DESCRIPTION

TITLE OF INVENTION: MATERIAL GENERATION DEVICE, MATERIAL GENERATION METHOD, AND PROGRAM

TECHNICAL FIELD

[0001]

The present invention relates to a material generation device, a material generation method, and a program.

BACKGROUND ART

[0002]

In recent years, generative artificial intelligence (generative AI) using deep learning, such as ChatGPT (registered trademark) of OpenAI (registered trademark) and Stable Diffusion of the CompVis group at Ludwig Maximilian University of Munich, has been developed. In Patent Document 1, an information processing device that generates, in a case of responding to question content with an answer derived by deep learning, an answer based on a plurality of pieces of information including text and an image related to the question content is disclosed.

Citation List

Patent Document

[0003]

Patent Document 1: Japanese Unexamined Patent Application, First Publication No. 2022-112541

SUMMARY OF INVENTION

Technical Problem

[0004]

However, in the information processing device disclosed in Patent Document 1, there is a problem that it may be impossible to obtain a material corresponding to characteristics of a person.

[0005]

The present disclosure has been made in view of such circumstances, and provides a material generation device, a material generation method, and a program capable of generating a material corresponding to characteristics of a person.

Solution to Problem

[0006]

The present disclosure has been made to achieve the above-described object, and an aspect of the present disclosure relates to a material generation device including: an input unit configured to receive input of a query that specifies a material to be generated; and a dialogue management unit configured to infer, based on the query, the material specified by the query by causing a plurality of artificial intelligence (AI) agents to hold a dialogue, in which information indicating characteristics of a person corresponding to at least one AI agent among the plurality of AI agents is included in the dialogue among the plurality of AI agents.

[0007]

According to the aspect of the present disclosure, in the material generation device described above, the plurality of AI agents are a plurality of dialogue-based document generation models, the input unit is a language input unit configured to receive the input of the query, the dialogue management unit is an information refining unit configured to infer a document in the material specified by the query by using the query as initial input and causing the plurality of dialogue-based document generation models to mutually repeat input and output to hold a dialogue, the material generation device

further comprises an image generation unit configured to generate an image based on the document that is an inference result by the information refining unit, and preference information including a preference or a thought pattern of a specific individual is added to the dialogue among the plurality of dialogue-based document generation models.

[0008]

According to the aspect of the present disclosure, the material generation device described above further includes an image input unit configured to receive input of an image serving as an example of the material to be generated, and an image recognition unit configured to input the image input to the image input unit to a trained model, cause the trained model to infer a feature value of the image input to the image input unit, and set the feature value as an image-derived context feature, and the image generation unit is configured to generate an image based on the image-derived context feature, in addition to the document that is the inference result by the information refining unit.

[0009]

According to the aspect of the present disclosure, the material generation device described above further includes a natural language processing unit configured to input the query to a trained model, cause the trained model to infer a feature value of the query, set the feature value as a language-derived context feature, and generate a generation query in which at least some of the query, the language-derived context feature, the image-derived context feature, the preference information, and the document that is the inference result by the information refining unit are combined, and a document generation unit configured to input the generation query to the dialogue-based document generation model, infer a document requested by the generation query, and set an inference result as an instruction query, and the image generation unit is configured to

input the instruction query to a trained image generation model and cause the trained image generation model to infer an image specified by the instruction query.

[0010]

According to the aspect of the present disclosure, in the material generation device described above, the dialogue among the plurality of AI agents includes generation of a hypothesis by one AI agent and criticism or critique of the generated hypothesis by another AI agent.

[0011]

According to the aspect of the present disclosure, in the material generation device described above, the plurality of AI agents correspond to a plurality of persons on a one-to-one basis, and each of the plurality of AI agents includes, in the dialogue, person information indicating characteristics of a corresponding person among the plurality of persons.

[0012]

According to the aspect of the present disclosure, in the material generation device described above, at least one of the plurality of persons is a virtual person, and the person information of the virtual person includes information indicating a role of the virtual person.

