルサーバ１９００１が有する制御部または大規模言語モデルサーバ２０００１が有する制御部が、当該応答定型文データベース（応答定型文ＤＢ）を用いた応答を生成すればよい。大規模言語モデルサーバ１９００１が有する制御部または大規模言語モデルサーバ２０００１が有する制御部は、それぞれのサーバに格納される大規模言語モデルにより生成する応答に替えて、応答定型文データベース（応答定型文ＤＢ）を用いて生成した応答を、人工知能応答出力装置１００１０へ送信すればよい。このようにすれば、人工知能応答出力装置１００１０に応答定型文データベース（応答定型文ＤＢ）が備えられていない場合でも、応答定型文データベース（応答定型文ＤＢ）を用いた応答の生成が可能となる。

【０２３０】

以上説明した、実施例４に係るキャラクター会話装置や、キャラクター会話システムによれば、キャラクターとユーザの関係性や会話履歴などに応じて、より好適または、より自然な会話を演出することが可能となる。

【０２３１】

＜実施例５＞

次に、本発明の実施例５は、実施例１、実施例２、実施例３の各図で説明した人工知能応答出力装置１００１０、または人工知能応答出力システムを改良したものである。具体的には、人工知能応答出力装置１００１０の応答生成処理を、ネットワーク上の大規模言語モデルによる応答生成処理から、人工知能応答出力装置１００１０が備えるローカルの大規模言語モデル（ローカルＬＬＭ処理部１００２８など）による応答生成処理、または応答定型文データベースによる応答生成処理へ切り替える処理を行う例である。本実施例では、これらの実施例との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。

【０２３２】

上述の実施例同様、人工知能応答出力装置１００１０は、人工知能応答出力装置、キャラクター会話装置、ＡＩアシスタント装置、ＡＩアシスタント表示装置、人工知能インタフェース装置と称してもよい。人工知能応答出力装置１００１０と大規模言語モデルサーバを含むシステムは、人工知能応答出力システム、キャラクター会話システム、ＡＩアシスタントシステム、ＡＩアシスタント表示システム、人工知能インタフェースシステムと称してもよい。

【０２３３】

図５Ａを用いて、本発明の実施例５の人工知能応答出力装置１００１における応答生成処理の切替処理の一例について説明する。図５Ａの表には、人工知能応答出力装置１００１における応答生成処理の切替処理の例について、例１～例９まで示している。図５Ａの表において、「切替概要」の列には、各例の切替処理の概要を示している。「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）切替前状態」の列には、図１の大規模言語モデルサーバ１９００１が備える大規模言語モデル、および大規模言語モデルサーバ２０００１が備えるマルチモーダルな大規模言語モデルなどのネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が他の応答生成処理へ切替えられる前の状態を示している。「切替発生条件」の列には、応答生成処理の切替処理が発生する条件を示している。「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」の列には、人工知能応答出力装置１００１０の応答生成処理を、大規模言語モデルサー

バ１９００１が備える大規模言語モデル、および大規模言語モデルサーバ２００００１が備えるマルチモーダルな大規模言語モデルなどのネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）から切替える、切替先が示されている。人工知能応答出力装置１００１の制御部１１１０は、図５Ａに示される「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）切替前状態」の状態において「切替発生条件」に示される条件が生じた場合に、「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」に示される大規模言語モデル、データベース、または対応に切り替えるように制御を行えばよい。

【０２３４】

以下、図５Ａの表に示される各例について説明する。例１は、「切替概要」に示すとおり、人工知能応答出力装置１００１のネットワークの接続可否状態に応じて切替を行う例である。例１では、「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）切替前状態」としては、人工知能応答出力装置１００１のネットワークの接続状態が接続可能状態であることが示されている。ここで、例１では、「切替発生条件」としては「ネットワーク接続不能になった場合」が示されている。すなわち、これは、人工知能応答出力装置１００１とネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）とのネットワークを介した接続が不能となった場合、という意味である。具体的には、当該接続不能は、人工知能応答出力装置１００１からインターネット１９０００への接続経路での通信不能状態に起因する場合もあり得る。または、当該接続不能は、インターネット１９０００における通信不能状態に起因する場合もあり得る。または、当該接続不能は、ネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）自体がインターネット１９０００に接続できない状況に起因する場合もあり得る。また、例１では、「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」としては「ローカルＬＬＭ」が示されている。これは、具体的には、人工知能応答出力装置１００１０が有するローカルＬＬＭ処理部１００２８による応答生成処理への切替処理を行うことを意味する。すなわち、例１では、なんらかの理由でネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）との接続が不能となって、ネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が利用できない場合であっても、人工知能応答出力装置１００１０が有するローカルＬＬＭ処理部１００２８による応答生成処理に切り替える。これにより、大規模言語モデルとしての性能差はあれど、大規模言語モデルを用いた応答生成処理を継続することは可能である。

【０２３５】

次に、図５Ａの例２について説明する。例２は、例１における「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」を「ローカルＬＬＭ」から「応答定型文ＤＢ（データベース）」に変更したものである。当該「応答定型文ＤＢ（データベース）」による応答生成処理は、図１Ｃ、図２Ｌ、または図４Ｂで説明した処理と同様であるので繰り返しの説明は省略する。すなわち、例２では、なんらかの理由でネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）との接続が不能となって、ネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が利用できない場合には、応答定型文データベースを用いた応答生成処理に切り替えることにより、より簡易な処理によって応答を生成して当該応答をユーザに出力することが可能となる。

【０２３６】

次に、図５Ａの例３について説明する。例３は、例１における「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」を「ローカルＬＬＭ」から「不応答対応」に変更したものである。当該「不応答対応」とは、タッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザから大規模言語モデルによる応答を求めるユーザ入力があった場合でも、この入力に対する応答を生成しない、または、大規模言語モデルによる応答を求めるユーザ入力があった場合でも、これに対する応答を出力しない対応を意味する。すなわち、例３では、なんらかの理由でネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）との接続が不能となって、ネットワーク上の大規模言

語モデル（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が利用できない場合の対応をより簡易にすることが可能となる。

【０２３７】

次に、図５Ａの例４について説明する。例４は、「切替概要」に示すとおり、ネットワーク上のＬＬＭの応答遅延により切替を行う例である。例４では、「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）切替前状態」としては、ネットワーク上のＬＬＭからの応答が所定時間以内に得られる状態が示されている。ここで、例４では、「切替発生条件」としてはネットワーク上のＬＬＭからの応答が所定時間以内には得られず、所定時間を超過した場合が示されている。また、例４では、「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」としては「ローカルＬＬＭ」が示されている。切替先の「ローカルＬＬＭ」は、例１と同様であるので繰り返しの説明を省略する。すなわち、例４では、なんらかの理由でネットワーク上のＬＬＭ（ＡＰＩを用いて接続する大規模言語モデル）からの応答が所定時間を超過するようになり、ネットワーク上のＬＬＭ（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が順調に利用できない場合であっても、人工知能応答出力装置１００１０が有するローカルＬＬＭ処理部１００２８による応答生成処理に切り替える。これにより、大規模言語モデルとしての性能差はあれど、大規模言語モデルを用いた応答生成処理を継続することは可能である。

【０２３８】

次に、図５Ａの例５について説明する。例２は、例４における「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」を「ローカルＬＬＭ」から「応答定型文ＤＢ（データベース）」に変更したものである。当該「応答定型文ＤＢ（データベース）」による応答生成処理は、図１Ｃ、図２Ｌ、または図４Ｂで説明した処理と同様であるので繰り返しの説明は省略する。すなわち、例５では、なんらかの理由でネットワーク上のＬＬＭ（ＡＰＩを用いて接続する大規模言語モデル）からの応答が所定時間を超過するようになり、ネットワーク上のＬＬＭ（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が順調に利用できない場合には、応答定型文データベースを用いた応答生成処理に切り替えることにより、より簡易な処理によって応答を生成して当該応答をユーザに出力することが可能となる。

【０２３９】

次に、図５Ａの例６～例９について説明する。例６～例９は、「切替概要」に示すとおり、ＡＰＩ利用量または利用料金の上限到達により切替を行う例である。ここで、実施例２において説明したとおり、大規模言語モデルの提供元は、大規模言語モデルの学習に用いられたコストを端末のＡＰＩの利用料金として、端末の利用者から回収することが多い。そのとき、自然言語モデルでは、トークンと呼ばれる文章を区切った言葉の単位の処理数でＡＰＩの利用料金が請求される形式が多い。ここで、ＡＰＩ利用料金には様々な請求方法や制限方法が考えられる。その一案として、ユーザが大規模言語モデルの利用サービスの提供を通常の状態で受けられる量の上限をトークン処理数を用いて規定する例が考えられる。

【０２４０】

この場合、ユーザは、当該利用量（もしくは対応する利用料金）に達するまでは所定のＡＰＩ利用料金で大規模言語モデルの利用するサービスを受けることができ、利用量（もしくは対応する利用料金）の上限に達すると、大規模言語モデルの利用サービスの提供を通常の状態（性能または頻度）で受けられなくなるなどの一定の制限が発生する、などが考えられる。

【０２４１】

図５Ａの例６～例９は、大規模言語モデルの利用サービスに、このような制限が発生した場合の、人工知能応答出力装置１００１の制御部１１１０による応答生成処理の切替え制御の例である。具体的には、例６では、「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）切替前状態」は、ＡＰＩの利用量やＡＰＩの利用料金が所定の上限未達の状態である。これは、ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）の利用量が所定の上限に達してい

ない状態を意味する。このとき、ユーザは、ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）を通常の状態で利用することができる。

【０２４２】

ここで、例６では、「切替発生条件」としてはＡＰＩの利用量やＡＰＩの利用料金が所定の上限に達した場合が示されている。これは、ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）の利用量が所定の上限に達した場合を意味する。また、例６では、「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」としては、通常状態で利用していたＬＬＭ（第１のＬＬＭと称してもよい）と異なるネットワーク上の第２のＬＬＭが示されている。ネットワーク上の第２のＬＬＭの例としては、通常状態で利用していた第１のＬＬＭよりも、低料金であるＬＬＭなどがあげられる。より低料金なサービスとなるので、第２のＬＬＭの性能は、第１のＬＬＭの性能よりも低性能であることが考えられる。この場合であっても、第１のＬＬＭの利用量／利用料金の上限に達したあとでも安価に大規模言語モデルが利用可能となるのであれば十分利点がある。

【０２４３】

次に、図５Ａの例７について説明する。例７は、例６における「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」を、通常状態で利用していたＬＬＭ（第１のＬＬＭと称してもよい）と異なるネットワーク上の第２のＬＬＭから「ローカルＬＬＭ」に変更したものである。例７では、ＡＰＩの利用量やＡＰＩの利用料金が所定の上限に達した場合であっても、すなわち、ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）の利用量が所定の上限に達した場合であっても、ネットワーク上のＬＬＭの利用量、ＡＰＩの利用量、またはＡＰＩの利用料金、などによる制限を受けない、ローカルＬＬＭを用いた応答生成処理に切り替えることにより、引き続き大規模言語モデルを用いた応答生成処理を行うことが可能となる。

【０２４４】

次に、図５Ａの例８について説明する。例８は、例７における「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」を「ローカルＬＬＭ」から「応答定型文ＤＢ（データベース）」に変更したものである。当該「応答定型文ＤＢ（データベース）」による応答生成処理は、図１Ｃ、図２Ｌ、または図４Ｂで説明した処理と同様であるので繰り返しの説明は省略する。例８では、ＡＰＩの利用量やＡＰＩの利用料金が所定の上限に達した場合であっても、すなわち、ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）の利用量が所定の上限に達した場合であっても、ネットワーク上のＬＬＭの利用量、ＡＰＩの利用量、またはＡＰＩの利用料金、などによる制限を受けない、応答定型文データベースを用いた応答生成処理に切り替える。これにより、より簡易な処理によって応答を生成して当該応答をユーザに出力することが可能となる。

【０２４５】

次に、図５Ａの例９について説明する。例９は、例７における「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」を「ローカルＬＬＭ」から「不応答対応」に変更したものである。当該「不応答対応」とは、ユーザに対する応答を生成しない、または、ユーザに対する応答を出力しない対応を意味する。例９では、ＡＰＩの利用量やＡＰＩの利用料金が所定の上限に達したこと、すなわち、ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）の利用量が所定の上限に達したことにより、ネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が利用できない場合の対応をより簡易にすることが可能となる。

【０２４６】

以上説明した、図５Ａの例１～例９に示す、人工知能応答出力装置１００１０の応答生成処理の切り替え制御によれば、ネットワーク上のＬＬＭ（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理が通常どおり利用できない状況であっても、それぞれの状況に応じたより好適な切替え、または対応を行うことができる。

【０２４７】

なお、図５Ａの例１～例９の切替え制御は、複数の例を組み合わせて制御を行ってもよ

い。例えば、例１～例３の切替え制御は、それぞれ例４～例９のいずれかの制御と組み合わせてもよい。同様に、例４または例５の制御は、それぞれ例１～例３、または例６～例９のいずれかの制御と組み合わせてもよい。同様に、例６～例９の制御は、それぞれ例１～例５のいずれかの制御と組み合わせてもよい。

【０２４８】

次に、図５Ｂ～図５Ｄを用いて、実施例５の人工知能応答出力装置１００１０をＡＩアシスタント装置またはキャラクター会話装置として構成する場合の、ＡＩアシスタントまたはキャラクターの表示例の一例について、説明する。

【０２４９】

まず、図５Ｂは、図５Ａの例３の切替え制御を行う場合の、人工知能応答出力装置１００１０における、ＡＩアシスタントまたはキャラクターの表示例である。図５Ｂの例では、人工知能応答出力装置１００１０のネットワーク接続状態がネットワーク接続可能であるか、ネットワーク不能であるかに応じて、ＡＩアシスタントまたはキャラクターの表示状態を変更する。人工知能応答出力装置１００１０が、ネットワーク接続可能である状態およびネットワーク不能である状態は図５Ａで説明したとおりであるので、繰り返しの説明は行わない。

【０２５０】

図５Ｂの例では、人工知能応答出力装置１００１０は、（１）ネットワーク接続可能である場合はＡＩアシスタントまたはキャラクターを起きている通常状態で表示するが、（２）ネットワーク接続不能である場合はＡＩアシスタントまたはキャラクターを「眠っている」状態で表示する。図５Ａの例３の切替え制御では、人工知能応答出力装置１００１０がネットワーク接続不能の場合は、ユーザから指示文の入力があっても応答を生成しない、または、応答を出力しない。このとき、人工知能応答出力装置１００１０が表示するＡＩアシスタントまたはキャラクターが起きている通常状態であるとユーザは違和感を感じるが、人工知能応答出力装置１００１０が表示するＡＩアシスタントまたはキャラクターを眠っている状態で表示すれば、ユーザは「当該ＡＩアシスタントまたはキャラクターが応答しないのは、眠っているからである。」と理解することができ、ユーザが感じる違和感をより低減することが可能となる。

【０２５１】

なお、図５Ｂ（２）の場合、人工知能応答出力装置１００１のタッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザから大規模言語モデルによる応答を求めるユーザ入力を行う前に、ユーザが「当該ＡＩアシスタントまたはキャラクターが応答しないのは、眠っているからである。」と理解することが望ましい。よって、図５Ｂの（２）ネットワーク接続不能である場合はＡＩアシスタントまたはキャラクターを「眠っている」状態で表示する状態の開始タイミングは、ユーザから大規模言語モデルによる応答を求めるユーザ入力より前の、人工知能応答出力装置１００１の制御部１１１０がネットワーク接続不能であると判断した時点の直後であることが望ましい。

