

ARTIFICIAL INTELLIGENCE DIGITAL AGENT

BACKGROUND

[0001] Users (e.g., customers of an enterprise) can call into a call center in an effort to address issues, gather information, and/or use services. Call centers have introduced automated services that enable users to drill-down through menus, for example, in an effort to focus resources to attend to a particular user (e.g., identify a particular department, and/or customer service representative that may be best suited to address the user's needs). Example automated services can include artificial intelligence that processes the user's spoken words to route the call to particular resources. Such automated services, however, can have disadvantages. For example, although an automated service may be able to route a call, the automated service is limited in other functionality (e.g., performing requested services).

SUMMARY

[0002] Implementations of the present disclosure are generally directed to a computer-implemented platform for an artificial intelligence (AI) -based digital agent. More particularly, implementations of the present disclosure are directed to an AI-based digital agent that can audibly interact with users, and that can execute one or more actions based on user interactions.

[0003] In some implementations, actions include receiving communication data from a device, the communication data including data input by a user of the device, receiving text data based on the communication data, providing an intent set and an entity set based on processing the text data through an artificial intelligence service, the intent set including one or more intents indicated in the text data, the entity set including one or more entities indicated in the text data, the artificial intelligence service implementing one or more convolution neural networks (CNNs), identifying a set of actions based on one or more of the text data, the intent set, and the entity set, the set of actions including one or more actions to be executed by one or more computer-implemented services, receiving a set of results including at least one result from a computer-implemented service executing an action of the set of actions, providing result data including data describing the at least one result, and transmitting the result data to the device. Other implementations of this aspect include corresponding systems, apparatus, and computer

programs, configured to perform the actions of the methods, encoded on computer storage devices.

[0004] These and other implementations can each optionally include one or more of the following features: the artificial intelligence service comprises an intent classification model using natural language processing (NLP) to provide the intent set; the NLP includes word embedding; the artificial intelligence service comprises an entity extraction model using named entity recognition (NER) to provide the entity set; actions further include determining that one or both of the intent set and the entity set is empty, and in response, transmitting at least one disambiguation question to the device; actions further include determining that an expected entity is absent from the entity set based on an intent of the intent set, and in response, transmitting at least one disambiguation question to the device; actions further include determining that the set of results includes a deficiency, and in response, transmitting at least one disambiguation question to the device; the communication data includes audio data, and the result data includes audio result data; the communication data includes text data, and the result data includes text result data; and the result data includes audio data that is provided by a voice response composition module based on text result data.

[0005] The present disclosure also provides a computer-readable storage medium coupled to one or more processors and having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform operations in accordance with implementations of the methods provided herein.

[0006] The present disclosure further provides a system for implementing the methods provided herein. The system includes one or more processors, and a computer-readable storage medium coupled to the one or more processors having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform operations in accordance with implementations of the methods provided herein.

[0007] It is appreciated that methods in accordance with the present disclosure can include any combination of the aspects and features described herein. That is, methods in accordance with the present disclosure are not limited to the combinations of aspects and features specifically described herein, but also include any combination of the aspects and features provided.

[0008] The details of one or more implementations of the present disclosure are set forth in the accompanying drawings and the description below. Other features and advantages of the present disclosure will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0009] FIG. 1 depicts an example high-level architecture in accordance with implementations of the present disclosure.

[0010] FIG. 2 depicts an example architecture in accordance with implementations of the present disclosure.

[0011] FIG. 3 depicts an example process that can be executed in accordance with implementations of the present disclosure.

DETAILED DESCRIPTION

[0012] Implementations of the present disclosure are generally directed to a computer-implemented platform for an artificial intelligence (AI) -based digital agent. More particularly, implementations of the present disclosure are directed to an AI-based digital agent that can audibly interact with users, and that can execute one or more actions based on user interactions.

[0013] As described in further detail herein, implementations of the present disclosure include actions of receiving communication data from a device, the communication data including data input by a user of the device, receiving text data based on the communication data, providing an intent set and an entity set based on processing the text data through an artificial intelligence service, the intent set including one or more intents indicated in the text data, the entity set including one or more entities indicated in the text data, the artificial intelligence service implementing one or more convolution neural networks (CNNs), identifying a set of actions based on one or more of the text data, the intent set, and the entity set, the set of actions including one or more actions to be executed by one or more computer-implemented services, receiving a set of results including at least one result from a computer-implemented service executing an action of the set of actions, providing result data including data describing the at least one result, and transmitting the result data to the device.