[0013]

According to the aspect of the present disclosure, the material generation device described above further includes a dialogue strategy storage unit configured to store dialogue strategy information indicating a dialogue process and persons to participate in a dialogue, and the dialogue management unit is configured to cause the AI agents selected from among the plurality of AI agents to hold the dialogue in accordance with

the dialogue process specified by the dialogue strategy information acquired from the dialogue strategy storage unit.

[0014]

According to the aspect of the present disclosure, the material generation device described above further includes a storage unit configured to store information related to the material, and each of the plurality of AI agents is configured to acquire reference information related to the dialogue from the storage unit and include the acquired reference information in the dialogue.

[0015]

According to the aspect of the present disclosure, in the material generation device described above, the input unit is configured to receive input of multimodal data related to the material, the dialogue management unit is configured to cause the plurality of AI agents to hold the dialogue based on the query and the multimodal information, and the dialogue among the plurality of AI agents is multimodal.

[0016]

According to the aspect of the present disclosure, the material generation device described above further includes an output generation unit configured to generate the material based on a result of the dialogue among the plurality of AI agents, and an evaluation/verification unit configured to evaluate and verify the material generated by the output generation unit.

[0017]

Another aspect of the present disclosure relates to a material generation method including: a first step of receiving input of a query that specifies a material to be generated; and a second step of inferring, based on the query, the material specified by the query by causing a plurality of AI agents to hold a dialogue, in which information

indicating characteristics of a person, including a virtual person, corresponding to at least one AI agent among the plurality of AI agents is added to the dialogue among the plurality of AI agents.

[0018]

Still another aspect of the present disclosure relates to a program causing a computer to function as: an input unit configured to receive input of a query that specifies a material to be generated; and a dialogue management unit configured to infer, based on the query, the material specified by the query by causing a plurality of AI agents to hold a dialogue, in which information indicating characteristics of a person, including a virtual person, corresponding to at least one AI agent among the plurality of AI agents is added to the dialogue among the plurality of AI agents.

Advantageous Effects of Invention

[0019]

The material generation device, the material generation method, and the program corresponding to the present disclosure can generate the material according to the characteristics of the person.

BRIEF DESCRIPTION OF DRAWINGS

[0020]

[FIG. 1] A schematic block diagram showing a configuration of a material generation device 10 according to a first embodiment of the present disclosure.

[FIG. 2] A schematic block diagram showing an example configuration of an image recognition unit 110 according to the first embodiment.

[FIG. 3] A schematic block diagram showing an example configuration of a natural language processing unit 130 according to the first embodiment.

[FIG. 4] A schematic block diagram showing an example configuration of an information refining unit 140 according to the first embodiment.

[FIG. 5] A schematic block diagram showing an example configuration of an image generation unit 170 according to the first embodiment.

[FIG. 6] A flowchart showing an operation example of the material generation device 10 according to the first embodiment.

[FIG. 7] A schematic diagram showing a use case of the material generation device 10 according to the first embodiment.

[FIG. 8] A schematic block diagram showing material generation device 20 according to a second embodiment of the present disclosure.

[FIG. 9] A schematic block diagram showing a configuration of a hypothesis generation unit 208 according to the second embodiment.

[FIG. 10] An explanatory diagram showing a hardware configuration of the material generation device 10 according to each embodiment.

DESCRIPTION OF EMBODIMENTS

[0021]

<First Embodiment>

Hereinafter, a first embodiment of the present disclosure will be described with reference to the accompanying drawings. FIG. 1 is a schematic block diagram showing a configuration of a material generation device 10 according to the first embodiment of the present disclosure. The material generation device 10 is a device that generates a material consisting of a sentence and an image in accordance with a request of a user. The material generation device 10 may be implemented by one or a plurality of computers reading and executing a program. The plurality of computers may be connected to be capable of communicating with each other via a general-purpose or

dedicated network. The material generation device 10 includes an image input unit 100, an image recognition unit 110, a language input unit 120 (input unit), a natural language processing unit 130, an information refining unit 140 (dialogue management unit), a storage unit 150, a document generation unit 160, an image generation unit 170, a display processing unit 180, and a display unit 190.