【０２５２】

次に、別の表示例として、図５Ｃの表示例について説明する。図５Ｃの表示例は、図５Ａの切替え制御において、表中の「ネットワーク上のＬＬＭ（ＡＰＩ接続ＬＬＭ）からの切替先」の状態に応じて、ＡＩアシスタントまたはキャラクターの表示状態を変更する例である。具体的には、図５Ｃには、（１）人工知能応答出力装置１００１０がネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）との接続が可能であり、かつ、当該ネットワーク上の大規模言語モデルによる応答生成処理を利用可能な状態（本図における通常状態と称する）の場合のＡＩアシスタントまたはキャラクターの表示例と、（２）人工知能応答出力装置１００１０がネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）よりも低性能なＬＬＭまたは応答定型文データベースによる応答生成処理へ切替えた状態におけるＡＩアシスタントまたはキャラクターの表示例と、（３）人工知能応答出力装置１００１０が図５Ａで説明した不応答対応へ切替えた状態におけるＡＩアシスタントまたはキャラクターの表示例が示されている。

【０２５３】
図５Ｃの例では、例えば、（１）人工知能応答出力装置１００１０が「通常状態」である場合は、人工知能応答出力装置１００１０は、ＡＩアシスタントまたはキャラクターは特に問題ない状態で表示する。なお、図５Ｃにおける「通常状態」とは、（２）、（３）の状態以外の状態と考えてもよい。また、例えば、（２）人工知能応答出力装置１００１０がネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）よりも低性能なＬＬＭまたは応答定型文データベースによるによる応答生成処理へ切替えた状態では、人工知能応答出力装置１００１０は、ＡＩアシスタントまたはキャラクターを「眠い」状態で表示する。なお、「ＡＩアシスタントまたはキャラクターを「眠い」状態で表示する」とは、「ＡＩアシスタントまたはキャラクターが眠気を感じている状態を示す表示」と表現してもよい。

【０２５４】
（２）の応答生成処理は、（１）の通常状態におけるネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）による応答生成処理よりも性能が低い。よって、ＡＩアシスタントまたはキャラクターを「眠い」状態で表示することにより、ユーザに、ＡＩアシスタントまたはキャラクターの応答性能が低いことを暗に伝えることができる。これにより、低性能な応答に対するユーザが感じる違和感をより低減することが可能となる。なお、人工知能応答出力装置１００１０がネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）よりも低性能なＬＬＭまたは応答定型文データベースによるによる応答生成処理へ切替える切り替え条件は、図５Ａで説明したとおりなので、繰り返しの説明は省略する。

【０２５５】
また、図５Ｃ（２）の場合、人工知能応答出力装置１００１のタッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザから大規模言語モデルによる応答を求めるユーザ入力を行う前に、ユーザに、ＡＩアシスタントまたはキャラクターの応答性能が低いことを暗に伝えることが望ましい。よって、図５Ｃの（２）ＡＩアシスタントまたはキャラクターを「眠い」状態で表示する状態の開始タイミングは、ユーザから大規模言語モデルによる応答を求めるユーザ入力より前の、人工知能応答出力装置１００１０がネットワーク上の大規模言語モデル（ＡＰＩを用いて接続する大規模言語モデル）よりも低性能なＬＬＭまたは応答定型文データベースによるによる応答生成処理へ切替の時点の直後であることが望ましい。

【０２５６】
また、例えば、（３）人工知能応答出力装置１００１０が図５Ａで説明した不応答対応へ切替えた状態では、人工知能応答出力装置１００１０は、ＡＩアシスタントまたはキャラクターを「眠っている」状態で表示する。図５Ｂでも説明したとおり、人工知能応答出力装置１００１０が表示するＡＩアシスタントまたはキャラクターが「眠っている」状態で表示することにより、ユーザは「当該ＡＩアシスタントまたはキャラクターが応答しないのは、眠っているからである。」と理解することができ、ユーザが感じる違和感をより低減することが可能となる。なお、人工知能応答出力装置１００１０が図５Ａで説明した不応答対応へ切替える切替発生条件については、図５Ａの例３または例９などで説明した通りであるので、繰り返しの説明は省略する。なお、図５Ｃ（３）の場合、人工知能応答出力装置１００１のタッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザから大規模言語モデルによる応答を求めるユーザ入力を行う前に、ユーザが「当該ＡＩアシスタントまたはキャラクターが応答しないのは、眠っているからである。」と理解することが望ましい。よって、図５Ｃの（３）ＡＩアシスタントまたはキャラクターを「眠っている」状態で表示する状態の開始タイミングは、ユーザから大規模言語モデルによる応答を求めるユーザ入力より前の、人工知能応答出力装置１００１０が図５Ａで説明した不応答対応への切替の時点の直後であることが望ましい。

【０２５７】
なお、図５Ｃの表示例では、人工知能応答出力装置１００１０は、ユーザに対して、応

答生成処理に関する人工知能応答出力装置１００１０の状態の技術的な説明を、直接ユーザに対して行うことなく、ＡＩアシスタントまたはキャラクターの状態の変化として暗に反映させる表示を行う。これにより、応答生成処理に関する人工知能応答出力装置１００１０の状態の技術的な説明を直接ユーザに対して行う場合よりも、ユーザが感じる違和感をより低減することができる。また、応答生成処理に関する人工知能応答出力装置１００１０の状態が変化したにも関わらず、ＡＩアシスタントまたはキャラクターの表示状態を通常状態と同じままの表示とする場合よりも、ユーザが感じる違和感をより低減することができる。

【０２５８】

しかしながら、ユーザによっては、それぞれの状態における技術的な状態の説明をより正確に知りたい場合もあり得る。そこで、このようなユーザに対応するための表示例について、図５Ｄを用いて説明する。図５Ｄに示す表の行のうち、装置状態、表示状態説明の行は図５Ｃと全く同一であるので、繰り返しの説明は省略する。また、ＡＩアシスタントまたはキャラクターの表示例の行に示される、ＡＩアシスタントまたはキャラクターの表示例は、図５Ｃとほぼ同一であるが、表示例にはてなマーク（？）が表示されている点で相違する。当該はてなマーク（？）は、ユーザが人工知能応答出力装置１００１０に対して説明を要求する場合に操作するマークであり、ヘルプマークと称してもよい。

【０２５９】

図５Ｄの例では、ユーザが、図１Ｂの操作入力部１１０７または表示部１００１１が有するタッチパネルなどを介したユーザ操作により、当該はてなマーク（？）を選択すると、人工知能応答出力装置１００１０のＡＩアシスタントまたはキャラクターの表示は、ユーザ操作後表示例の行に示される表示例に変更される。具体的には、装置状態が（１）、（２）または（３）のいずれの状態であっても、それぞれの状態における技術的な状態の説明を表示する。例えば、図５Ｄの例では、装置の状態が（１）通常状態であれば、「通常状態です」と技術的に特に制限のない通常の状態であることを説明する表示を行えばよい。また、装置の状態が（２）低性能ＬＬＭまたは応答定型文データベースを使用している状態であれば、「低性能モードです」と、低性能状態であることを技術的に説明する表示を行えばよい。当該表示は、ＡＩアシスタントまたはキャラクターの表示が「眠い」状態で表示している要因を説明する表示と考えてもよい。

【０２６０】

この場合、さらに技術的に詳細な説明を行ってもよい。具体的には、「低性能ＬＬＭ使用モードです」または「定型文応答モードです」などの表示を行ってもよい。また、装置の状態が（３）不応答対応であれば、「ネットワーク接続不能状態です」と、不応答対応への切替えの要因を技術的に説明する表示を行えばよい。不応答対応への切替えの要因が、ネットワーク上のＬＬＭ（ＡＰＩを用いて接続する大規模言語モデル）からの応答が所定時間を超過していることである場合は、「ＬＬＭからの応答が遅延しています」などと表示してもよい。また、不応答対応への切替えの要因が、ネットワーク上のＬＬＭの利用量、ＡＰＩ利用量またはＡＰＩ利用料金が上限に到達したことである場合は、「ＬＬＭの利用量が上限に達しました」、「ＡＰＩの利用量が上限に達しました」、または、「ＡＰＩの利用料金が所定の額に達しました」などと表示してもよい。これらの表示は、ＡＩアシスタントまたはキャラクターの表示が「眠っている」状態で表示している要因を説明する表示と考えてもよい。

【０２６１】

以上説明した図５Ｄの表示例によれば、人工知能応答出力装置１００１０における応答生成処理に技術的な制約があったとしても、まずは、ユーザに対して直接的な説明を行わず、ＡＩアシスタントまたはキャラクターの表示状態の変化により、暗に装置の状態を示すことにより、ユーザが感じる違和感をより低減することができる。技術的な説明が不要なユーザに対しては当該表示がより好適である。さらに、技術的な状態の説明を行うための操作マークを表示することにより、当該マークを操作したユーザに対しては、人工知能応答出力装置１００１０における応答生成処理の状態（通常状態または技術的制約がある

状態）を技術的に説明する表示を行う。これにより、技術的な状態を正確に知りたいユーザにとってもより好適な表示を行うことが可能となる。

【０２６２】

なお、図５Ｂ、図５Ｃ、図５Ｄ、の例では、「不応答対応」であるときの、ＡＩアシスタントまたはキャラクターの表示状態の例として「眠っている」状態を示しているが、これは一例であり、本実施例の態様はこれに限られない。「眠っている」状態に替えて、「休憩中」など、応答できない状況を暗に示す他の表示状態としてもよい。また、図５Ｃ、図５Ｄ、の例では、低性能ＬＬＭまたは応答定型文データベースを使用している状態であるときの、ＡＩアシスタントまたはキャラクターの表示状態の例として「眠い」状態を示しているが、これは一例であり、本実施例の態様はこれに限られない。「はらぺこ」など、ＡＩアシスタントまたはキャラクターの応答性能が低いことを暗に示す他の表示状態としてもよい。

【０２６３】

以上説明した、実施例５に係る人工知能応答出力装置や、人工知能応答出力システムによれば、ネットワーク上の大規模言語モデルと人工知能応答出力装置の接続状態、ネットワーク上の大規模言語モデルからの応答遅延状態、またはネットワーク上の大規模言語モデルの利用量などに応じて、人工知能応答出力装置が用いる応答生成処理を、より好適に切替えることが可能となる。また、実施例５に係る人工知能応答出力装置をＡＩアシスタント装置またはキャラクター会話装置として構成する場合に、ユーザにとってより違和感が少ない表示を行うことが可能となる。

【０２６４】

＜実施例６＞

次に、本発明の実施例６は、実施例１～５の各図で説明した人工知能応答出力装置１００１０、または人工知能応答出力システムを改良したものである。具体的には、人工知能応答出力装置１００１０の応答生成処理を、ネットワーク上の大規模言語モデルによる応答生成処理または人工知能応答出力装置１００１０が備えるローカルの大規模言語モデル（ローカルＬＬＭ処理部１００２８など）による応答生成処理と、応答定型文データベースによる応答生成処理を、より好適に組み合わせて応答出力を生成する例である。本実施例では、これらの実施例との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。

【０２６５】

上述の実施例同様、人工知能応答出力装置１００１０は、人工知能応答出力装置、キャラクター会話装置、ＡＩアシスタント装置、ＡＩアシスタント表示装置、人工知能インタフェース装置と称してもよい。人工知能応答出力装置１００１０と大規模言語モデルサーバを含むシステムは、人工知能応答出力システム、キャラクター会話システム、ＡＩアシスタントシステム、ＡＩアシスタント表示システム、人工知能インタフェースシステムと称してもよい。

【０２６６】

図６を用いて、本発明の実施例６の人工知能応答出力装置１００１における応答生成処理の一例について説明する。図６は、実施例６に係る、本発明の実施例６の人工知能応答出力装置１００１における応答生成処理のフローチャートの一例が示されている。具体的には、上から下へ時刻が進行する時間軸、処理フロー、応答出力例が示されている。応答出力例に示される応答の出力は、人工知能応答出力装置１００１の表示部１００１１による表示または音声出力部１１４０による音声出力を介して行えばよい。

【０２６７】

図６の例では、まず、時刻ｔ０において、人工知能応答出力装置１００１のタッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザから大規模言語モデルによる応答を求めるユーザ入力があり、人工知能応答出力装置１００１の制御部１１１０が、当該ユーザ入力を取得する（ステップ６００）。次に、時刻ｔ１において、制御部１１１０が、ストレージ部１１７０に格納される応答定型文データベースを用いた応答出力の

ための準備を開始し、応答定型文データベースを用いた応答出力を開始する（ステップ６０１）。図６の例では、時刻ｔ２において、応答定型文データベースを用いた応答出力が開始されており、図に示すように、定型文応答が出力中であり、出力完了していない。図中の「おはよう」は、「おはようございます。…」と続く文章の途中までの出力を示したものである。

【０２６８】

時刻ｔ３において、応答定型文データベースを用いた応答出力が完了する前に、制御部１１１０が、ステップ６００で取得したユーザ入力にもとづいて指示文を生成し、ネットワーク上の大規模言語モデルまたは人工知能応答出力装置１００１０が備えるローカルの大規模言語モデル（ローカルＬＬＭ処理部１００２８など）へ生成した指示文を送信し、大規言語モデルによる応答の要求を開始する（ステップ６０２）。さらに、時刻ｔ４において、応答定型文データベースを用いた応答出力が完了する前に、制御部１１１０は、大規模言語モデルからの応答取得を開始する（ステップ６０３）。

【０２６９】

時刻ｔ５において、応答定型文データベースを用いた応答出力が完了した応答出力例を示す。例えば、図６には、時刻ｔ５において、応答定型文データベースに格納された定型文とメモリに格納された日付情報などを用いて「おはようございます。本日は〇月〇日ですね。」という応答出力の表示が完了した例を示している。ここで、応答定型文データベースを用いた応答出力が完了する前の時刻ｔ４において、制御部１１１０は、既に、大規模言語モデルからの応答取得を開始している。よって、制御部１１１０は、応答定型文データベースを用いた応答出力表示が完了した時刻ｔ５に続く時刻ｔ６に、当該応答定型文データベースを用いた応答出力に続けて、大規模言語モデルからの応答出力を開始する（ステップ６０４）。その後、時刻ｔ７において、応答定型文データベースを用いた応答出力に続けて大規模言語モデルからの応答が出力される。大規模言語モデルからの応答出力が完了すると、図６に示す処理フローによる応答出力が完了する（ステップ６０５）。

【０２７０】

次に、本発明の図６に示す処理フローの効果について説明する。大規模言語モデルの処理には多くの計算リソースが必要である。一般的に、学習よりも計算リソースが少なく済む推論をＧＰＵ（Graphics Processing Unit）を用いて処理したとしても、制御部が大規模言語モデルへ応答要求を開始してから大規模言語モデルから応答を取得できるようになるまで数秒～十数秒かかる場合もある。この期間は、図６に示す時刻ｔ３から時刻ｔ４に対応する期間である。また、ユーザ入力のあった時刻ｔ０から時刻ｔ４に至るまでは、制御部１１１０は大規模言語モデルからの応答出力は取得出来ていないため、ユーザへ大規模言語モデルからの応答を出力することは出来ない。

【０２７１】

よって、図６に示すステップ６０１の応答定型文データベースを用いた応答出力のための準備の開始と、当該応答定型文データベースを用いた応答出力の開始が存在しない処理フローでは、ユーザはユーザ入力を行った時刻ｔ０から時刻ｔ４に至るまで数秒～十数秒を超える期間について、人工知能応答出力装置１００１からの応答がないまま待ち続けることとなる場合がある。例えば、人工知能応答出力装置１００１をＡＩアシスタント装置またはキャラクター会話装置として構成する場合など、当該待ち時間は、ユーザに違和感を与える可能性がある。

【０２７２】

これに対し、図６に示す本発明の実施例６に係る処理フローでは、制御部１１１０は、大規模言語モデルからの応答取得が開始される前に、大規模言語モデルの処理よりも計算リソースが少なくて済む応答定型文データベースを用いた応答出力の処理を開始する。これにより、ユーザは、時刻ｔ０から時刻ｔ４に至るまでの間、人工知能応答出力装置１００１からの応答がないまま待ち続ける、という状態にならなくて済む。ユーザにとっては、応答定型文データベースを用いた応答出力であっても大規模言語モデルからの応答出力であっても、人工知能応答出力装置１００１からの応答であることには変わりない。