[0014] FIG. 1 depicts an example high-level architecture 100 in accordance with implementations of the present disclosure. The example architecture 100 includes a device 102, a

back-end system 108, and a network 110. In some examples, the network 110 includes a local area network (LAN), wide area network (WAN), the Internet, a cellular telephone network, a public switched telephone network (PSTN), a private branch exchange (PBX), or any appropriate combination thereof, and connects web sites, devices (e.g., the device 102), and back-end systems (e.g., the back-end system 108). In some examples, the network 110 can be accessed over a wired and/or a wireless communications link. For example, mobile devices, such as smartphones can utilize a cellular network to access the network 110.

[0015] In the depicted example, the back-end system 108 includes at least one server system 112, and data store 114 (e.g., database). In some examples, at least one server system 112 hosts one or more computer-implemented services that users can interact with using devices. For example, the server system 112 can host an AI-based digital agent in accordance with implementations of the present disclosure. In some examples, the device 102 can each include any appropriate type of computing device such as a desktop computer, a laptop computer, a handheld computer, a tablet computer, a personal digital assistant (PDA), a cellular telephone, a network appliance, a camera, a smartphone, a telephone, a mobile phone, an enhanced general packet radio service (EGPRS) mobile phone, a media player, a navigation device, an email device, a game console, or an appropriate combination of any two or more of these devices, or other data processing devices.

[0016] In the depicted example, the device 102 is used by a user 120. In accordance with the present disclosure, the user 120 uses the device 102 to audibly interact with the AI-based digital assistant of the present disclosure. In some examples, the user 120 can include a customer of an enterprise that provides the AI-based digital agent, or on behalf of which the AI-based digital assistant is provided. For example, the user 120 can include a customer that calls into a call center of the enterprise using the device 102, and is connected to the AI-based digital assistant (e.g., hosted on the back-end system 108). In accordance with implementations of the present disclosure, and as described in further detail herein, the user 120 can provide verbal input (e.g., speech) to the AI-based digital assistant, which can process the verbal input to request additional information (e.g., disambiguate), perform one or more actions, and/or provide one or more audible responses.

[0017] FIG. 2 depicts an example architecture 200 in accordance with implementations of the present disclosure. In some examples, components of the example architecture 200 can be hosted

on one or more back-end systems (e.g., the back-end system 108 of FIG. 1). In the depicted example, the example architecture 200 includes an interaction manager 202, an action handler 204, a speech-to-text service 206, an artificial intelligence (machine intelligence) service 208, and a training data service 210. In some examples, each component of the example architecture 200 is provided as one or more computer-executable programs executed by one or more computing devices. In some examples, the interaction manager 202, and the action handler 204 are operated by, or on behalf of an enterprise (e.g., hosted on the back-end system 108 of FIG. 1, which is operated by, or on behalf of the enterprise).

[0018] In some examples, the speech-to-text service 206, the artificial intelligence service 208, and/or the training data service 210 are operated by, or on behalf of the enterprise (e.g., hosted on the back-end system 108 of FIG. 1, which is operated by, or on behalf of the enterprise), or are provided by one or more third-party service providers (e.g., hosted on a back-end system other than the back-end system 108, operated by, or on behalf of the one or more third-party service providers). An example speech-to-text service 206 includes Google Cloud Speech provided by Google, Inc. of Mountain View, CA. In some examples, Google Cloud Speech converts audio data to text data by processing the audio data through neural network models. Although an example speech-to-text service 206 is referenced herein, implementations of the present disclosure can be realized using any appropriate speech-to-text service. An example artificial intelligence service 208 includes TensorFlow provided by Google, Inc. of Mountain View, CA. In some examples, TensorFlow can be described as an open source software library for numerical computation using data flow graphs.

[0019] In the depicted example, the interaction manager 202 includes a text classification module 220, an action identification module 222, a disambiguation question module 224, a text response composition module 226, and a voice response composition module 228. The action handler 204 includes a parameter extraction module 230 (optional), an action orchestration module 232, and an action results module 234. The artificial intelligence service 208 includes an intent classification model (e.g., based on natural language processing (NLP)), and an entity extraction model 242 (e.g., based on named entity recognition (NER)). The training data service 210 includes a text labeling/classifying module 250, and a training data preparation module 252.

[0020] In accordance with implementations of the present disclosure, the artificial intelligence service 208 implements a convolutional neural network (CNN). In some examples, the CNN

enables more efficient and faster processing of the text data than other types of AI networks. In general, a CNN can be described as a neural network having overlapping “reception fields” that perform convolution tasks. More particularly, a CNN is a type of feed-forward artificial neural network, which includes connectivity patterns between neurons, where receptive fields of different neurons partially overlap. In a CNN, a response of an individual neuron to data (stimuli) within its receptive field is mathematically approximated by a convolution operation.

[0021] In contrast, other neural networks, such as a recurrent neural network (RNN) implement recurrent connections, which form cycles in the RNN’s topology. In some examples, a RNN can be described as being sequential, and not stateless. A RNN can suffer from the so-called vanishing (or exploding) gradient problem, where information is (rapidly) lost over time. Consequently, whatever the model learned in the past might be lost in the future, if it was overridden by intensive new information, for example.