[0022]

The image input unit 100 receives input of an image D1 that is an example of the material to be generated, by the user. The image input unit 100 may receive the input by the user using input means such as a keyboard or a mouse, or may receive the input of the image D1 by receiving the input from another device. The image recognition unit 110 inputs the image D1 input to the image input unit 100 to a trained model, causes the trained model to infer a feature value, and sets the feature value as an image-derived context feature D2. The image-derived context feature D2 may be one in which a feature value having a significant weight is identified from the feature value of the inference result. The trained model may be, for example, a hierarchical network such as an image encoder of Contrastive Language-Image Pre-training (CLIP) of OpenAI (registered trademark) or a Vision Transformer (ViT). The trained model may be provided in the image recognition unit 110 or may be provided in another device that is communicably connected to the material generation device 10 via the network or the like. In addition, the identifying of the feature value having the significant weight may be performed by an attention mechanism such as multi-head attention.

[0023]

The language input unit 120 receives input, by the user, of language (query D3) specifying the material to be generated. The language input unit 120 may receive the input by the user using input means such as a keyboard or a mouse, or may receive the

input of the query D3 by receiving the input from another device. The query D3 is text data, and may include text data that is an example of the material to be generated. The natural language processing unit 130 inputs the query D3 input to the language input unit 120 to the trained model, causes the trained model to infer a feature value, and sets the feature value as a language-derived context feature. The language-derived context feature may be a feature value having a significant weight identified from the feature value of the inference result. Further, the natural language processing unit 130 generates a generation query D4 that is language (text data) in which at least some of the query D3, the language-derived context feature, the image-derived context feature D2, complementary information D5 that is preference information described later, and a summary sentence D6 that is a document of the inference result by the information refining unit 140 are combined. The generation query D4 may be a prompt for causing the document generation unit 160 to infer the document specified by the generation query D4.

[0024]

The trained model of the natural language processing unit 130 may be, for example, a hierarchical network such as a text encoder of CLIP. The trained model may be provided in the natural language processing unit 130 or may be provided in another device that is communicably connected to the material generation device 10 via the network or the like. In addition, the identifying of the feature value having the significant weight may be performed by an attention mechanism. The trained model of the natural language processing unit 130 is paired with the trained model of the image recognition unit 110, and is trained to infer the same feature value as the feature value inferred from the image by the trained model of the image recognition unit 110 as the feature value of the natural language describing the image.

[0025]

The information refining unit 140 infers the document of the material specified by the query D3 by using the query D3 as the initial input and causing the two dialogue-based document generation models to mutually repeat input and output to hold a dialogue. The information refining unit 140 adds preference information including a preference or a thought pattern of a specific individual to the dialogue among the two dialogue-based document generation models. The two dialogue-based document generation models may be, for example, dialogue-based document generation networks such as ChatGPT (registered trademark) of OpenAI (registered trademark). The information refining unit 140 summarizes the inferred document, and inputs the result as the summary sentence D6 to the natural language processing unit 130. Further, the information refining unit 140 inputs the result of the summary of the inferred document as a language answer D9 to the display processing unit 180.

[0026]

Among the two dialogue-based document generation models, one may be requested to infer a detailed document and the other may be requested to infer a concise document. Further, the input to at least one of the two dialogue-based document generation models may be one in which a dialogue/inference history D8 stored in the storage unit 150 is added to the output of the other. The dialogue/inference history D8 may include an inference history D7 that is stored in the storage unit 150 by the information refining unit 140 as the inference result by the two dialogue-based document generation models. In addition, the dialogue-based document generation models may be provided in the information refining unit 140, or may be provided in another device that is communicably connected to the material generation device 10 via the network or the like.

[0027]

Further, the dialogue/inference history D8 stored in the storage unit 150 may include preference information (for example, a posting history of the individual to a social network service (SNS)) including the preference or the thought pattern of the specific individual. The specific individual may be a user who requests the generation of the material, or may be a person or a group of persons assumed to read the generated material. In addition, the information refining unit 140 may input the preference information, or the thought pattern extracted from the preference information by an encoder as a high-dimensional feature, or the high-dimensional feature decoded as the complementary information D5 to the natural language processing unit 130.