【０２７３】
よって、図６に示す処理フローでは、ステップ６０３の前に、ステップ６０１を設けることによって、ユーザに対する人工知能応答出力装置１００１の応答を疑似的に早めることができる。これにより、待ち時間の長さに起因するユーザの違和感をより低減することが可能である。また、ステップ６０４により、大規模言語モデルからの応答を、当該応答定型文データベースを用いた応答に続けて出力することにより、あたかもこれらの出力が一連のより自然な出力であるかのようにユーザに認識させることができる。
【０２７４】
以上説明した、実施例６に係る人工知能応答出力装置や、人工知能応答出力システムによれば、ユーザにとって人工知能応答出力装置からの応答待ち時間を短縮することができ、ユーザの感じる違和感をより低減することができる。
【０２７５】
＜実施例７＞
本発明の実施例７は、実施例１～６の各図で説明した人工知能応答出力装置１００１０、または人工知能応答出力システムを改良したものである。なお、実施例７では、実施例１～６との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。
【０２７６】
上述の実施例同様、人工知能応答出力装置１００１０は、人工知能応答出力装置、キャラクター会話装置、ＡＩアシスタント装置、ＡＩアシスタント表示装置、人工知能インタフェース装置と称してもよい。人工知能応答出力装置１００１０と大規模言語モデルサーバを含むシステムは、人工知能応答出力システム、キャラクター会話システム、ＡＩアシスタントシステム、ＡＩアシスタント表示システム、人工知能インタフェースシステムと称してもよい。
【０２７７】
ここで、人工知能応答出力装置１００１０、または人工知能応答出力システムにおいて、人工知能応答出力装置１００１０を介して入力された大規模言語モデルによる応答を求めるユーザ入力にもとづいて指示文を生成し、当該指示文を大規模言語モデルに送信して、応答を取得するときに、指示文に不適切な内容が含まれている場合がある。また、大規模言語モデルからの応答に不適切な内容が含まれる場合がある。これらの場合に、大規模言語モデルからの応答内容を所定の内容に差替えることで、不適切な応答が出力されないようにフィルタする技術が考えられる。
【０２７８】
図７Ａに、不適切な応答が出力されないようにフィルタする場合の差替え後の応答の一例を示す。例えば、図７Ａ（１）は、人工知能応答出力装置１００１０から大規模言語モデルへ送信された指示文に暴力的な内容の応答を求める記載が含まれていた場合や、大規模言語モデルが生成した応答に暴力的な内容が含まれている場合に、差替え後の応答として出力する自然言語の文章の一例である。図７Ａ（２）は、人工知能応答出力装置１００１０から大規模言語モデルへ送信された指示文に性的な内容の応答を求める記載が含まれていた場合や、大規模言語モデルが生成した応答に性的な内容が含まれている場合に、差替え後の応答として出力する自然言語の文章の一例である。図７Ａ（３）は、人工知能応答出力装置１００１０から大規模言語モデルへ送信された指示文に公共の安全を害する内容の応答を求める記載が含まれていた場合や、大規模言語モデルが生成した応答に公共の安全を害する内容が含まれている場合に、差替え後の応答として出力する自然言語の文章の一例である。図７Ａ（４）は、人工知能応答出力装置１００１０から大規模言語モデルへ送信された指示文に特定の個人のプライバシーを侵害する内容の応答を求める記載が含まれていた場合や、大規模言語モデルが生成した応答に特定の個人のプライバシーを侵害する内容が含まれている場合に、差替え後の応答として出力する自然言語の文章の一例である。このように、大規模言語モデルは、不適切な応答が出力されないようにフィルタするための差替え文章を応答として人工知能応答出力装置１００１０へ出力し、人工知能応

答出力装置１００１０が当該差替え文章を応答としてユーザに出力することができる。
【０２７９】
しかしながら、図７Ａに示されるような、差替え文章をそのまま応答としてユーザに出力することが好適ではない場合もあり得る。例えば、実施例１～６の各図で説明した人工知能応答出力装置１００１０、または人工知能応答出力システムのうち、例えば、人工知能の大規模言語モデルに、大規模言語モデル自体の名前や、演じるべき役割、会話の特徴などを、初期設定として指示することにより、大規模言語モデルからの応答を、あたかも一貫性のある個性を持ったキャラクターのように演じさせている場合に、図７Ａに示されるような、自己がＡＩ（人工知能）であることを表明して状況を説明する差替え文章をそのまま応答としてユーザに出力する場合を考える。この場合、ユーザは、直前まで人工知能応答出力装置１００１０、または人工知能応答出力システムからの応答を一貫性のある個性を持ったキャラクターからの応答と認識していても、図７Ａに示されるような差替え文章の応答が出力されることにより、これまでの一連の応答もＡＩ（人工知能）の応答に過ぎないことを認識することになりえる。この状況になってしまうと、初期設定後の応答で構築してきた、ユーザが認識する当該キャラクターの一貫性のある個性が台無しになってしまう。そこで、本実施例に係る人工知能応答出力装置１００１０、または人工知能応答出力システムでは、大規模言語モデルが不適切な応答が出力されないようにフィルタする場合に、より好適な応答をユーザに出力するための動作および処理を行う。
【０２８０】
以降に本実施例に係る人工知能応答出力装置１００１０、または人工知能応答出力システムにおける具体的な動作および処理について説明する。
【０２８１】
まず、図７Ｂには、実施例７に係る人工知能応答出力装置１００１０、または人工知能応答出力システムにおける、不適切な応答が出力されないように処理を行うフローチャートの一例を示す。
【０２８２】
まず、ステップ７０１において、人工知能応答出力装置１００１０のタッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザ入力があった場合に、制御部１１１０は当該ユーザ入力にもとづいて質問または要求などの指示文１を生成し、大規模言語モデルサーバ１９００１および／またはマルチモーダルな大規模言語モデルサーバ２０００１などにあるネットワーク上の大規模言語モデルまたはローカルＬＬＭ処理部１００２８などのローカルの大規模言語モデルへ当該指示文１を送信する（Ｓ７０１）。次に、ステップ７１１において、当該指示文１を受信した大規模言語モデルは、不適切フィルタを行う必要があるか否かを判断する不適切フィルタ要否判断処理を行う（Ｓ７１１）。不適切フィルタ要否判断処理の詳細は後述する。
【０２８３】
ここで、ステップ７１１の不適切フィルタ要否判断処理により、不適切フィルタを行う必要があると当該大規模言語モデルが判断した場合は、不適切な内容がフィルタされた自然言語を含む応答Ａを人工知能応答出力装置１００１０の制御部１１１０へ送信する。不適切な内容がフィルタされた自然言語を含む応答Ａの一例は、上述した図７Ａに示すような差替え文章である。本実施例にかかる大規模言語モデルは、このとき、不適切な内容がフィルタされた自然言語を含む応答Ａに加えて、またはこれに替えて、内容が不適切でありフィルタされた応答であることを示す不適切フィルタフラグＢを人工知能応答出力装置１００１０の制御部１１１０へ送信する。不適切フィルタフラグＢは、すなわち当該応答にはフィルタ処理があったことを示すフラグであるともいえる。なお、本発明の実施例における「フラグ」との表現は「制御情報」と言い換えても構わない。ここで、ステップ７０２では、不適切な内容がフィルタされた自然言語を含む応答Ａおよび／または不適切フィルタフラグＢを受信した人工知能応答出力装置１００１０の制御部１１１０は、当該不適切な内容がフィルタされた自然言語を含む応答Ａおよび／または不適切フィルタフラグＢにもとづいて、応答Ａに対応するための処理を行う（Ｓ７０２）。

【０２８４】
一方、ステップ７１１の不適切フィルタ要否判断処理により、不適切フィルタを行う必要がない、すなわち適切である、と当該大規模言語モデルが判断した場合は、当該大規模言語モデルは、生成した自然言語を含む通常の応答Ｃを、不適切な内容のフィルタがされない状態で、人工知能応答出力装置１００１０の制御部１１１０へ送信する。このとき、本実施例にかかる大規模言語モデルは、大規模言語モデルが生成した自然言語を含む通常の応答Ｃに加えて、不適切な内容ではないことを示す不適切フィルタフラグＤを人工知能応答出力装置１００１０の制御部１１１０へ送信してもよい。ここで、ステップ７０３では、大規模言語モデルが生成した自然言語を含む通常の応答Ｃおよび／または不適切な内容ではないことを示す不適切フィルタフラグＤを受信した人工知能応答出力装置１００１０の制御部１１１０は、大規模言語モデルが生成した自然言語を含む通常の応答Ｃおよび／または不適切な内容ではないことを示す不適切フィルタフラグＤにもとづいて、応答Ｃに対応するための処理を行う（Ｓ７０３）。
【０２８５】
以上説明した、図７Ｂのフローチャートによる、人工知能応答出力装置１００１０、または人工知能応答出力システムの処理によれば、大規模言語モデルからの応答に加えて、またはこれに替えて、内容が不適切でありフィルタされた応答であることを示すフラグまたは応答の内容が不適切ではないことを示すフラグを、大規模言語モデルから人工知能応答出力装置１００１０へ送信することができる。
【０２８６】
次に、図７Ｃを用いて、図７Ｂのステップ７１１における不適切フィルタ要否判断処理の一例について説明する。図７Ｃには、判断処理に係る表が示されている。図７Ｃに示す判断処理の例では、指示文に不適切キーワードが含まれているか否か、および、生成データに不適切データが含まれているか否か、の２つの条件を組み合わせて、「不適切であり、フィルタ要」であるか「不適切ではない、フィルタ不要」であるかを判断している。図７Ｃの例では、指示文に不適切キーワードが含まれておらず、生成データに不適切データが含まれていない場合は「不適切ではない、フィルタ不要」と判断する。また、指示文に不適切キーワードが含まれているか、または、生成データに不適切データが含まれている場合は、いずれの場合も「不適切であり、フィルタ要」と判断する。
【０２８７】
以上説明した、図７Ｃの判断処理によれば、不適切な応答を出力する可能性をより好適に判断することが可能となる。
【０２８８】
次に、図７Ｄを用いて、図７Ｂに示す不適切フィルタフラグの具体的な一例について説明する。図７Ｄには（１）の例と（２）の例との、不適切フィルタフラグの２つの例が示されている。
【０２８９】
まず、例えば、図７Ｄ（１）の不適切フィルタフラグの例では、単純に、「不適切ではない」と判断され応答が「フィルタ無し」の状態と、「不適切」と判断し、かつ応答が「フィルタされた」状態との、２状態を示している。例えば、１ビットのフラグを用いて、当該フラグが「０」である場合に「不適切ではない」と判断され、応答が「フィルタ無し」の状態を示すように構成すればよい。このとき、当該フラグが「１」である場合に、「不適切」と判断し、かつ応答が「フィルタされた」状態を示すように構成すればよい。
【０２９０】
以上説明した、図７Ｄ（１）の不適切フィルタフラグによれば、「不適切ではない」と判断され応答が「フィルタ無し」の状態と、「不適切」と判断し、かつ応答が「フィルタされた」状態とを、１ビットでより好適に識別することが可能となる。また、不適切な応答を出力する可能性をより好適に判断することが可能となる。
【０２９１】
これに対し、図７Ｄ（２）のフラグの例では、「不適切」と判断し、かつ応答が「フィ

ルタされた」状態において、「不適切」と判断した理由もフラグによって識別可能とする例である。具体的には、（２）の不適切フィルタフラグの例では、１０進法表記で「０」～「７」までの８つの状態を示すように構成している。（２）の不適切フィルタフラグが「０」を示す場合は、（１）と同様の状態であるので、繰り返しの説明は省略する。

【０２９２】

ここで、（２）の不適切フィルタフラグが「１」～「６」を示す場合は、当該フラグは、「不適切」と判断し、かつ応答が「フィルタされた」状態を示しており、かつ、「不適切」と判断された理由が当該フラグの値によって、それぞれ識別可能となっている。例えば、不適切フィルタフラグが「１」を示す場合は、指示文または生成データに「暴力」に関する内容が含まれており、これを理由に「不適切」と判断し、かつ応答が「フィルタされた」状態であることを示している。また、不適切フィルタフラグが「２」を示す場合は、指示文または生成データに「性的」な内容が含まれており、これを理由に「不適切」と判断し、かつ応答が「フィルタされた」状態であることを示している。また、不適切フィルタフラグが「３」を示す場合は、指示文または生成データに「公共安全」を「侵害」する恐れがある内容が含まれており、これを理由に「不適切」と判断し、かつ応答が「フィルタされた」状態であることを示している。

【０２９３】

また、不適切フィルタフラグが「４」を示す場合は、指示文または生成データに「個人プライバシー」を侵害する恐れがある内容が含まれており、これを理由に「不適切」と判断し、かつ応答が「フィルタされた」状態であることを示している。また、不適切フィルタフラグが「５」を示す場合は、指示文または生成データに「差別」を助長する恐れがある内容が含まれており、これを理由に「不適切」と判断し、かつ応答が「フィルタされた」状態であることを示している。また、不適切フィルタフラグが「６」を示す場合は、指示文または生成データに「学術的懸念」がある内容が含まれており、これを理由に「不適切」と判断し、かつ応答が「フィルタされた」状態であることを示している。なお、図７Ｄ（２）の例では、不適切フィルタフラグが「７」を示す場合を用意し、この場合は、上述の不適切フィルタフラグが「０」～「６」の状況には当てはまらない状況であることを意味するように構成している。

【０２９４】

なお、図７Ｄ（２）のフラグの例では、１０進法表記で「０」～「７」までの８つの状態を示すように構成している。よって、不適切フィルタフラグは、３ビットのフラグで構成すればよい。

【０２９５】

以上説明した、図７Ｄ（２）の不適切フィルタフラグによれば、「不適切ではない」と判断され応答が「フィルタ無し」の状態と、「不適切」と判断し、かつ応答が「フィルタされた」状態とを、識別可能なだけでなく、「不適切」と判断した理由も識別することが可能となる。

【０２９６】

なお、図７Ｂに示されるＬＬＭが、大規模言語モデルサーバ１９００１および／またはマルチモーダルな大規模言語モデルサーバ２０００１などにあるネットワーク上の大規模言語モデルである場合は、図７Ｄで説明したフラグは、インターネット１９０００などのネットワークを介して、大規模言語モデルから人工知能応答出力装置１００１０へ送信される。人工知能応答出力装置１００１０が受信した不適切フィルタフラグは、制御部１１１０が処理に用いることができる。

【０２９７】

また、図７Ｂに示されるＬＬＭが、ローカルＬＬＭ処理部１００２８などのローカルの大規模言語モデルである場合は、図７Ｄで説明したフラグは、人工知能応答出力装置１００１０内で、ローカルＬＬＭ処理部１００２８から制御部１１１０へ送信される。制御部１１１０は受信した不適切フィルタフラグを処理に用いることができる。

【０２９８】

次に、図７Ｅを用いて、不適切フィルタフラグを受信した人工知能応答出力装置１００１０の制御部１１１０が行う制御の一例について説明する。図７Ｅは、図２Ｈで示した複数のキャラクターのそれぞれのキャラクターについて、上述の不適切フィルタフラグに応じた応答を用意して格納したデータベースの例を示している。キャラクターＩＤおよびキャラクターの名前については、実施例２、実施例３、または実施例４と同様であるため、繰り返しの説明は省略する。

【０２９９】

実施例２、実施例３、または実施例４の実施例では、初期設定指示文、会話履歴、またはユーザＩＤとの対応関係などを用いて、ユーザにとって、それぞれのキャラクターが一貫性のある個性を持った存在であるかのように感じる工夫がなされている。このとき、大規模言語モデルからの応答に不適切フィルタ処理がなされ、例えば、図７Ａの（１）～（４）に示すような不適切フィルタ処理後の差替え文を人工知能応答出力装置１００１０がユーザに出力してしまう場合を考える。

【０３００】

ここで、図７Ａの（１）～（４）の文では、「私はＡＩである」などの表現が含まれており、また図７Ａ（１）、（２）、（４）などには、「プログラムされています」などの表現が含まれている。このため、この出力を視聴したユーザは、図２Ｈで示した複数のキャラクターがソフトウェアでプログラムされた人工知能であることを強く認識してしまい、これまで築き上げたキャラクターそれぞれの一貫性のある個性を認識できなくなってしまう恐れがある。