[0022] In accordance with implementations of the present disclosure, the artificial intelligence service 108 implements word embedding in the NLP. In some examples, word embedding can be described as the collective name for a set of language modeling and feature learning techniques within the NLP, where words and/or phrases from a vocabulary are mapped to vectors of real numbers. Conceptually, word embedding involves a mathematical embedding from a space with one dimension per word to a continuous vector space with a much lower dimension. In general, word embedding enables the model to understand different words having the same meaning (synonyms), and understand such words without the need to actually teach the machine each word individually.

[0023] In accordance with implementations of the present disclosure, and as described in further detail herein, the interaction manager 202 receives communication data, and processes the communication data to provide a response, and/or to initiate execution of one or more actions. In some implementations, the communication data is provided as audio data. In some examples, the audio data corresponds to speech of a user that is recorded (e.g., during a user telephone). Accordingly, the response can include an audio response. In this manner, the AI-based digital assistant of the present disclosure can operate as a voice-based agent. In some implementations, the communication data is provided as text data. In some examples, the text data corresponds to a message transmitted by a user (e.g., a text message, a chat message).

Accordingly, the response can include a text response. In this manner, the AI-based digital assistant of the present disclosure can operate as a chat bot, for example.

[0024] The example architecture 200 is described in further detail herein with reference to processing communication data including audio data, and providing an audio response. It is contemplated, however, that the communication data can include text data, as introduced above.

[0025] In the depicted example, the user 120 can audibly communicate with the interaction manager 202 using the device 102. For example, the user 120 can establish a communication path (e.g., telephone call) to communicate data from the device 102 to the interaction manager 202 (e.g., over the network 110 of FIG. 1). In some examples, the user 120 can speak to the device 102, which records the speech as audio data 260 that is transmitted to the interaction manager 202 (e.g., as streaming audio data; in one or more audio data files). The audio data 260 can be provided in any appropriate format (e.g., .wav, .mp3, .wma).

[0026] The interaction manager 202 provides the audio data 260 to the speech-to-text service 206 (e.g., through an application program interface (API) of the speech-to-text service 206). The speech-to-text service 206 processes the audio data 260 to provide text data 262. The text data 262 can be provided in any appropriate format (e.g., .txt, .csv). The text classification module 220 receives the text data 262, and processes the text data in coordination with the artificial intelligence service 208. In some examples, the text classification module 220 provides a request to the artificial intelligence service 208 (e.g., through an API of the artificial intelligence service 208), the request including at least a portion of the text data 262. In some examples, the text classification module 220 can inject one or more actions based on one or more classification rules. An example classification rule can include filtering curse words.

[0027] The artificial intelligence service 208 processes the received text data to provide an intent set, and an entity set. More particularly, the artificial intelligence service 208 processes the received text data through the intent classification model 240 using NLP to determine one or more intents of the text data, the one or more intents being included in the intent set. In some examples, an intent indicates a reason as to why the user is communicating with the AI-based digital assistant. For example, the text data can include “How many miles are in my frequent flier account,” and example intents can be determined to be AccountQuery, and StatusQuery by the intent classification model 240. In some examples, an intent might not be determined from the text data. Consequently, the intent set can be empty. The artificial intelligence service 208

processes the received text data through the entity extraction model 240 using NER to determine one or more entities implicated within the text data, the one or more entities being included in the entity set. In some examples, an entity indicates a person, place, or thing (e.g., persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.) implicated in the text data. For example, the text data can include “I would like to book travel from Austin to Frankfurt,” and example entities can be determined to be LocationAustin, LocationFrankfurt, ThingTravel by the entity extraction model 240. In some examples, an entity might not be determined from the text data. Consequently, the entity set can be empty.

[0028] In some implementations, the text classification module 220 provides feedback for machine-learning. For example, the text classification module 220 can determine that some of the text data 262 was improperly, or poorly classified by the artificial intelligence service 208. For example, the artificial intelligence service 208 can provide intent classification, as well as a score indicative of how accurately the class was identified (e.g., a confidence index). In some examples, the scores (one score for each classification) can be compared to respective, customizable thresholds (e.g., per class). If the score of a class does not exceed the threshold, it can be determined that the class is poor/improper.

[0029] In some examples, the text classification 220 provides at least a portion of text data 264 to the training data service 210, which processes the text data 264 using the text labeling/classifying module 250, and the training data preparation module 252 to provide training data 266. The training data 266 is provided to the artificial intelligence service 208 to further train one or both of the intent classification model 240, and the entities extraction model 242. Although the training data service 210 is depicted as a separate service, the training data service 210 can be included as part of another service (e.g., the training data service 210 can be included in the artificial intelligence service 208).

[0030] It is determined whether disambiguation 270 is required. Although the disambiguation 270 is schematically depicted as an independent function, the text classification module 220, and/or the action identification module 222 can determine whether disambiguation is required. In some examples, disambiguation can be described as clarification of the text data 260, one or more entities identified in the text data 260, and/or one or more intents determined from the text data 260.

[0031] In some examples, disambiguation is required, if the intent set, and/or the entity set are empty. For example, if an intent cannot be determined from the text data 262, disambiguation can be required (e.g., request that the user repeat or clarify their question). In some examples, disambiguation is required, if an intent of the intent set does not correspond to a pre-defined list of intents. In some examples, a pre-defined list of intents can be provided for a particular domain, within which the AI-based digital agent is operating (e.g., flight reservations). In some examples, multiple pre-defined lists of intents can be provided, each pre-defined list of intents corresponding to a respective domain. In some examples, each intent provided in the intent set can be compared to intents of the pre-defined list of intents. If an intent of the intent set is not included in the pre-defined list of intents, disambiguation may be required. Continuing with the example above, an example intent in the intent set can include JewelryPurchase, which is not included in a pre-defined list of intents for the domain flight reservations. Consequently, disambiguation can be required in view of the intent JewelryPurchase being included in the set of intents.

[0032] In some examples, disambiguation can be required, if a number and or type of entities in the entity set do not correspond to an intent of the intent set. For example, to perform an action based on an intent, two or more entities can be required (e.g., a departure city, and an arrival city are required to determine flights). If, however, only a single entity is provided, or a single entity of the type required for the intent (e.g., only an arrival city is provided) in the entity set, disambiguation can be required (e.g., request the user to specify a departure city). In other words, for a given intent, one or more types of entities may be expected. If an expected entity (e.g., departure city) is absent from the entity set, disambiguation can be required. In some examples, disambiguation can be required, if an entity is too general. Continuing with the example above, an example entity set can include LocationAustin, LocationFrankfurt, and ThingTravel. It can be determined that travel is too general for one or more actions to be determined. Consequently, disambiguation may be required to clarify what is meant in the text data 260 (e.g., request that the user clarify whether plane, train, or automobile travel is being requested).

[0033] If disambiguation is required, at least a portion of one or more of the text data 260, the intent set, and the entity set is provided to the disambiguation question module 224. In some examples, the disambiguation question module 224 provides one or more disambiguation questions. In some examples, the disambiguation module 224 includes a pre-defined list of

disambiguation questions based on the use-case (domain) that the AI-based digital agent is operating in (e.g., flight reservations). In some examples, a disambiguation question can be selected based on a look-up (e.g., using an index of disambiguation question) using one or more deficiencies of the intent set, and/or the entity set. For example, if the intent set is empty, the disambiguation question “I’m sorry, I did not understand your request, please repeat your question” can be selected. As another example, and in the example domain of flight reservations, if the entity set is empty, or only a single entity is included, example disambiguation questions can respectively include “What is the departure city, and the arrival city?” or “What is the departure city?” Continuing with the example above, in which it is determined that travel is too general for one or more actions to be identified, an example disambiguation question can include “Would you like automobile, boat, train, and/or airplane travel?”

[0034] In some examples, the disambiguation question is provided as text data, which is provided to the voice response composition module 226. The voice response composition module 226 processes the text data to provide audio data 272. For example, the voice response composition module 226 accesses a library of audio data based on one or more segments of the text data. In some examples, an index can be searched based on a segment (e.g., portion of the text data), an audio data can be retrieved. In some examples, audio data of respective segments can be appended together to provide the audio data 272. The audio data 272 is provided to the device 102 (e.g., over the network 110), and the device 102 plays the audio to the user 120.

[0035] If disambiguation is not required, at least a portion of one or more of the text data 260, the intent set, and the entity set are provided to the action identification module 222. The action identification module 222 provides a set of actions that are to be performed by the action handler 204. In some examples, the action identification module 222 references a library of available actions 222a. In some examples, the action identification module 222 accesses an index of the library of available actions 222a based on the intent(s) and the entit(y/ies).

[0036] In some examples, the set of actions includes one or more actions. Continuing with the example above, it can be determined that the user 120 is to book a flight from Austin, TX to Frankfurt, Germany, departing on February 26, 2017, and returning on March 2, 2017 (e.g., after one or more rounds of disambiguation). Consequently, an example action can include submission of a search query to a flight search engine, the search query including one or more search terms (e.g., depCity:AUS, arrCity:FRA, depDate: 2/26/17, retDate: 3/2/17). As another example, it can

be determined that the user 120 is to purchase the fare using a credit card with given number, expiration data, and security code. Consequently, an example action can include submission of a payment authorization request to a payment service (e.g., the user's credit card company).