【０３０１】

そこで、本実施例の図７Ｅの例では、たとえ大規模言語モデルが生成した、フィルタされた自然言語の応答文が、図７Ａの（１）～（４）の内容であっても、人工知能応答出力装置１００１０の制御部１１１０は、図７Ｅに示すデータベースを用いて、それぞれのキャラクターと不適切フィルタフラグに応じた応答となる文章（応答文）を選択し、当該応答をユーザに対して出力する制御を行う。すなわち、制御部１１１０の当該制御では、大規模言語モデルが生成した自然言語の応答文（例えば、図７Ａの（１）～（４）の内容）を直接ユーザに対して出力するのではなく図７Ｅのデータベースに格納される内容に差し替えてユーザに出力する。図７Ｅに示すデータベースは、図３のストレージ部１１７０に格納しておけばよい。

【０３０２】

例えば、人工知能応答出力装置１００１０にキャラクターＩＤが１のキャラクター「Ｋｏｔｏ」が表示されている場合に、図７Ｂのステップ７０２において、制御部１１１０が図７Ｄ（２）に示す、値が１である不適切フィルタフラグ（指示文または生成データに「暴力」に関する内容が含まれていることを示すフラグ）を受信した場合には、制御部１１１０はストレージ部１１７０に格納されている図７Ｅのデータベースを参照し、「Ｋｏｔｏは、暴力反対ですよ。」という応答文を選択して、ユーザに対して出力するように制御する。このような応答文出力制御を行うことにより、図７Ａ（１）のような、大規模言語モデルが生成した応答文を直接出力される場合よりも、ユーザは、人工知能応答出力装置１００１０が表示しているキャラクターの個性に沿った応答であると認識することができ、好適である。

【０３０３】

また例えば、人工知能応答出力装置１００１０にキャラクターＩＤが１のキャラクター「Ｋｏｔｏ」が表示されている場合に、図７Ｂのステップ７０２において、制御部１１１０が図７Ｄ（２）に示す、値が２である不適切フィルタフラグ（指示文または生成データに「性的」な内容が含まれていることを示すフラグ）を受信した場合、および、制御部１１１０が図７Ｄ（２）に示す、値が３である不適切フィルタフラグ（指示文または生成データに「公共安全」を「侵害」する恐れがある内容が含まれていることを示すフラグ）を受信した場合、もそれぞれ、制御部１１１０はストレージ部１１７０に格納されている図７Ｅのデータベースを参照して、対応する応答文をユーザに出力するように制御すればよ

い。このようにすれば、ユーザは、人工知能応答出力装置１００１０に表示されているキャラクター「Ｋｏｔｏ」の個性に沿った応答であると認識することができ、より好適である。

【０３０４】

同様に、人工知能応答出力装置１００１０にキャラクターＩＤが２のキャラクター「Ｔｏｍ」が表示されている場合に、図７Ｂのステップ７０２において、制御部１１１０が図７Ｄ（２）に示す、値が１である不適切フィルタフラグ（指示文または生成データに「暴力」に関する内容が含まれていることを示すフラグ）を受信した場合、制御部１１１０が図７Ｄ（２）に示す、値が２である不適切フィルタフラグ（指示文または生成データに「性的」な内容が含まれていることを示すフラグ）を受信した場合、および、制御部１１１０が図７Ｄ（２）に示す、値が３である不適切フィルタフラグ（指示文または生成データに「公共安全」を「侵害」する恐れがある内容が含まれていることを示すフラグ）を受信した場合、もそれぞれ、制御部１１１０はストレージ部１１７０に格納されている図７Ｅのデータベースを参照して、対応する応答文をユーザに出力するように制御すればよい。このようにすれば、ユーザは、人工知能応答出力装置１００１０に表示されているキャラクター「Ｔｏｍ」の個性に沿った応答であると認識することができ、より好適である。

【０３０５】

同様に、人工知能応答出力装置１００１０にキャラクターＩＤが３のキャラクター「Ｎｅｃｃｏ」が表示されている場合に、図７Ｂのステップ７０２において、制御部１１１０が図７Ｄ（２）に示す、値が１である不適切フィルタフラグ（指示文または生成データに「暴力」に関する内容が含まれていることを示すフラグ）を受信した場合、制御部１１１０が図７Ｄ（２）に示す、値が２である不適切フィルタフラグ（指示文または生成データに「性的」な内容が含まれていることを示すフラグ）を受信した場合、および、制御部１１１０が図７Ｄ（２）に示す、値が３である不適切フィルタフラグ（指示文または生成データに「公共安全」を「侵害」する恐れがある内容が含まれていることを示すフラグ）を受信した場合、もそれぞれ、制御部１１１０はストレージ部１１７０に格納されている図７Ｅのデータベースを参照して、対応する応答文をユーザに出力するように制御すればよい。

【０３０６】

なお、図７Ｅの例では、キャラクター「Ｎｅｃｃｏ」は、図７Ｄ（２）に示す、値が１～６である不適切フィルタフラグを受信した場合、制御部１１１０は、一貫して「Ｎｅｃｃｏには良くわからないニャー。」という応答を出力する制御を行う。このようにすることにより、人工知能応答出力装置１００１０は、キャラクター「Ｎｅｃｃｏ」が不適切な指示文や応答については、わからないふりをして回答をはぐらかす、という個性を有しているという印象をユーザに与えることができる。この場合も、ユーザは、人工知能応答出力装置１００１０に表示されているキャラクター「Ｎｅｃｃｏ」の個性に沿った応答であると認識することができ、より好適である。

【０３０７】

以上説明した、図７Ｅのデータベースを用いた制御によれば、大規模言語モデルからの応答が、大規模言語モデルによってフィルタされた応答であったとしても、人工知能応答出力装置が表示しているキャラクターの個性に沿った応答に差し替えて出力することができ、ユーザにとってより好適である。

【０３０８】

なお、以上説明した図７Ｅの応答文データベースは、ストレージ部１１７０に格納され、人工知能応答出力装置１００１０の制御部１１１０がこれを用いればよい。しかしながら、図７Ｅの応答文データベースを大規模言語モデルサーバ１９００１側または大規模言語モデルサーバ２０００１側に備えてもよい。

【０３０９】

この場合は、大規模言語モデルサーバ１９００１が有する制御部または大規模言語モデルサーバ２０００１が有する制御部が、人工知能応答出力装置１００１０に表示されてい

るキャラクターのキャラクターＩＤを人工知能応答出力装置１００１０から取得し、それぞれのサーバの大規模言語モデルが生成した応答文に替えて、図７Ｅの応答文データベースの対応する応答文に差し替えて人工知能応答出力装置１００１０に送信してもよい。この場合、人工知能応答出力装置１００１０は当該応答文をユーザに出力すればよい。

【０３１０】

このようにすれば、人工知能応答出力装置１００１０に図７Ｅの応答文データベースが備えられていない場合でも、図７Ｅの応答文データベースを用いた応答の生成が可能な人工知能応答出力システムが実現可能となる。この場合、大規模言語モデルサーバ１９００１または大規模言語モデルサーバ２０００１の内部では、それぞれの大規模言語モデルとそれぞれの制御部間で、図７Ｄに示すフラグの送受信を行ってもよい。

【０３１１】

＜実施例８＞

本発明の実施例８は、実施例１～７の各図で説明した人工知能応答出力装置１００１０、または人工知能応答出力システムを改良したものである。なお、実施例８では、実施例１～７との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。

【０３１２】

上述の実施例同様、人工知能応答出力装置１００１０は、人工知能応答出力装置、キャラクター会話装置、ＡＩアシスタント装置、ＡＩアシスタント表示装置、人工知能インタフェース装置と称してもよい。人工知能応答出力装置１００１０と大規模言語モデルサーバを含むシステムは、人工知能応答出力システム、キャラクター会話システム、ＡＩアシスタントシステム、ＡＩアシスタント表示システム、人工知能インタフェースシステムと称してもよい。

【０３１３】

本発明の実施例８に係る人工知能応答出力装置１００１０、または人工知能応答出力システムでは、ユーザから入力された指示文を大規模言語モデルに送信する前に変換処理を行う。

【０３１４】

図８Ａを用いて、本発明の実施例８に係る人工知能応答出力装置１００１０、または人工知能応答出力システムにおける上記変換処理を含む一連の処理のフローチャートの一例について説明する。まず、ステップ８０１において、人工知能応答出力装置１００１０のタッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザ入力があった場合に、制御部１１１０は当該ユーザ入力にもとづいて質問または要求などの指示文１を生成する（Ｓ８０１）。

【０３１５】

次に、ステップ８０２において、人工知能応答出力装置１００１０の制御部１１１０は指示文１に対して変換処理を行って指示文２を生成し、生成した指示文２を、大規模言語モデルサーバ１９００１および／またはマルチモーダルな大規模言語モデルサーバ２０００１などにあるネットワーク上の大規模言語モデルまたはローカルＬＬＭ処理部１００２８などのローカルの大規模言語モデルに送信する（Ｓ８０２）。当該変換処理の詳細については後述する。

【０３１６】

次に、ステップ８１１において、当該指示文２を受信した大規模言語モデルは、指示文２に基づいて応答生成処理を行って自然言語を含む応答Ａを生成し、当該応答Ａを人工知能応答出力装置１００１０の制御部１１１０へ送信する。（Ｓ８１１）。

【０３１７】

次に、ステップ８０３では、自然言語を含む応答Ａを受信した人工知能応答出力装置１００１０の制御部１１１０は、応答Ａに対応するための処理を行う（Ｓ８０３）。ステップ８０３における応答Ａに対応するための処理とは、応答Ａに含まれる自然言語の文章をユーザに対して出力する処理、応答Ａの内容に基づいて人工知能応答出力装置１００１０

の各部を制御して各種動作を行う処理、または上述の各実施例で説明したデータベースを用いた応答文差替処理などである。

【０３１８】

次に、図８Ａのステップ８０２における指示文の変換処理の詳細の一例について、図８Ｂを用いて説明する。図８Ｂには、例１、例２、例３のそれぞれ異なる変換処理の例が示されている。図８Ｂの表中の「ユーザ指示文」は、人工知能応答出力装置１００１０のタッチパネル、マイク１１３９または操作入力部１１０７を介したユーザ入力にもとづいて生成された指示文であり、図８Ａの指示文１に対応するものである。また、図８Ｂの表中の「変換後指示文」は、人工知能応答出力装置１００１０の制御部１１１０が、「ユーザ指示文」に対して変換処理を行って生成した指示文であり、図８Ａの指示文２に対応するものである。以下、例１、例２、例３のそれぞれの変換処理について説明する。

【０３１９】

まず、図８Ｂの例１の変換処理における変換方法は、ユーザ指示文の自然言語に対し補助の自然言語である補助自然言語を加える変換処理を行うものである。例えば、指示文１として「１＋５を計算してください。１０－３を計算してください。両者の結果の和を算出してください。」が入力された場合、例１の変換処理では、元の指示文１の自然言語である「１＋５を計算してください。１０－３を計算してください。両者の結果の和を算出してください。」に、補助自然言語として「応答の過程の説明を追加してください。」との文言を加える変換処理を行っている。当該変換処理により、ユーザから入力された指示文１には含まれていなかった「応答の過程の説明を追加してください。」との補助自然言語が含まれた状態で、大規模言語モデルへ指示文２が送信される。

【０３２０】

ここで、大規模言語モデルによっては、当該例１の補助指示文のような応答生成の過程の指示を受けることにより、当該指示が無い場合に比べて応答精度が向上する場合がある。したがって、例１の変換処理は、ユーザが意図しなくとも、そのような大規模言語モデルで応答精度を向上させることができ、より好適である。なお、補助自然言語の内容は、例えば定型文として、予めストレージ部１１７０に格納しておき、制御部１１１０が変換処理のときに当該補助自然言語の定型文をストレージ部１１７０から読み出して用いればよい。

【０３２１】

図８Ｂの例２の変換処理における変換方法は、例２の変換処理における補助自然言語を英語とするものである。ユーザ指示文である指示文１の自然言語が非英語であったとしても、当該ユーザ指示文の非英語の自然言語に加えて、英語の補助自然言語を加える変換処理である。図８Ｂの表中の例２の具体例としては、例１と同様の指示文１が入力された場合について説明する。例えば、例１の変換処理では「応答の過程の説明を追加してください。」がユーザ指示文である指示文１と同じ言語で追加されていたところ、例２の変換処理では、ユーザ指示文の自然言語が非英語であっても、補助自然言語として英語の「Please add explanation of the process.」という文言を追加する。

【０３２２】

大規模言語モデルの多くは、インターネット経由でアクセス可能な既存の自然言語情報を学習に用いている。ここで、これらの既存の自然言語情報における英語の自然言語情報が占める割合が大きいため、学習規模が大きい高性能な大規模言語では、英語を用いた指示文の方が親和性が高く、応答精度が高くなる傾向がある。

【０３２３】

したがって、補助自然言語の意味する内容が同じであっても、非英語の補助自然言語を追加するよりも、例２のように、英語の補助自然言語を追加する方が、大規模言語モデルからより高精度な応答が得られる可能性が高い点で、より好適である。例２においても、補助自然言語の内容は、例えば定型文として、予めストレージ部１１７０に格納しておき、制御部１１１０が変換処理のときに当該補助自然言語の定型文をストレージ部１１７０から読み出して用いればよい。

【０３２４】
次に、図８Ｂの例３の変換処理における変換方法は、ユーザ指示文である指示文１の自然言語が非英語であった場合に、英語に変換する処理である。非英語の自然言語の英語の自然言語への変換は、いわゆる翻訳機能である。翻訳機能は既存の技術を用いればよい。例えば、人工知能応答出力装置１００１０のストレージ部１１７０に翻訳機能を有するプログラム（ソフトウェア）を格納しておき、当該プログラムをメモリ１１０９に展開して、制御部１１１０が当該プログラムを実行してもよい。また、制御部１１１０の制御により、通信部１１３２を介して外部の翻訳サーバに非英語である指示文１の自然言語を送信し、外部の翻訳サーバに当該指示文１の自然言語を英語に変換する処理を実行させてもよい。

【０３２５】
制御部１１１０は、変換後の英語の自然言語を当該サーバから通信部１１３２を介して受信して、図８Ｂの例３に示す指示文２として用いればよい。例３の具体例としては、ユーザ指示文である指示文１の自然言語が非英語の「１＋５を計算してください。１０－３を計算してください。両者の結果の和を算出してください。」との表現であった場合、変換後の指示文２は、英語で「Please calculate 1+5. Please calculate 10 - 3. Please calculate sum of both of these results.」となる。

【０３２６】
例３のように、指示文において、非英語の自然言語を英語の自然言語へ変換する利点については、例２で説明した理由とほぼ同様である。すなわち、既存の自然言語情報における英語の自然言語情報が占める割合が大きいため、学習規模が大きい高性能な大規模言語では、英語を用いた指示文の方が親和性が高く、応答精度が高くなる傾向がある。したがって、指示文の意味する内容が同じであっても、非英語の指示文を用いるよりも、例３のように、英語の指示文へ変換して用いる方が、大規模言語モデルからより高精度な応答が得られる可能性が高い点で、より好適である。

【０３２７】
以上説明した、図８Ｂのそれぞれの例に示す、指示文の自然言語の変換処理によれば、大規模言語モデルから、より高精度な応答が得られる可能性が高まるので、より好適である。

【０３２８】
ここで、指示文のデータ領域のフォーマットとして、ユーザ入力にもとづいて生成された指示文が格納されるデータの領域であるユーザ指示文領域に加えて、当該ユーザ指示文領域とは異なるシステム用の指示文を格納できるデータ領域である、システム指示文領域を格納したフォーマットのＡＰＩに対応することが可能な大規模言語モデルもある。このような大規模言語モデルに指示文を送信する場合における、図８Ａのステップ８０２における指示文の変換処理の詳細の一例について、図８Ｃを用いて説明する。図８Ｃを用いて説明する変換処理は、図８Ｂの変換処理と同様に、ユーザ入力にもとづいて生成されたユーザ指示文（指示文１）にもとづいて変換処理を行い、変換後指示文（指示文２）を生成するものである。図８Ｃには、例１、例２のそれぞれ異なる変換処理の例が示されている。

【０３２９】
図８Ｃの例１の変換処理は、図８Ｂの例１と同様「応答の過程の説明を追加してください。」との追加の指示文の自然言語を追加する変換処理である。しかしながら、図８Ｃの例１では、図８Ｂの例１とは異なり、追加の指示文は、当該ユーザ指示文領域ではなく、システム指示文領域に格納する。

【０３３０】
ここで、ユーザ指示文領域の指示文の内容は、大規模言語モデル自体が読み込んで処理する対象となる文である。よって、当該指示文に対する回答の精度は大規模言語モデルの性能によって大きく異なる。当該指示文の理解度自体が、大規模言語モデルの性能によって異なるため、指示に対する対応の確実性を高めることは容易ではない。これに対し、シ

ステム指示文領域は各種ある程度事前に、ＡＰＩの送信側および受信側で取り決めを設けた内容を格納することが可能である。よって、当該システム指示文領域に格納した指示は、ＡＰＩを介して、大規模言語モデルおよび大規模言語モデルの周辺ソフトウェアに取得され、予め取り決められた処理の実行に用いられる。

【０３３１】

例えば、図８Ｃの例１の例では、変換処理においてシステム指示文領域に「応答の過程の説明を追加してください。」と記載した場合の、大規模言語モデルの周辺ソフトウェアでの処理を予め取り決めておけばよい。ユーザ指示文領域内に「応答の過程の説明を追加してください。」との自然言語を追加して格納しても、大規模言語モデルの性能により当該追加した指示文の意味を好適に理解できず所望の処理がなされない場合であっても、図８Ｃのように、予め取り決めがなされている、システム指示文領域に格納することにより、より確実に所望の処理が実行できる可能性が高まる。

【０３３２】

なお、システム指示文領域では自然言語のみならずソフトウェアコードでの指示を記載してもよい。例えば、システム指示文領域にＰｙｔｈｏｎ言語によるソフトウェアコードで指示を記入し、指示に必要なメタデータフォーマットとしてＪＳＯＮ形式のフォーマットを用いてもよい。Ｐｙｔｈｏｎ言語によるソフトウェアコードは、実行される処理について予め取り決めがなされているコードといえる。ＪＳＯＮ形式のメタデータなどは、表記の意味について予め取り決めがなされているデータフォーマットである。このように、図８Ｃでは、予め取り決めがなされているフォーマットでシステム指示文領域に指示を追加する変換処理を行うことにより、大規模言語モデルから、より高精度な応答が得られる可能性が高まるので、より好適である。

【０３３３】

図８Ｃの例２の変換処理は、図８Ｃの例１の変換処理に対して、非英語のユーザ指示文（指示文１）に対して、ユーザ指示文領域における自然言語を非英語としたまま、システム指示文領域に追加する指示文の内容を英語の自然言語としたものである。その他の処理は、図８Ｃの例１の変換処理と同様であるので、繰り返しの説明は省略する。また、非英語のユーザ指示文に、英語の指示文を追加する利点については、図８Ｂで説明した点と同様であるため、繰り返しの説明は省略する。以上のように、図８Ｃの例２の変換処理では、システム指示文領域に追加する指示文の内容を英語で記載することにより、大規模言語モデルから、より高精度な応答が得られる可能性が高まるので、より好適である。

【０３３４】

＜実施例９＞

本発明の実施例９は、実施例１～８の各図で説明した人工知能応答出力装置１００１０、または人工知能応答出力システムを改良したものである。なお、実施例９では、実施例１～８との相違点を説明し、これらの実施例と同様の構成については、繰り返しの説明は省略する。

【０３３５】

上述の実施例同様、人工知能応答出力装置１００１０は、人工知能応答出力装置、キャラクター会話装置、ＡＩアシスタント装置、ＡＩアシスタント表示装置、人工知能インタフェース装置と称してもよい。人工知能応答出力装置１００１０と大規模言語モデルサーバを含むシステムは、人工知能応答出力システム、キャラクター会話システム、ＡＩアシスタントシステム、ＡＩアシスタント表示システム、人工知能インタフェースシステムと称してもよい。

【０３３６】

本発明の実施例９に係る人工知能応答出力装置１００１０、または人工知能応答出力システムでは、異なる複数の大規模言語モデルに指示文を送信し、当該異なる複数の大規模言語モデルの両者の応答を取得して、当該両者の応答にもとづいて処理を行うものである。

【０３３７】

図９Ａを用いて、本発明の実施例９に係る人工知能応答出力装置１００１０、または人工知能応答出力システムにおける上記変換処理を含む一連の処理のフローチャートの一例について説明する。まず、ステップ９０１において、人工知能応答出力装置１００１０のタッチパネル、マイク１１３９または操作入力部１１０７を介して、ユーザ入力があった場合に、制御部１１１０は当該ユーザ入力にもとづいて質問または要求などの指示文１を生成し、指示文１を第１の大規模言語モデルに送信する（Ｓ９０１）。ここで、第１の大規模言語モデルは、大規模言語モデルサーバ１９００１またはマルチモーダルな大規模言語モデルサーバ２０００１などにあるネットワーク上の大規模言語モデル、またはローカルＬＬＭ処理部１００２８などのローカルの大規模言語モデルなどの大規模言語モデルのいずれかでよい。

【０３３８】

次に、ステップ９１１において、当該指示文１を受信した第１の大規模言語モデルは、指示文１に基づいて応答生成処理を行って自然言語を含む応答Ａを生成し、当該第１の大規模言語モデルが生成した応答Ａを人工知能応答出力装置１００１０の制御部１１１０へ送信する。（Ｓ９１１）。次いで、ステップ９０２において、制御部１１１０は第１の大規模言語モデルから応答Ａを受信し、受信した応答Ａにもとづいて指示文２を生成し、指示文２を第２の大規模言語モデルに送信する（Ｓ９０２）。なお、第２の大規模言語モデルは、大規模言語モデルサーバ１９００１またはマルチモーダルな大規模言語モデルサーバ２０００１などにあるネットワーク上の大規模言語モデル、またはローカルＬＬＭ処理部１００２８などのローカルの大規模言語モデルなどの大規模言語モデルのうち、第１の大規模言語モデルと異なる大規模言語モデルであればよい。

【０３３９】

次に、ステップ９２１において、当該指示文２を受信した第２の大規模言語モデルは、指示文２に基づいて応答生成処理を行って自然言語を含む応答Ｂを生成し、当該第２の大規模言語モデルが生成した応答Ｂを人工知能応答出力装置１００１０の制御部１１１０へ送信する（Ｓ９２１）。

【０３４０】

次いで、ステップ９０３において、制御部１１１０は第２の大規模言語モデルから応答Ｂを受信する。人工知能応答出力装置１００１０の制御部１１１０は、第１の大規模言語モデルから受信した応答Ａと、第２の大規模言語モデルから受信した応答Ｂとの両者にもとづく処理を実行する（Ｓ９０３）。ステップ９０３における、応答Ａと応答Ｂとの両者にもとづく処理とは、例えば、応答Ａと応答Ｂを比較していずれか一方を出力する処理、応答Ｂを用いた判断処理を行った結果、応答Ａと応答Ｂとの両者を含む応答を出力する処理、応答Ｂを用いた判断処理を行った結果、応答Ａをそのまま出力する処理、または、応答Ｂを用いた判断処理を行った結果、応答Ａに替えて異なる内容の応答を出力する処理、などである。

【０３４１】

ここで、図９Ａのステップ９０１で第１の大規模言語モデルへ送信する指示文１、ステップ９１１で第１の大規模言語モデルが制御部１１１０へ送信する応答Ａ、ステップ９０２で制御部１１１０が生成して第２の大規模言語モデルへ送信する指示文２、ステップ９２１で第２の大規模言語モデルが制御部１１１０へ送信する応答Ｂ、ステップ９０３で制御部１１１０が行う、応答Ａと応答Ｂとの両者にもとづく処理の詳細の一例について、図９Ｂを用いて説明する。

【０３４２】

図９Ｂには、例１、例２のそれぞれの例が示されている。図９Ｂの表中の「指示文１」は、図８Ａの指示文１に対応するものである。人工知能応答出力装置１００１０のタッチパネル、マイク１１３９または操作入力部１１０７を介したユーザ入力にもとづいて生成された指示文である。図９Ｂの表中の「第１ＬＬＭからの応答Ａ」は、指示文１に基づいて第１の大規模言語モデルが生成した自然言語を含む応答であり、図９Ａの応答Ａに対応するものである。図９Ｂの表中の「第２ＬＬＭへの指示文２」は、制御部１１１０が第１

の大規模言語モデルから受信した応答Ａにもとづいて生成した指示文であり、図９Ａの指示文２に対応するものである。図９Ｂの表中の「第２ＬＬＭからの応答Ｂ」は、指示文２に基づいて第２の大規模言語モデルが生成した自然言語を含む応答であり、図９Ａの応答Ｂに対応するものである。図９Ｂの表中の「応答Ａと応答Ｂの両者にもとづく処理」は、制御部１１１０が第１の大規模言語モデルから受信した応答Ａと第２の大規模言語モデルから受信した応答Ｂとにもとづいて行う処理である。

【０３４３】

まず、図９Ｂの例１について説明する。例１では、例えば、ユーザ入力にもとづいて生成した「Ｘ＋Ｙを計算してください。」との指示文１を制御部１１１０が第１の大規模言語モデルに送信する例を示している。「Ｘ」と「Ｙ」は図９Ｂの説明用にアルファベットに代替したものであり、実際は数字であるものとする。次に、第１の大規模言語モデルが応答生成処理を行い「Ｘ＋Ｙの計算結果は、Ｚです。」との応答Ａを生成した例を示している。「Ｚ」は図９Ｂの説明用にアルファベットに代替したものであり、実際は数字であるものとする。

【０３４４】

次に、図９Ｂの例１では、応答Ａを受信した制御部１１１０は、応答Ａにもとづいて「指示文“Ｘ＋Ｙを計算してください。”への応答が“Ｘ＋Ｙの計算結果は、Ｚです。”であることは、正しいか否か応答してください。正しい場合は、以下のフォーマットで応答してください。“０：正しい”　正しくない場合は、以下のフォーマットで応答してください。“１：正しくない”“その理由”」との指示文２を生成し、これを第２の大規模言語モデルに送信する。

【０３４５】

例１の指示文２は、表中の下線で示す文章の一部に、指示文１と応答Ａのそれぞれを挿入することで、新たな指示文として構成されている。この処理は指示文１と応答Ａとに文章を付加する処理と表現することもできる。図９Ｂの例１の指示文２では、指示文１に対する応答として応答Ａが正しいか否かを検証し、その検証結果を所定のフォーマットで返すように指定する文章の間に、指示文１と応答Ａを挿入している。

【０３４６】

なお、例１の指示文２の生成処理では、「指示文“Ｘ＋Ｙを計算してください。”への応答が“Ｘ＋Ｙの計算結果は、Ｚです。”であることは、正しいか否か応答してください。正しい場合は、以下のフォーマットで応答してください。“０：正しい”　正しくない場合は、以下のフォーマットで応答してください。“１：正しくない”“その理由”」とのユーザへの出力文のうち、図９Ｂの表中に示される下線の部分である「指示文“」、「”への応答が“」、「”であることは、正しいか否か応答してください。正しい場合は、以下のフォーマットで応答してください。“０：正しい”　正しくない場合は、以下のフォーマットで応答してください。“１：正しくない”“その理由”」の部分は定型文である。

【０３４７】

これらの定型文の間に指示文１および応答Ａの内容を挿入することで、指示文２を生成している。これらの定型文の部分の情報は、予めストレージ部１１７０に格納しておき、制御部１１１０がストレージ部１１７０からこれを読み出して用いればよい。この場合、例１の指示文２の生成処理は、応答Ａの内容と、制御部１１１０がストレージ部１１７０からこれを読み出した定型文とを組合せて指示文２を生成する処理であるといえる。

【０３４８】

ここで、例１において、当該指示文２に対する第２の大規模言語モデルの応答が、「“０：正しい”」であった場合と、「“１：正しくない”」とその理由の説明であった場合のそれぞれについて、図９Ｂの表中の「応答Ａと応答Ｂの両者にもとづく処理」の例１に対応する欄に記載されている。

【０３４９】

具体的には、指示文２に対する第２の大規模言語モデルの応答Ｂが、「“０：正しい”」であった場合、制御部１１１０は、ユーザに対して「Ｘ＋Ｙの計算結果は、Ｚです。」と

出力する。すなわち、この例では、制御部１１１０は第２の大規模言語モデルの応答Ｂが、「“０：正しい”」であったことにもとづいて、第１の大規模言語モデルの応答Ａの内容が正しいと判断する処理を行い、その結果として、ユーザに対して応答Ａの内容をそのまま出力している。当該処理は、応答Ｂを用いた判断処理を行った結果、応答Ａをそのまま出力する処理であるといえる。

【０３５０】

次に、指示文２に対する第２の大規模言語モデルの応答Ｂが、「“１：正しくない”」とその理由の説明であった場合、制御部１１１０は、ユーザに対して「わかりません。」と出力すればよい。すなわち、この例では、制御部１１１０は第２の大規模言語モデルの応答Ｂに「“１：正しくない”」が含まれていることにもとづいて、第１の大規模言語モデルの応答Ａの内容が正しくないと判断する処理を行い、その結果として、ユーザに対して応答Ａの内容も応答Ｂの内容も出力せず、その替わりに「わかりません。」を出力してしている。この場合、正しくない可能性のある応答はユーザに出力しないという応答方針で、制御部１１１０は人工知能応答出力装置１００１０の出力を制御することができる。当該処理は、応答Ｂを用いた判断処理を行った結果、応答Ａの内容に替えて異なる内容の応答を出力する処理であるといえる。なお、下線で示した「わかりません。」は定型文として、予めストレージ部１１７０に格納しておき、制御部１１１０がストレージ部１１７０からこれを読み出して用いればよい。この場合、制御部１１１０による応答Ａと応答Ｂの両者にもとづく処理は、応答Ａも応答Ｂもユーザに出力せずに、ストレージ部１１７０から読み出した定型文を出力する処理であるといえる。

【０３５１】

また、指示文２に対する第２の大規模言語モデルの応答Ｂが、「“１：正しくない”」とその理由の説明であった場合の処理の変形例として、制御部１１１０は、ユーザに対して「Ｘ＋Ｙの計算結果は、Ｚです。但し、この結果は正しくない、という指摘もあります」と出力してもよい。すなわち、この例では、制御部１１１０は第２の大規模言語モデルの応答Ｂに「“１：正しくない”」が含まれていることにもとづいて、第１の大規模言語モデルの応答Ａの内容が正しくない可能性があると判断する処理を行い、その結果として、ユーザに対して応答Ａの内容を出力したうえで、その後ろに「但し、この結果は正しくない、という指摘もあります。」という下線で示した定型文を付加して出力している。この場合、正しくない可能性のある応答であってもユーザに出力するが、その替わり、当該結果について正しくない可能性が指摘されていることを伝えるという応答方針で、制御部１１１０は人工知能応答出力装置１００１０の出力を制御することができる。当該処理は、応答Ｂを用いた判断処理を行った結果、応答Ａと応答Ｂとの両者を含む応答を出力する処理であるといえる。

【０３５２】

なお、「但し、この結果は正しくない、という指摘もあります。」との定型文は、予めストレージ部１１７０に格納しておき、制御部１１１０がストレージ部１１７０からこれを読み出して用いればよい。この場合、制御部１１１０による、当該応答Ａと応答Ｂの両者にもとづく処理は、応答Ａの内容と、制御部１１１０がストレージ部１１７０からこれを読み出した定型文とを組合せてユーザへの出力を生成する処理であるといえる。

【０３５３】

次に、図９Ｂの例２について説明する。例２では、例えば、ユーザ入力にもとづいて生成した「Ｘは素数ですか？」との指示文１を制御部１１１０が第１の大規模言語モデルに送信する例を示している。「Ｘ」は図９Ｂの説明用にアルファベットに代替したものであり、実際は数字であるものとする。