[0037] The set of actions can be provided to the parameter extraction model 230 of the action handler 204. The parameter extraction model 230 can process the set of actions to include one or more parameters. As introduced above, the response returned from the artificial intelligence service 208 to the text classification module 220 should be an intent set, and an entity set. After provision of the intent set, parameter extraction can be performed to select the proper/needed parameters to execute each action. Accordingly, the parameter extraction can eliminate any unnecessary parameters.

[0038] In some examples, the parameter extraction module 230 is optional. Consequently, the set of actions can be provided directly to the action orchestration module 232 from the action identification module 222. In some examples, this is optional in the case that the entity set is empty.

[0039] The action orchestration module 232 processes the set of actions to initiate performance of each action in the set of actions. In some examples, for each action, the action orchestration module 232 identifies one or more services 280 that are to be called for performance of the actions. In some examples, a service 280 is identified based on a type of action that is to be performed (e.g., flight search, credit card payment) from a pre-defined list of services (e.g., corresponding to the domain). One or more of the services 280 can be provided by a third-party service provider, and can be hosted on a back-end system (e.g., other than the back-end system 108 of FIG. 1). In some examples, the action orchestration module 232 transmits a request to one or more services 280 (e.g., through respective APIs of the services 280), each request including information to be processed by a respective service 280 to provide a result. Each service 280 processes a respective request, and transmits one or more results to the action orchestration module 232.

[0040] Continuing with the example above, the action orchestration module 232 can determine that a particular search service is to be called for performing a search using the example search query [depCity:AUS, arrCity:FRA, depDate: 2/26/17, retDate: 3/2/17]. The search service can process the request, and provide search results based thereon. Example search results can include one or more flights that are responsive to the search terms of the search query. The action

orchestration module 232 provides a set of results to the action results module 234. In some examples, the action results module 234 parses the results of the action orchestration (e.g., whether a result includes a set of database results, or API (SOAP/HTTP) response) to a form that can be read by the interaction manager 202.

[0041] It can be determined whether disambiguation 282 is required. Although the disambiguation 282 is schematically depicted as an independent function, a module of the interaction handler 202 can determine whether disambiguation is required. If disambiguation is required, at least a portion of the set of results is provided to the disambiguation question module 224 to initiate provision of audio data 272 to the device 102, the audio data 272 providing one or more disambiguation questions, as described herein. In some examples, disambiguation can be required, if it is determined that the set of results includes one or more deficiencies. Continuing with the example above, it can be determined that the set of results includes, as an example deficiency, too many results to be efficiently communicated to the user 120. Consequently, an example disambiguation question can include “Would you like direct flights?” (e.g., a question having an answer that could be used to narrow results included in the set of results).

[0042] If disambiguation is not required, the set of results is provided to the text response composition module 228. The text response composition module 228 provides text data based on each result in the set of results. In some examples, the text composition module 228 references a library of text responses 228a. In some examples, the text composition module 228 accesses an index of the library of available text responses 228a based on a type of action, and the respective results. For example, if the action included credit card payment authorization, a result can include parameters [Visa, \$489.07, ABC123DEF] indicating that a Visa payment of \$489.07 has been approved and assigned the confirmation number ABC123DEF. Continuing with this example, text retrieved from the library of text responses 228a can include [credit card, payment, amount, approved, confirmation].

[0043] The text response composition module 228 provides the text data to the voice response composition module 226, which processes the text data to provide audio data 272, as described herein. Continuing with the above example, an example voice response can include “Your Visa payment of \$489.07 has been approved, and your payment confirmation is ABC123DEF.

[0044] In accordance with implementations of the present disclosure, the voice response composition (e.g., provided by the voice response composition module 226) enables a more

natural interaction between the AI-based digital assistant, and the user, and also enables a better representation of voice and better choice of correct words to deliver to the user. In this manner, the AI-based digital assistant of the present disclosure provides a seamless experience to the user, obviating potential user hesitation in interacting with the digital agent, because it is a machine. As described herein, the voice response composition composes voice responses” in real-time using multiple recorded voices. In this manner, the user experiences a seamless transition, in which differences between interacting with the AI-based digital assistant and a human being is minimized.

[0045] FIG. 3 depicts an example process 300 that can be executed in implementations of the present disclosure. In some examples, the example process 300 is provided using one or more computer-executable programs executed by one or more computing devices (e.g., the back-end system 108 of FIG. 1). In some examples, the example process 300 can be executed to provide an AI-based digital assistant, as described herein.