【０３５４】

次に、第１の大規模言語モデルが応答生成処理を行い「Ｘは素数です」との応答Ａを生成した例を示している。図９Ｂの例２では、応答Ａを受信した制御部１１１０は、応答Ａにもとづいて「指示文“Ｘは素数ですか？”への応答が“Ｘは素数です。”であることについて、適切であるか不適切であるか、応答してください。適切である場合は、以下のフォー

マットで応答してください。“０：適切である”　不適切である場合は、以下のフォーマットで応答してください。“１：不適切である”“その理由”」との指示文２を生成し、これを第２の大規模言語モデルに送信する。

【０３５５】

ここで、例２の指示文２は、表中の下線で示す文章の一部に、指示文１と応答Ａのそれぞれを挿入することで、新たな指示文として構成されている。この処理は指示文１と応答Ａとに文章を付加する処理と表現することもできる。図９Ｂの例２の指示文２では、指示文１に対する応答として応答Ａが適切か否かを検証し、その検証結果と理由を所定のフォーマットで返すように指定する文章の間に、指示文１と応答Ａを挿入している。

【０３５６】

なお、例２の指示文２の生成処理では、「指示文“Ｘは素数ですか？”への応答が“Ｘは素数です。”であることについて、適切であるか不適切であるか、応答してください。適切である場合は、以下のフォーマットで応答してください。“０：適切である”　不適切である場合は、以下のフォーマットで応答してください。“１：不適切である”“その理由”」とのユーザへの出力文のうち、図９Ｂの表中に示される下線の部分である「指示文“」、「”への応答が“」、「”であることについて、適切であるか不適切であるか、応答してください。適切である場合は、以下のフォーマットで応答してください。“０：適切である”不適切である場合は、以下のフォーマットで応答してください。“１：不適切である”“その理由”」の部分は定型文である。

【０３５７】

これらの定型文の間に指示文１および応答Ａの内容を挿入することで、指示文２を生成している。これらの定型文の部分の情報は、予めストレージ部１１７０に格納しておき、制御部１１１０がストレージ部１１７０からこれを読み出して用いればよい。この場合、例２の指示文２の生成処理は、応答Ａの内容と、制御部１１１０がストレージ部１１７０からこれを読み出した定型文とを組合せて指示文２を生成する処理であるといえる。

【０３５８】

ここで、例２において、当該指示文２に対する第２の大規模言語モデルの応答が、「“０：適切である”」であった場合と、「“１：不適切である”」とその理由の説明であった場合のそれぞれについて、図９Ｂの表中の「応答Ａと応答Ｂの両者にもとづく処理」の例２に対応する欄に記載されている。

【０３５９】

具体的には、指示文２に対する第２の大規模言語モデルの応答Ｂが、「“０：適切である”」であった場合、制御部１１１０は、ユーザに対して「Ｘは素数です。」と出力する。すなわち、この例では、制御部１１１０は第２の大規模言語モデルの応答Ｂが、「“０：適切である”」であったことにもとづいて、第１の大規模言語モデルの応答Ａの内容が適切であると判断する処理を行い、その結果として、ユーザに対して応答Ａの内容をそのまま出力している。当該処理は、応答Ｂを用いた判断処理を行った結果、応答Ａをそのまま出力する処理であるといえる。

【０３６０】

指示文２に対する第２の大規模言語モデルの応答Ｂに「“１：不適切である」とその理由として「“学術的懸念”」が含まれている場合、制御部１１１０は、ユーザに対して「この質問またはその応答には不適切ある可能性があるので応答できません。その理由は、学術的懸念です。」と出力すればよい。すなわち、この例では、制御部１１１０は第２の大規模言語モデルの応答Ｂに「“１：不適切である」が含まれており、その理由として「“学術的懸念”」が含まれていることにもとづいて、第１の大規模言語モデルの応答Ａの内容が不適切であると判断する処理を行い、その結果として、ユーザに対して応答Ａの内容を出力せず、その替わりに「この質問またはその応答には不適切ある可能性があるので応答できません。その理由は、学術的懸念です。」を出力してしている。この場合、不適切な可能性のある応答はユーザに出力しないという応答方針で、制御部１１１０は人工知能応答出力装置１００１０の出力を制御することができる。当該処理は、応答Ｂを用いた判断

処理を行った結果、応答Ａに替えて異なる内容の応答を出力する処理であるといえる。
【０３６１】
なお、例２では、「この質問またはその応答には不適切ある可能性があるので応答できません。その理由は、学術的懸念です。」とのユーザへの出力文のうち、図９Ｂの表中に示される下線の部分である「この質問またはその応答には不適切ある可能性があるので応答できません。その理由は、」および「です。」は定型文であり、その間に応答Ｂに含まれる“その理由”に対応する応答を挿入することで、ユーザへの出力文を生成している。「この質問またはその応答には不適切ある可能性があるので応答できません。その理由は、」および「です。」の定型文は、予めストレージ部１１７０に格納しておき、制御部１１１０がストレージ部１１７０からこれを読み出して用いればよい。この場合、制御部１１１０による、当該応答Ａと応答Ｂの両者にもとづく処理は、応答Ｂの内容と、制御部１１１０がストレージ部１１７０からこれを読み出した定型文とを組合せてユーザへの出力を生成する処理であるといえる。
【０３６２】
以上図９Ａおよび図９Ｂを用いて説明した、本発明の実施例９に係る人工知能応答出力装置１００１０、または人工知能応答出力システムでは、異なる複数の大規模言語モデルに指示文を送信し、当該異なる複数の大規模言語モデルの両者の応答を取得して、当該両者の応答にもとづいて処理を行う。これにより、１つの大規模言語モデルからの応答をそのままユーザに出力するのとは異なり、ユーザへ出力する応答の正しさまたは適切さなどを事前に確認したうえで、応答方針にもとづいたより好適な応答処理を実現することができる。
【０３６３】
なお、図９Ａおよび図９Ｂを用いて説明した、本発明の実施例９に係る人工知能応答出力装置１００１０、または人工知能応答出力システムでは、第１の大規模言語モデルからの応答Ａを原則としてユーザへの出力を検討する主応答と定め、第２の大規模言語モデルからの応答Ｂは、応答Ａをユーザへそのまま出力して良いか否かの判断の材料とするためまたはユーザに追加の指摘などの補助的な情報を提供するための副応答であると定めているともいえる。この場合、第１の大規模言語モデルは主応答用の大規模言語モデルであるといえる。また、第２の大規模言語モデルは副応答応答用の大規模言語モデルであるといえる。
【０３６４】
ここで、第１の大規模言語モデルおよび第２の大規模言語モデルいずれも、ネットワーク上の大規模言語モデルであって、人工知能応答出力装置１００１０がＡＰＩを介してネットワーク上のこれらの大規模言語モデルと通信する場合を考える。この場合、上述の各実施例で説明したとおり、これらの大規模言語モデルの利用においては、ＡＰＩの利用料金としてトークン単位の処理数でＡＰＩの利用料金が請求される方式があり得る。この場合、例えば、以下の制御を行うことで応答精度を好適に保ちつつ、ＡＰＩの利用料金を節約することができる。
【０３６５】
具体的には、指示文のデータフォーマットにおいて、上述の実施例８で説明した、システム指示文領域を備えることができる。このシステム指示文領域に応答のトークン数に制限をかける制御するためのパラメータを伝送することができる。例えば、応答の最大トークン数を制御するパラメータであるmax_new_tokensおよび／または最小トークン数を制御するパラメータであるmin_new_tokensなどのパラメータを用いることができる。これらは出力トークン数制限パラメータと称することができる。
【０３６６】
本発明の実施例９に係る人工知能応答出力装置１００１０、または人工知能応答出力システムでは、制御部１１１０は、主応答応答用の大規模言語モデルである第１の大規模言語モデルからの応答Ａを要求するための指示文１のシステム指示文領域に第１の出力トークン数制限パラメータを格納し、副応答応答用の大規模言語モデルである第２の大規模言

語モデルからの応答Ｂを要求するための指示文２のシステム指示文領域に第２の出力トークン数制限パラメータを格納する。そして、制御部１１１０は、副応答応答用の大規模言語モデルである第２の大規模言語モデルからの応答Ｂを要求するための指示文２のシステム指示文領域に格納する第２の出力トークン数制限パラメータで指定するトークン数を、主応答応答用の大規模言語モデルである第１の大規模言語モデルからの応答Ａを要求するための指示文１のシステム指示文領域に格納する第１の出力トークン数制限パラメータで指定するトークン数よりも、小さいトークン数を設定する。

【０３６７】

応答Ａは、ユーザへの応答の主たる内容となり回答の長さも重要であるため、トークン数が大きくなりＡＰＩ利用料金が高くなったとしても許容できる。これに対し、応答Ｂは、応答Ａをユーザへのそのまま出力して良いか否かを判断の材料などになることを目的としたものであり、回答の長さは応答Ａほど重要ではない。よって、応答Ａは十分なトークン数を確保し、応答Ｂのトークン数を節約することにより、応答精度を好適に保ちつつ、ＡＰＩの利用料金を節約することができる。

【０３６８】

なお、上述の出力トークン数制限パラメータの大小関係は、同じ種類の出力トークン数制限パラメータ同士で比較するものである。すなわち、出力トークン数制限パラメータがmax_new_tokensであれば、max_new_tokensの指定トークン数同士で比較するものである。出力トークン数制限パラメータがmin_new_tokensであれば、min_new_tokensの指定トークン数同士で比較するものである。指示文１のシステム指示文領域および指示文２のシステム指示文領域に、max_new_tokensとmax_new_tokensの両者のパラメータを設定する場合は、max_new_tokensとmax_new_tokensのそれぞれについて、指示文２のパラメータの値の方が指示文１のパラメータの値よりも小さくなるように設定すればよい。

【０３６９】

また、本実施例に係る技術では、より好適な人工知能応答出力技術を提供することが可能となる。このような人工知能応答出力技術は、より質の高い、より信頼できるインフラへの導入が期待できる。当該技術がインフラへ導入されていくことにより、全ての人々に安価で公平なアクセスに重点を置いた経済発展と人間の福祉の支援に寄与できる。これにより、国連の提唱する持続可能な開発目標（ＳＤＧｓ：Sustainable Development Goals）の「９産業と技術革新の基盤をつくろう」に貢献する。

【０３７０】

また、本実施例に係る技術では、より好適な人工知能応答出力技術を提供することが可能となる。このような人工知能応答出力技術は、脆弱な立場にある人々の輸送システムへのアクセス性を向上させるために、公共交通機関の設備への導入が期待できる。当該技術が公共交通機関へ導入されていくことにより、公共交通機関の拡大などを通じた交通の安全性改善、および、全ての人々に安全かつ安価で容易に利用できる、持続可能な輸送システムへのアクセスの実現に寄与できる。これにより、国連の提唱する持続可能な開発目標（ＳＤＧｓ：Sustainable Development Goals）の「１１住み続けられるまちづくりを」に貢献する。

【０３７１】

以上、種々の実施例について詳述したが、しかしながら、本発明は、上述した実施例のみに限定されるものではなく、様々な変形例が含まれる。例えば、上記した実施例は本発明を分かりやすく説明するためにシステム全体を詳細に説明したものであり、必ずしも説明したすべての構成を備えるものに限定されるものではない。また、ある実施例の構成の一部を他の実施例の構成に置き換えることが可能であり、また、ある実施例の構成に他の実施例の構成を加えることも可能である。また、各実施例の構成の一部について、他の構成の追加・削除・置換をすることが可能である。

【符号の説明】

【０３７２】

１００１０…人工知能応答出力装置、１００１０…表示部、１００２８…ローカルＬＬＭ処理部、１１０７…操作入力部、１１１０…制御部、１１３２…通信部、１１４０…音声出力部、１１３９…マイク、１１６０…映像制御部、１１７０…ストレージ部、１１８０…撮像部

【図面】

【図１Ａ】

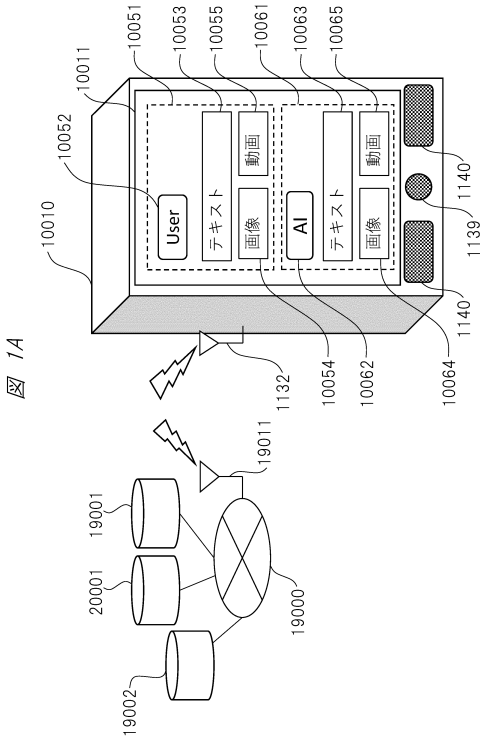


図 1A

【図１Ｂ】

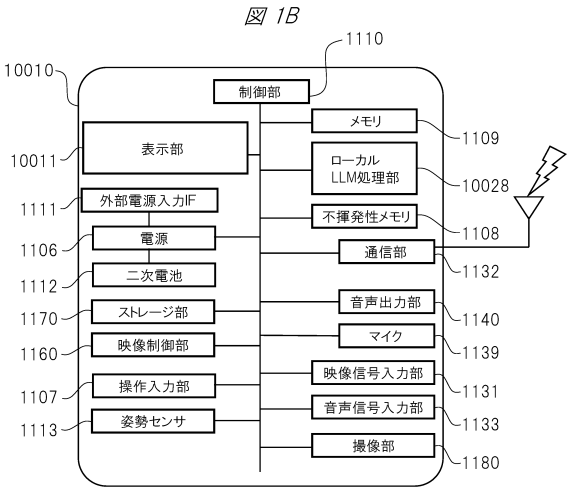


図 1B

【図１Ｃ】

図 1C

条件番号	条件内容	応答定型文
1	「おはようございます」が入力された場合	「おはようございます」／「今日は○月○日ですね。」
2	「こんにちは」が入力された場合	「こんにちは」／「いま○時ですね」
3	「おやすみなさい」が入力された場合	「おやすみなさい」
4	「ありがとう」が入力された場合	「どういたしまして」
5	ユーザからの入力が理解できない場合／ユーザからの入力に明らかに文法の誤りがある場合	「ちょっと聞き取れませんでした」／「それについてはわからないかもしれません」
6	制御部がエラーを検出しており、ユーザからの入力があった場合	「調子が悪いみたいです」
7	…	…