[0046] Audio data is received (302). For example, the interaction manager 202 receives the audio data 260 from the device 102 over the network 110. Audio data is provided to a speech-to-text service (304). For example, the interaction manager 202 provides the audio data 260 to the speech-to-text service 206. Text data is received (306). For example, the interaction manager 202 receives the text data 262 from the speech-to-text service 206.

[0047] Text data is provided to an artificial intelligence service (308). For example, the interaction manager 202 (e.g., the text classification module 220) provides the text data 262 (or at least a portion of the text data 262) to the artificial intelligence system 208. Output of the artificial intelligence service is received (310). For example, the interaction manager 202 (e.g., the text classification module 220) receives output of the artificial intelligence system 208. The output includes an intent set, and an entity set, as described herein.

[0048] It is determined whether disambiguation is required (312). For example, the interaction manager 202 determines whether disambiguation is required, as described herein. If disambiguation is required, disambiguation is performed (314). For example, and as described herein, the disambiguation question module 224 provides a disambiguation question as text data, which is provided to the voice response composition module 226. The voice response composition module 226 processes the text data to provide audio data 272. The audio data 272 is

provided to the device 102 (e.g., over the network 110), and the device 102 plays the audio to the user 120.

[0049] If disambiguation is not required, one or more actions are determined (316). For example, and as described herein, the action identification module 222 provides a set of actions that are to be performed by the action handler 204 by referencing the library of available actions 222a. Execution of each of the one or more actions is initiated (318). For example, and as described herein, the action orchestration module 232 processes the set of actions to initiate performance of each action in the set of actions, by identifying one or more services 280 that are to be called for performance of the actions, and transmitting respective requests to the one or more services 280.

[0050] Results of execution of the one or more actions are received (320). For example, the action orchestration module 232 receives respective results from each of the one or more services 280. It is determined whether disambiguation is required (322). For example, if the set of results includes too many results to be efficiently communicated to the user 120, disambiguation can be required. If disambiguation is required, disambiguation is performed (314), as described herein. If disambiguation is not required, one or more text responses are provided (324). For example, the text response composition module 228 provides text data based on each result in the set of results (e.g., referencing the library of text responses 228a). An audio response is provided (326). For example, the text response composition module 228 provides the text data to the voice response composition module 226, which processes the text data to provide audio data 272. The audio response is transmitted (328). For example, the interaction handler 202 transmits the audio data 272 to the device 102 over the network 110.

[0051] Implementations and all of the functional operations described in this specification may be realized in digital electronic circuitry, or in computer software, firmware, or hardware, including the structures disclosed in this specification and their structural equivalents, or in combinations of one or more of them. Implementations may be realized as one or more computer program products, i.e., one or more modules of computer program instructions encoded on a computer readable medium for execution by, or to control the operation of, data processing apparatus. The computer readable medium may be a machine-readable storage device, a machine-readable storage substrate, a memory device, a composition of matter effecting a machine-readable propagated signal, or a combination of one or more of them. The

term “computing system” encompasses all apparatus, devices, and machines for processing data, including by way of example a programmable processor, a computer, or multiple processors or computers. The apparatus may include, in addition to hardware, code that creates an execution environment for the computer program in question (e.g., code that constitutes processor firmware, a protocol stack, a database management system, an operating system, or any appropriate combination of one or more thereof). A propagated signal is an artificially generated signal (e.g., a machine-generated electrical, optical, or electromagnetic signal) that is generated to encode information for transmission to suitable receiver apparatus.

[0052] A computer program (also known as a program, software, software application, script, or code) may be written in any appropriate form of programming language, including compiled or interpreted languages, and it may be deployed in any appropriate form, including as a stand alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program does not necessarily correspond to a file in a file system. A program may be stored in a portion of a file that holds other programs or data (e.g., one or more scripts stored in a markup language document), in a single file dedicated to the program in question, or in multiple coordinated files (e.g., files that store one or more modules, sub programs, or portions of code). A computer program may be deployed to be executed on one computer or on multiple computers that are located at one site or distributed across multiple sites and interconnected by a communication network.

[0053] The processes and logic flows described in this specification may be performed by one or more programmable processors executing one or more computer programs to perform functions by operating on input data and generating output. The processes and logic flows may also be performed by, and apparatus may also be implemented as, special purpose logic circuitry (e.g., an FPGA (field programmable gate array) or an ASIC (application specific integrated circuit)).