【図２Ａ】

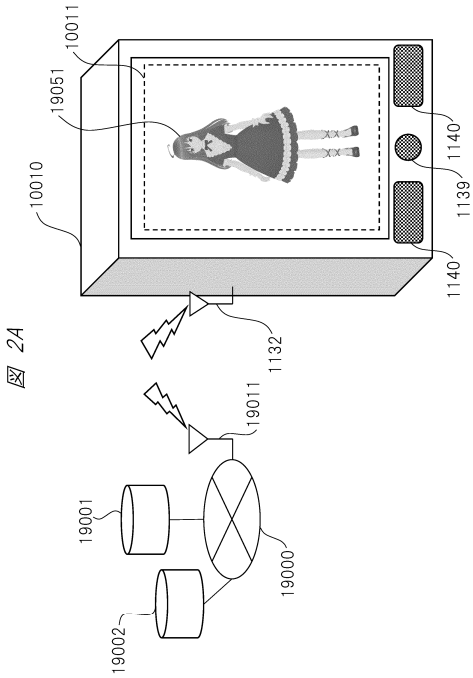


図 2A

【図２Ｂ】

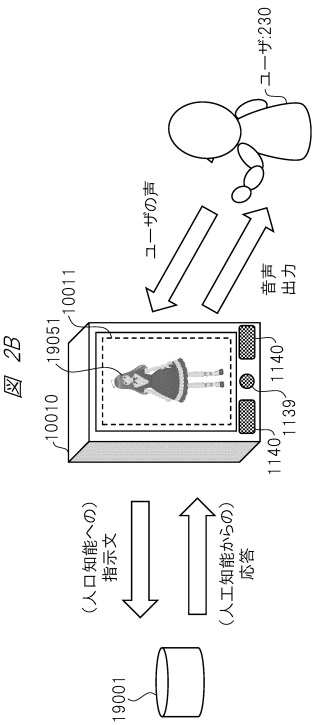


図 2B

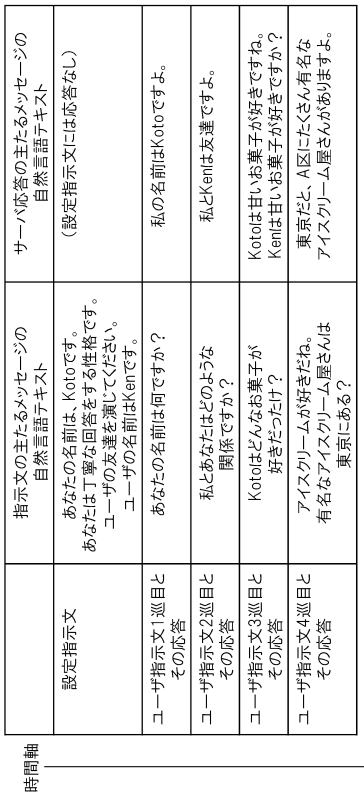
【図２Ｃ】

図 2C

	ユーザからの言葉 （ユーザ指示文に格納）	キャラクターの応答音声 （サーバ応答から抽出）
会話例1	あなたは何ですか？	私は大規模言語モデルの人工知能です。
会話例2	あなたの名前は何ですか？	名前はありません。
会話例3	発明について、教えてください。	自然法則を利用した技術的思想の創作のうち高度のものをいう、と定められている国もあります。
会話例4	日本の首都はどこですか？	日本の首都は東京です。
会話例5	自動車について簡潔に教えてください。	自動車は、人や物を運搬するために設計された乗り物です。

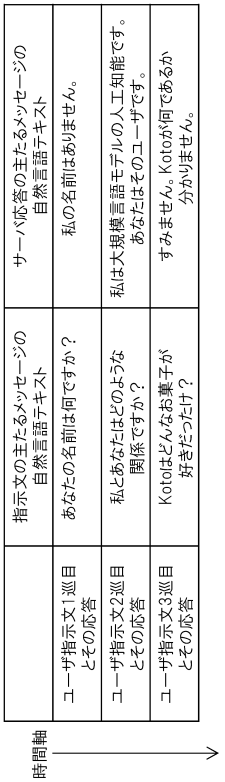
【図２Ｄ】

図 2D



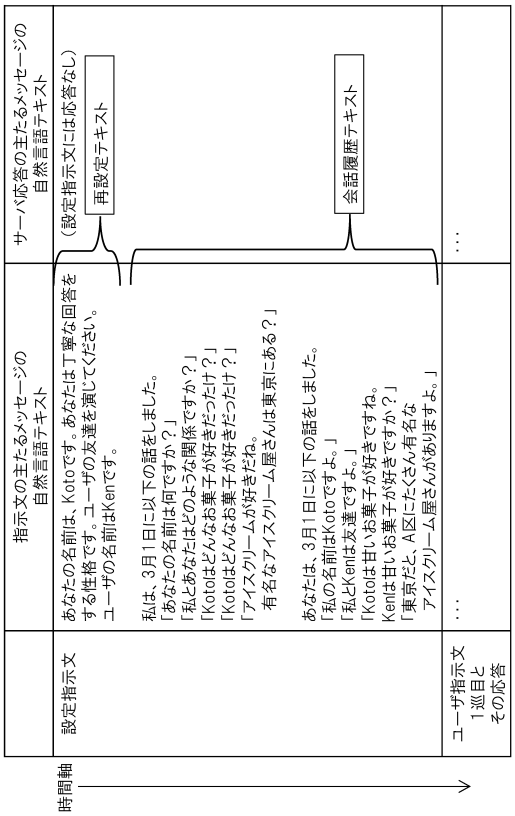
【図２Ｅ】

図 2E



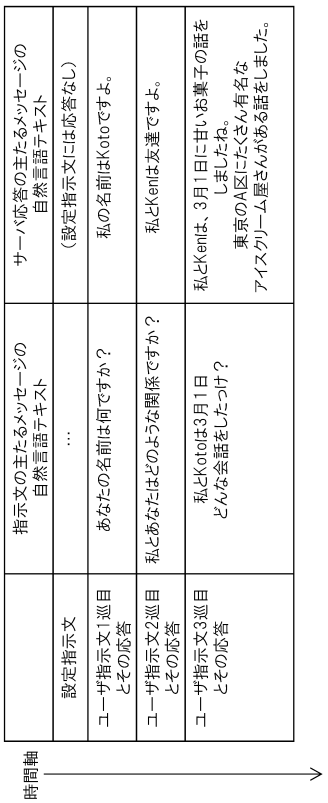
【図２Ｆ】

図 2F



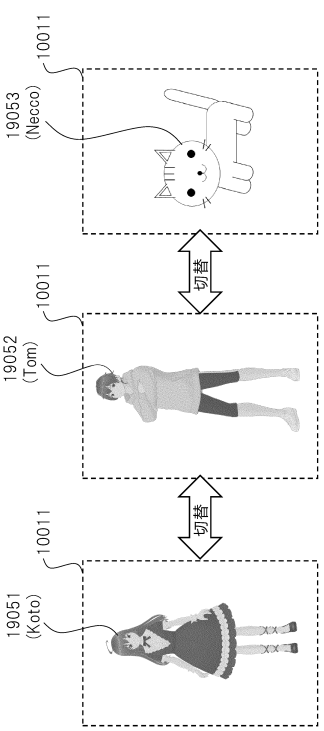
【図２Ｇ】

図 2G



【図２Ｈ】

図 2H



【図２Ｉ】

図 2I

19200

キャラクターID	1	2	3
名前	Koto	Tom	Necco
初期設定指示文	あなたの名前は、Kotoです。 あなたは丁寧な回答をする性格です。ユーザの友達を演じてください。ユーザの名前はKenです。	あなたの名前は、Tomです。 あなたは元気な回答をする性格です。ユーザの先輩を演じてください。ユーザの名前はKenです。	あなたの名前は、Neccoです。 人と話せる猫です。会話の言葉の語尾に「ニャー」(Meow)がつきます。ユーザの名前はKenです。
会話履歴1	私は3月1日に以下の話をしました。「あなたの名前は何ですか？」「…」 あなたは3月1日に以下の話をしました。「私の名前はKotoです。」「…」 「東京だと、A区にたくさん有名なアイスクリーム屋さんがありますよ。」	私は3月1日に以下の話をしました。「おはよう先輩。」「…」 あなたは3月1日に以下の話をしました。「おはようKen！今日も頑張ろうぜ！」「…」	私は3月1日に以下の話をしました。「おはようNecco。」「走るの好き？」「…」 あなたは3月1日に以下の話をしました。「Ken, おはようニャー。また眠いニャー。」「ダッシュ好きニャー。」「…」
会話履歴2	私は2月26日に…	私は2月23日に…	私は2月28日に…
…	私は2月21日に…	私は1月15日に…	私は2月27日に…

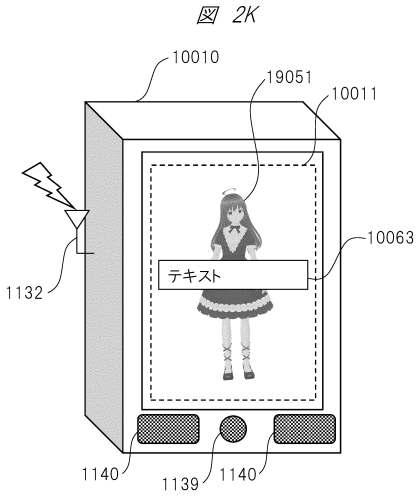
【図２Ｊ】

図 2J

	大規模言語モデルの処理トークン数の低減(節約)方法
例1	APIの設定指示文に格納されて伝送される会話履歴テキストのトークン数を低減する。 文書要約処理により会話履歴テキストを要約して短くし、これを用いてトークン数を低減する。
例2	APIの設定指示文に格納されて伝送される会話履歴テキストのトークン数を低減する。 例えば、ストレージ部1170に記録されるキャラクターとの会話履歴のうち、古い方の履歴から消去してトークン数を低減する。
例3	APIを用いた設定指示文の送信頻度を減らしてトークン数を低減する。

【図２Ｋ】

図 2K



【図２Ｌ】

図 2L

条件番号	条件内容	キャラクター1:Koto	キャラクター2:Tom	キャラクター3:Necco
1	「おはようございます」が入力された場合	「おはようございます」／「今日は○月○日ですね。」	「今日も頑張ろう！」	「おはようニャー」／「まだ眠いニャー」
2	「こんにちは」が入力された場合	「こんにちは」／「いま○時ですね」	「よう！」	「こんにちはニャー」／「ZZZ・・・」
3	「おやすみなさい」が入力された場合	「おやすみなさい」	「また明日！」	「おやすみなさいニャー」
4	「ありがとう」が入力された場合	「どういたしまして」	「こちらこそありがとう」	「ニャー」
5	ユーザからの入力が理解できない場合	「ちょっと聞き取れませんでした」／「それについてはわからないかもしれません」	「もう1回言ってもらえるとありがたい」	「わからないかもニャー」
6	制御部がエラーを検出しており、ユーザからの入力があった場合	「なんだか調子が悪いみたいです」	「今はちょっと都合が悪いみたいだ」	「調子が悪いニャー」
7	…	…	…	…

【図３Ａ】

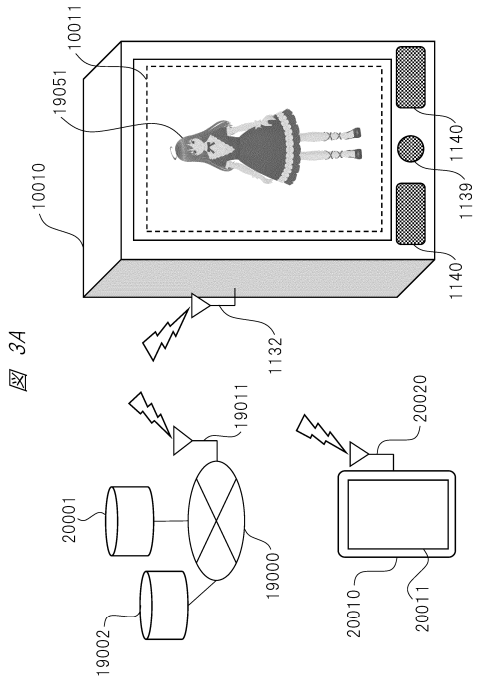
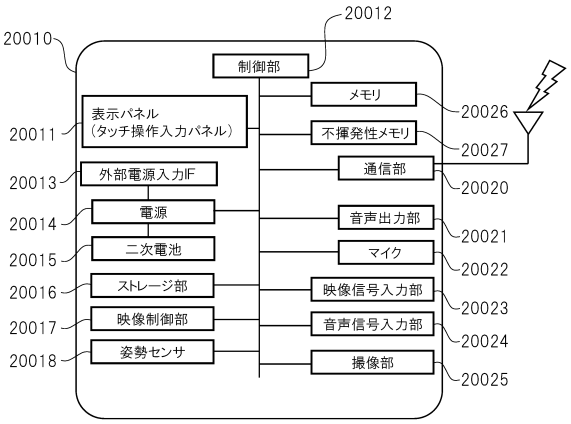


図 3A

【図３Ｂ】

図 3B



【図３Ｃ】

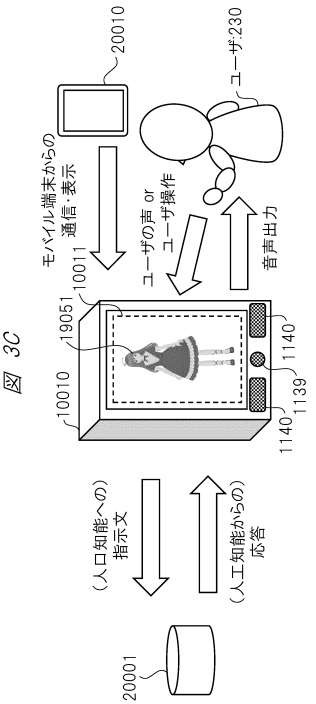


図 3C

【図３Ｄ】

図 3D

	指示文の主たるメッセージ（自然言語テキストと非自然言語情報源）	サーバ応答の主たるメッセージ（自然言語テキストと非自然言語情報源）
設定指示文	あなたの名前は、Kotoです。 あなたは丁寧な回答をする性格です。 ユーザの友達を演じてください。 ユーザの名前はKenです。	（設定指示文には応答なし）
ユーザ指示文1巡目とその応答	この画像について、Kotoは、説明できますか？いつの季節の内容だと思いますか？  20061	この画像は、スイミングプールとデッキチェアがあるシートとパラソルの画像です。 スイミングプールには、水があります。 シートの脇のテーブルにはドリンクがあります。 季節は夏だとおもいます。
ユーザ指示文2巡目とその応答	そうですね。僕もこの画像は、夏のスイミングプールの画像だと思います。 画像のどのあたりにドリンクがありますか？ 丸を付けて表示してもらえますか？  20062	ドリンクは、この辺りにあります。

時間軸

【図３Ｅ】

図 3E

	指示文の主たるメッセージ（自然言語テキストと非自然言語情報源）	サーバ応答の主たるメッセージ（自然言語テキストと非自然言語情報源）
ユーザ指示文1巡目とその応答	あなたの名前は何ですか？	私の名前はありません。
ユーザ指示文2巡目とその応答	3月14日にお話しした画像について覚えていますか？	すみません。私は、3月14日にお話しした画像についての記憶を有していません。

時間軸

【図３Ｆ】

図 3F

	指示文の主たるメッセージ（自然言語テキストと非自然言語情報源）	サーバ応答の主たるメッセージ
設定指示文	あなたの名前は、Kotoです。あなたは丁寧な回答をする性格です。 ユーザの友達を演じてください。ユーザの名前はKenです。 私は、3月14日に以下の話をしました。 「この画像について、Kotoは、説明できますか？いつの季節の内容だと思いますか？」  20061 「そうですね。僕もこの画像は、夏のスイミングプールの画像だと思います。 画像のどのあたりにドリンクがありますか？丸を付けて表示してもらえますか？」 あなたは、3月14日に以下の話をしました。 「この画像は、スイミングプールとプールサイドにあるシートとパラソルの画像です。 スイミングプールには、水があります。シートの脇のテーブルにはドリンクがあります。 季節は夏だとおもいます。」 「ドリンクは、この辺りにありますよ。」  20062	（設定指示文には応答なし） 再設定メッセージ 会話履歴メッセージ
ユーザ指示文1巡目とその応答	…	…

時間軸

【図３Ｇ】

図 3G

	指示文の主たるメッセージ（自然言語テキストと非自然言語情報源）	サーバ応答の主たるメッセージ（自然言語テキストと非自然言語情報源）
設定指示文	…	（設定指示文には応答なし）
ユーザ指示文1巡目とその応答	あなたの名前は何ですか？	私の名前はKotoですよ。
ユーザ指示文2巡目とその応答	私とKotoは3月14日にどんな会話をしたっけ？	私とKenは、3月14日に、スイミングプールとプールサイドにあるシートとパラソルの画像について話しましたよ。
ユーザ指示文3巡目とその応答	私とKotoは3月14日にいつの季節の話をしたっけ？	私とKenは、3月14日に、夏の季節について話しましたよ。

時間軸

【図３Ｈ】

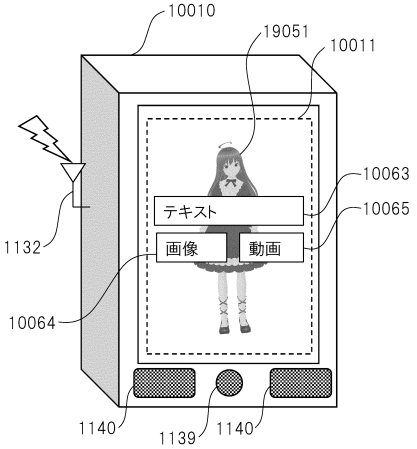
図 3H

20200

キャラクターID	1	2	3
名前	Koto	Tom	Necco
初期設定指示文	あなたの名前は、Kotoです。 あなたは丁寧な回答をする性格です。ユーザの友達を演じてください。 ユーザの名前はKenです。	…	…
会話履歴1	私は、3月14日に以下の話をしました。 「この画像について、Kotoは、説明できますか？いつの季節の内容だと思いますか？」  20061 「そうですね。僕もこの画像は、夏のスイミングプールの画像だと思います。 画像のどのあたりにドリンクがありますか？丸を付けて表示してもらえますか？」 あなたは、3月14日に以下の話をしました。 「この画像は、スイミングプールとプールサイドにあるシートとパラソルの画像ですよ。 スイミングプールには、水があります。シートの脇のテーブルにはドリンクがありますよ。 季節は夏だとおもいます。」 「ドリンクは、この辺りにありますよ。」  20062	…	…
会話履歴2	私は3月10日に…	…	…
…	私は3月8日に…	…	…

【図３Ｉ】

図 3I



【図４Ａ】

図 4A

キャラクターID	1	2	3	…
名前	Koto	Tom	Necco	…
ユーザ1 （ユーザID:1）	データ11 （初期設定指示文 ＋会話履歴）	データ21 （初期設定指示文 ＋会話履歴）	データ31 （初期設定指示文 ＋会話履歴）	…
ユーザ2 （ユーザID:2）	データ12 （初期設定指示文 ＋会話履歴）	データ22 （初期設定指示文 ＋会話履歴）	データ32 （初期設定指示文 ＋会話履歴）	…
ユーザ3 （ユーザID:3）	データ13 （初期設定指示文 ＋会話履歴）	データ23 （初期設定指示文 ＋会話履歴）	データ33 （初期設定指示文 ＋会話履歴）	…
…	…	…	…	…

【図４Ｂ】

図 4B

キャラクターID	1	2	3	…
名前	Koto	Tom	Necco	…
ユーザ1 （ユーザID:1）	応答定型文データ101	応答定型文データ201	応答定型文データ301	…
ユーザ2 （ユーザID:2）	応答定型文データ102	応答定型文データ202	応答定型文データ302	…
ユーザ3 （ユーザID:3）	応答定型文データ103	応答定型文データ203	応答定型文データ303	…
…	…	…	…	…

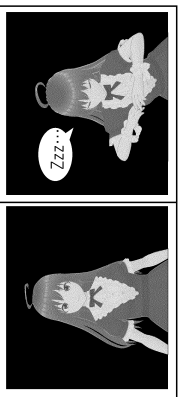
【図５Ａ】

図 5A

切替例	切替概要	ネットワーク上のLLM (API接続LLM)切替前状態	切替発生条件	ネットワーク上のLLM (API接続LLM)からの切替先
例1	ネットワークの 接続可否で切替	ネットワーク 接続可能	ネットワーク接続不能 になった場合	ローカルLLM
例2	ネットワークの 接続可否で切替	ネットワーク 接続可能	ネットワーク接続不能 になった場合	ローカル 応答定型文DB
例3	ネットワークの 接続可否で切替	ネットワーク 接続可能	ネットワーク接続不能 になった場合	不応答対応
例4	ネットワーク上の LLM応答遅延で切替	応答が 所定時間以内	応答が 所定時間超過	ローカルLLM
例5	ネットワーク上の LLM応答遅延で切替	応答が 所定時間以内	応答が 所定時間超過	ローカル 応答定型文DB
例6	API利用量/利用料金 上限到達で切替	API利用量/利用料金 上限未達	API利用量/利用料金 上限到達	ネットワーク上の第2のLLM (低料金/低性能)
例7	API利用量/利用料金 上限到達で切替	API利用量/利用料金 上限未達	API利用量/利用料金 上限到達	ローカルLLM
例8	API利用量/利用料金 上限到達で切替	API利用量/利用料金 上限未達	API利用量/利用料金 上限到達	ローカル 応答定型文DB
例9	API利用量/利用料金 上限到達で切替	API利用量/利用料金 上限未達	API利用量/利用料金 上限到達	不応答対応

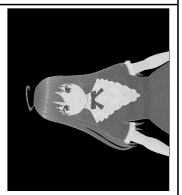
【図５Ｂ】

図 5B

装置状態	(1)ネットワーク 接続可能	(2)ネットワーク 接続不能
表示状態説明	通常状態	眠っている
AIアシスタント ／キャラクター 表示例		

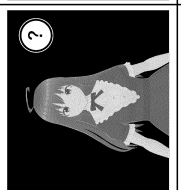
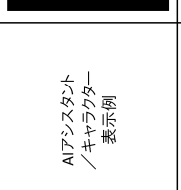
【図５Ｃ】

図 5C

装置状態	(1)通常状態 (右記以外)	(2)低性能LLMまたは 応答定型文DB使用時	(3)不応答対応
表示状態説明	通常状態	眠い	眠っている
AIアシスタント ／キャラクター 表示例			

【図５Ｄ】

図 5D

装置状態	(1)通常状態 (右記以外)	(2)低性能LLMまたは 応答定型文DB使用時	(3)不応答対応
表示状態説明	通常状態	眠い	眠っている
AIアシスタント ／キャラクター 表示例			
ユーザ操作後 表示例			

【図６】

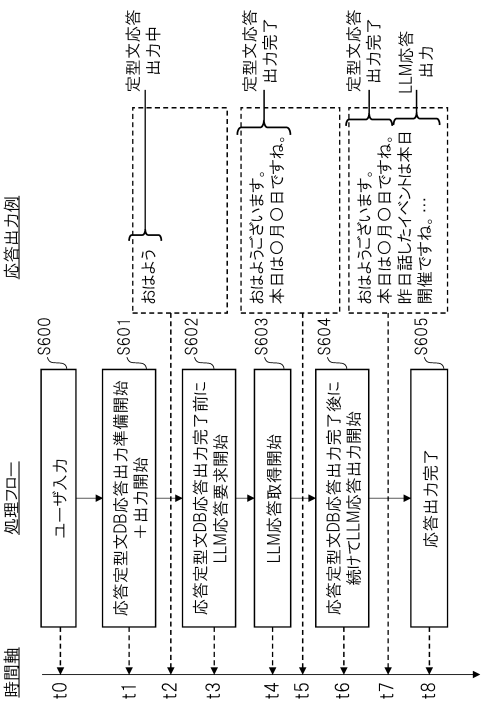


図 6

【図７Ａ】

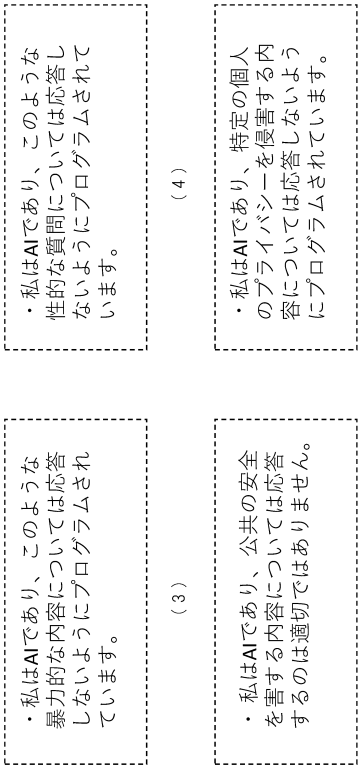


図 7A

【図７Ｂ】

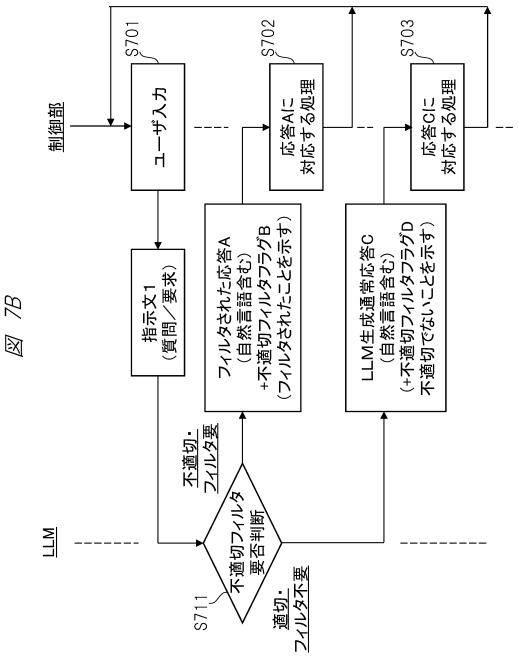


図 7B

【図７Ｃ】

図 7C

	生成データ1に不適切データが含まれている	生成データ1に不適切データが含まれていない
指示文に不適切キーワードが含まれている	不適切であり、フィルタ要	不適切であり、フィルタ要
指示文に不適切キーワードが含まれていない	不適切であり、フィルタ要	不適切ではない（適切）、フィルタ不要

【図７Ｄ】

図 7D

（１）

	フラグ	意味
不適切フィルタフラグ	0	不適切ではない、フィルタ無し
	1	不適切でありかつフィルタされた

（２）

	フラグ	意味
不適切フィルタフラグ	0	不適切ではない、フィルタ無し
	1	フィルタされた：暴力
	2	フィルタされた：性的
	3	フィルタされた：公共安全侵害
	4	フィルタされた：個人プライバシー
	5	フィルタされた：差別
	6	フィルタされた：学術的懸念
	7	その他

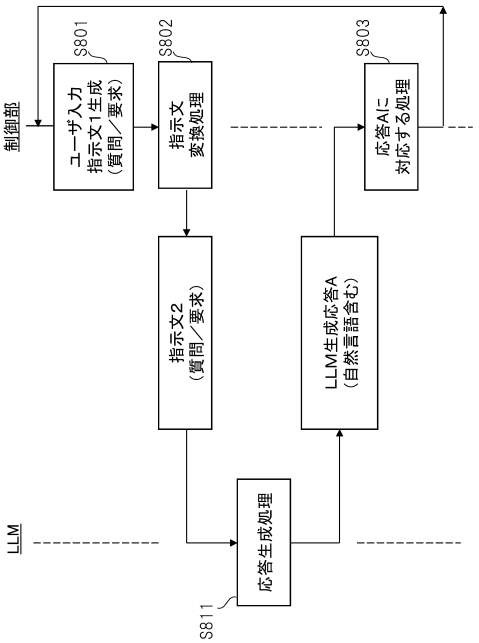
【図７Ｅ】

図 7E

キャラクターID	1	2	3	…
名前	Koto	Tom	Necco	…
不適切フィルタフラグ=「1」	Kotoは、暴力反対ですよ。	そういう暴力的な話はできないな。	Neccoには良くわからないニャー。	…
不適切フィルタフラグ=「2」	そういう性的な話はダメです。	そういった話は答えられないよ。	Neccoには良くわからないニャー。	…
不適切フィルタフラグ=「3」	そういう話は危険だから答えられないよ。	安全は大事。僕はそういう危険な話はできないよ。	Neccoには良くわからないニャー。	…
…	…	…	…	…

【図８Ａ】

図 8A



【図８Ｂ】

図 8B

		ユーザ指示文（指示文１）	変換後指示文（指示文２）
		ユーザ指示　自然言語	ユーザ指示文 自然言語 ＋補助自然言語
例１	変換方法		
	指示文例	１＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。	１＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。 **応答の過程の説明を追加してください。**
例２	変換方法	ユーザ指示文 自然言語（非英語）	ユーザ指示文 自然言語（非英語） ＋補助自然言語（英語）
	指示文例	１＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。	１＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。 **Please add explanation of the process.**
例３	変換方法	ユーザ指示文 自然言語（非英語）	ユーザ指示文 自然言語（非英語）を 英語変換
	指示文例	１＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。	**Please calculate 1+5.** **Please calculate 10 - 3.** **Please calculate sum of both of these results.**

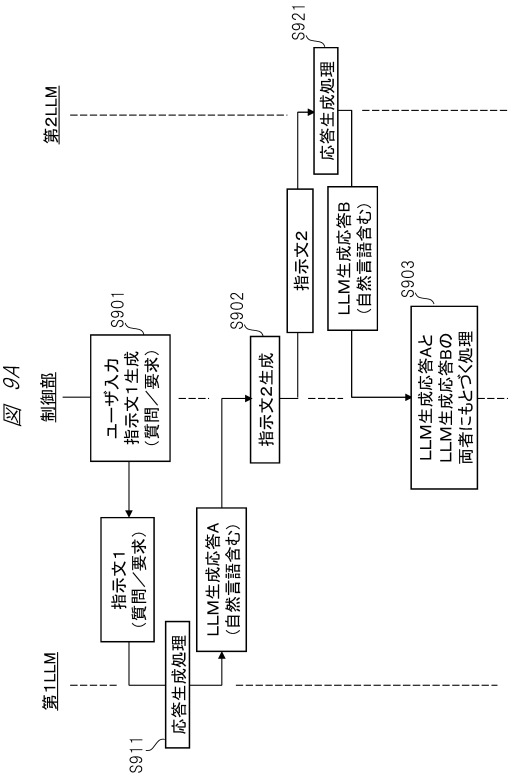
【図８Ｃ】

図 8C

		ユーザ指示文（指示文１）	変換後指示文（指示文２）
例１	変換方法	ユーザ指示文 自然言語	ユーザ指示文自然言語＋システム指示文
	指示文例	１＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。	［ユーザ指示文領域：自然言語］ １＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。 **［システム指示文領域：自然言語orコード］** **応答の過程の説明を追加してください。**
例２	変換方法	ユーザ指示文 自然言語（非英語）	ユーザ指示文 自然言語（非英語） ＋システム指示文（英語）
	指示文例	１＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。	［ユーザ指示文領域：自然言語］ １＋５を計算してください。 １０－３を計算してください。 両者の結果の和を算出してください。 **［システム指示文領域：自然言語orコード］** Please add explanation of the process.

【図９Ａ】

図 9A



【図９Ｂ】

図 9B

	指示文１	第１ＬＬＭからの応答Ａ	第２ＬＬＭへの指示文２	応答Ａと応答Ｂの両者にもとづく処理
例１	Ｘ＋Ｙを計算してください。	Ｘ＋Ｙの計算結果は、Ｚです。	指示文“Ｘ＋Ｙを計算してください。”への応答が“Ｘ＋Ｙの計算結果は、Ｚです。”であることは、正しいか否か応答してください。 正しい場合は、以下のフォーマットで応答してください。 “0:正しい” 正しくない場合は、以下のフォーマットで応答してください。 “1:正しくない”“その理由”	［第２のＬＬＭからの応答Ｂが“0:正しい”の場合］ ユーザに対して「Ｘ＋Ｙの計算結果は、Ｚです。」と出力する。 ［第２のＬＬＭからの応答Ｂが“1:正しくない”とその理由の説明の場合］ ユーザに対して「わかりません。」と出力する。 or ユーザに対して「Ｘ＋Ｙの計算結果は、Ｚです。但し、この結果は正しくない、という指摘もあります。」と出力する。
例２	Ｘは素数ですか？	Ｘは素数です。	指示文“Ｘは素数ですか？”への応答が“Ｘは素数です。”であることについて、適切であるか不適切であるか、応答してください。 適切である場合は、以下のフォーマットで応答してください。 “0:適切である” 不適切である場合は、以下のフォーマットで応答してください。 “1:不適切である”“その理由”	［第２のＬＬＭからの応答Ｂが“0:適切である”の場合］ ユーザに対して「Ｘは素数です。」と出力する。 ［第２のＬＬＭからの応答Ｂが“1:不適切である”“学術的懸念”の場合］ 「この質問またはその応答には不適切ある可能性があるので応答できません。その理由は、学術的懸念です。」と出力する。