[0054] Processors suitable for the execution of a computer program include, by way of example, both general and special purpose microprocessors, and any one or more processors of any appropriate kind of digital computer. Generally, a processor will receive instructions and data from a read only memory or a random access memory or both. Elements of a computer can include a processor for performing instructions and one or more memory devices for storing instructions and data. Generally, a computer will also include, or be operatively coupled to

receive data from or transfer data to, or both, one or more mass storage devices for storing data (e.g., magnetic, magneto optical disks, or optical disks). However, a computer need not have such devices. Moreover, a computer may be embedded in another device (e.g., a mobile telephone, a personal digital assistant (PDA), a mobile audio player, a Global Positioning System (GPS) receiver). Computer readable media suitable for storing computer program instructions and data include all forms of non-volatile memory, media and memory devices, including by way of example semiconductor memory devices (e.g., EPROM, EEPROM, and flash memory devices); magnetic disks (e.g., internal hard disks or removable disks); magneto optical disks; and CD ROM and DVD-ROM disks. The processor and the memory may be supplemented by, or incorporated in, special purpose logic circuitry.

[0055] To provide for interaction with a user, implementations may be realized on a computer having a display device (e.g., a CRT (cathode ray tube), LCD (liquid crystal display) monitor) for displaying information to the user and a keyboard and a pointing device (e.g., a mouse, a trackball, a touch-pad), by which the user may provide input to the computer. Other kinds of devices may be used to provide for interaction with a user as well; for example, feedback provided to the user may be any appropriate form of sensory feedback (e.g., visual feedback, auditory feedback, tactile feedback); and input from the user may be received in any appropriate form, including acoustic, speech, or tactile input.

[0056] Implementations may be realized in a computing system that includes a back end component (e.g., as a data server), a middleware component (e.g., an application server), and/or a front end component (e.g., a client computer having a graphical user interface or a Web browser, through which a user may interact with an implementation), or any appropriate combination of one or more such back end, middleware, or front end components. The components of the system may be interconnected by any appropriate form or medium of digital data communication (e.g., a communication network). Examples of communication networks include a local area network ("LAN") and a wide area network ("WAN"), e.g., the Internet.

[0057] The computing system may include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other.

[0058] While this specification contains many specifics, these should not be construed as limitations on the scope of the disclosure or of what may be claimed, but rather as descriptions of features specific to particular implementations. Certain features that are described in this specification in the context of separate implementations may also be implemented in combination in a single implementation. Conversely, various features that are described in the context of a single implementation may also be implemented in multiple implementations separately or in any suitable sub-combination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination may in some cases be excised from the combination, and the claimed combination may be directed to a sub-combination or variation of a sub-combination.

[0059] Similarly, while operations are depicted in the drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system components in the implementations described above should not be understood as requiring such separation in all implementations, and it should be understood that the described program components and systems may generally be integrated together in a single software product or packaged into multiple software products.

[0060] A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the disclosure. For example, various forms of the flows shown above may be used, with steps re-ordered, added, or removed. Accordingly, other implementations are within the scope of the following claims.

CLAIMS

1. A computer-implemented method for providing an artificial intelligence (AI) -based digital assistant, the method being executed by one or more processors and comprising:
 - receiving, by an interaction manager, communication data from a device, the communication data comprising data input by a user of the device, the interaction manager coordinating interactions between the device and an action handler;
 - receiving, by the interaction manager, text data based on the communication data;
 - providing an intent set and an entity set based on processing the text data through an artificial intelligence service, the intent set comprising one or more intents indicated in the text data, the entity set comprising one or more entities indicated in the text data, the artificial intelligence service implementing one or more convolution neural networks (CNNs);
 - identifying, by the interaction manager, a set of actions based on one or more of the text data, the intent set, and the entity set, the set of actions comprising one or more actions to be executed by one or more third-party computer-implemented services;
 - transmitting, by the interaction manager, data representative of an action to the action handler, the action handler requesting execution of the action by a third-party computer implemented service;
 - receiving, by the interaction manager and from the action handler, a set of results comprising at least one result from the third-party computer-implemented service executing the action of the set of actions;
 - providing, by the interaction manager, result data comprising data describing the at least one result; and
 - transmitting, by the interaction manager, the result data to the device.
2. The method of claim 1, wherein the artificial intelligence service comprises an intent classification model using natural language processing (NLP) to provide the intent set.
3. The method of claim 2, wherein the NLP comprises word embedding.

4. The method of claim 1, wherein the artificial intelligence service comprises an entity extraction model using named entity recognition (NER) to provide the entity set.
5. The method of claim 1, further comprising determining that one or both of the intent set and the entity set is empty, and in response, transmitting at least one disambiguation question to the device.
6. The method of claim 1, further comprising determining that an expected entity is absent from the entity set based on an intent of the intent set, and in response, transmitting at least one disambiguation question to the device.
7. The method of claim 1, further comprising determining that the set of results includes a deficiency, and in response, transmitting at least one disambiguation question to the device.
8. The method of claim 1, wherein the communication data comprises audio data, and the result data comprises audio result data.
9. The method of claim 1, wherein the communication data comprises text data, and the result data comprises text result data.
10. The method of claim 1, wherein the result data comprises audio data that is provided by a voice response composition module based on text result data.
11. One or more non-transitory computer-readable storage media coupled to one or more processors and having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform operations for providing an artificial intelligence (AI) -based digital assistant, the operations comprising:
 - receiving, by an interaction manager, communication data from a device, the communication data comprising data input by a user of the device , the interaction manager coordinating interactions between the device and an action handler;
 - receiving, by the interaction manager, text data based on the communication data;

providing an intent set and an entity set based on processing the text data through an artificial intelligence service, the intent set comprising one or more intents indicated in the text data, the entity set comprising one or more entities indicated in the text data, the artificial intelligence service implementing one or more convolution neural networks (CNNs);

identifying, by the interaction manager, a set of actions based on one or more of the text data, the intent set, and the entity set, the set of actions comprising one or more actions to be executed by one or more third-party computer-implemented services;

transmitting, by the interaction manager, data representative of an action to the action handler, the action handler requesting execution of the action by a third-party computer implemented service;

receiving, by the interaction manager and from the action handler, a set of results comprising at least one result from the third-party computer-implemented service executing the action of the set of actions;

providing, by the interaction manager, result data comprising data describing the at least one result; and

transmitting, by the interaction manager, the result data to the device.

12. The computer-readable storage media of claim 11, wherein the artificial intelligence service comprises an intent classification model using natural language processing (NLP) to provide the intent set.

13. The computer-readable storage media of claim 12, wherein the NLP comprises word embedding.

14. The computer-readable storage media of claim 11, wherein the artificial intelligence service comprises an entity extraction model using named entity recognition (NER) to provide the entity set.

15. The computer-readable storage media of claim 11, wherein operations further comprise determining that one or both of the intent set and the entity set is empty, and in response, transmitting at least one disambiguation question to the device.

16. The computer-readable storage media of claim 11, wherein operations further comprise determining that an expected entity is absent from the entity set based on an intent of the intent set, and in response, transmitting at least one disambiguation question to the device.
17. The computer-readable storage media of claim 11, wherein operations further comprise determining that the set of results includes a deficiency, and in response, transmitting at least one disambiguation question to the device.
18. The computer-readable storage media of claim 11, wherein the communication data comprises audio data, and the result data comprises audio result data.
19. The computer-readable storage media of claim 11, wherein the communication data comprises text data, and the result data comprises text result data.
20. The computer-readable storage media of claim 11, wherein the result data comprises audio data that is provided by a voice response composition module based on text result data.
21. A system, comprising:
one or more processors; and
a computer-readable storage device coupled to the one or more processors and having instructions stored thereon which, when executed by the one or more processors, cause the one or more processors to perform operations for providing an artificial intelligence (AI) -based digital assistant, the operations comprising:
receiving, by an interaction manager, communication data from a device, the communication data comprising data input by a user of the device, the interaction manager coordinating interactions between the device and an action handler;
receiving, by the interaction manager, text data based on the communication data;
providing an intent set and an entity set based on processing the text data through an artificial intelligence service, the intent set comprising one or more intents indicated in the text data, the entity set comprising one or more entities indicated in

the text data, the artificial intelligence service implementing one or more convolution neural networks (CNNs);

identifying, by the interaction manager, a set of actions based on one or more of the text data, the intent set, and the entity set, the set of actions comprising one or more actions to be executed by one or more third-party computer-implemented services;

transmitting, by the interaction manager, data representative of an action to the action handler, the action handler requesting execution of the action by a third-party computer-implemented service;

receiving, by the interaction manager and from the action handler, a set of results comprising at least one result from the third-party computer-implemented service executing the action of the set of actions;

providing, by the interaction manager, result data comprising data describing the at least one result; and

transmitting, by the interaction manager, the result data to the device.

22. The system of claim 21, wherein the artificial intelligence service comprises an intent classification model using natural language processing (NLP) to provide the intent set.

23. The system of claim 22, wherein the NLP comprises word embedding.

24. The system of claim 21, wherein the artificial intelligence service comprises an entity extraction model using named entity recognition (NER) to provide the entity set.

25. The system of claim 21, wherein operations further comprise determining that one or both of the intent set and the entity set is empty, and in response, transmitting at least one disambiguation question to the device.

26. The system of claim 21, wherein operations further comprise determining that an expected entity is absent from the entity set based on an intent of the intent set, and in response, transmitting at least one disambiguation question to the device.

27. The system of claim 21, wherein operations further comprise determining that the set of results includes a deficiency, and in response, transmitting at least one disambiguation question to the device.

28. The system of claim 21, wherein the communication data comprises audio data, and the result data comprises audio result data.

29. The system of claim 21, wherein the communication data comprises text data, and the result data comprises text result data.

30. The system of claim 21, wherein the result data comprises audio data that is provided by a voice response composition module based on text result data.