[0028]

The document generation unit 160 inputs the generation query D4 to the dialogue-based document generation model, infers the document requested by the generation query D4, and sets the inference result as the instruction query D10. The dialogue-based document generation model may be, for example, a large-scale language model such as ChatGPT (registered trademark) or BERT. In addition, the dialogue-based document generation model may be provided in the document generation unit 160, or may be provided in another device that is communicably connected to the material generation device 10 via the network or the like. The instruction query D10 may be a prompt for causing the image generation unit 170 to infer the image corresponding to the document requested by the generation query D4.

[0029]

The image generation unit 170 generates the image based on at least the document that is the inference result by the information refining unit 140. Further, the image may be generated based on the image-derived context feature. Specifically, the

image generation unit 170 inputs the instruction query D10 generated by using this information to the trained image generation model, causes the trained image generation model to infer the image specified by the instruction query D10, and sets the inference result as an image answer D11. The trained image generation model may be, for example, an image synthesis model that generates an image from text, such as Stable Diffusion developed by the CompVis group of the University of Munich, or may be a combination of a text encoder of CLIP and a diffusion decoder corresponding to the text encoder.

[0030]

The display processing unit 180 acquires the language answer D9 from the information refining unit 140 and the image answer D11 from the image generation unit 170, and displays the language answer D9 and the image answer D11 on the display unit 190. The display unit 190 is a display such as a liquid-crystal display or an organic electro-luminescence (EL) display.

[0031]

FIG. 2 is a schematic block diagram showing an example configuration of the image recognition unit 110 according to the present embodiment. The image recognition unit 110 includes a trained deep feature value extraction unit 111 and a context analysis unit 112. The trained deep feature value extraction unit 111 inputs the image D1 to a trained model such as an image encoder of CLIP or ViT, and causes the trained model to infer the feature value. The context analysis unit 112 identifies the feature value having a significant weight by an attention mechanism from among the feature values inferred by the trained deep feature value extraction unit 111, and sets the feature value as the image-derived context feature D2. As the attention mechanism, self-attention and cross-attention are used. The self-attention learns a relationship

between elements at different positions in the sequence data and assigns a weight. That is, the self-attention learns similar portions or features in the same image space and assigns a weight. The cross-attention can learn a correlation between different inputs, and can learn similar features between different images provided as input in a case of an image, and can obtain features between a plurality of texts provided as inputs in a global manner in the case of a language. In addition, acquisition of important features of different modalities such as language and an image can also be realized by the cross-attention. Both the self-attention and the cross-attention have a feature that the model itself can automatically acquire a relationship between different elements and automatically learn the importance of information.

[0032]

FIG. 3 is a schematic block diagram showing an example configuration of the natural language processing unit 130 according to the present embodiment. The natural language processing unit 130 includes a trained deep feature value extraction unit 131, a context analysis unit 132, and a prompt generation unit 133. The trained deep feature value extraction unit 131 inputs the query D3 to a trained model such as a text encoder of CLIP, and causes the trained model to infer the feature value. The context analysis unit 132 identifies the feature value having a significant weight by an attention mechanism from among the feature values inferred by the trained deep feature value extraction unit 131, and sets the feature value as the language-derived context feature. As the attention mechanism, self-attention and cross-attention are used.

[0033]

The prompt generation unit 133 inputs the context feature obtained by combining the language-derived context feature and the image-derived context feature D2 to the trained model, and infers the context language which is the text data

representing the context feature. Synthesis of the language-derived context feature and the image-derived context feature D2 may be, for example, element-wise addition or multiplication of the language-derived context feature and the image-derived context feature D2, concatenation thereof as coupling of different tensors, or learning of relationships between elements of different modalities by cross-attention.

[0034]

That is, when the element of the language-derived context feature is denoted by $T1_i$, the element of the image-derived context feature D2 is denoted by $T2_i$, and i is an integer from 0 to N-1 (N is the number of elements of each feature), the combined context feature may be denoted by $T_i = T1_i \times T2_i$ or $T_i = T1_i + T2_i$. The trained model may be, for example, a hierarchical network such as a text decoder corresponding to CLIP. The prompt generation unit 133 combines the query D3, the complementary information D5, the summary sentence D6, and the context language, and generates the generation query D4 which is the text data for instructing the document generation unit 160 to generate the instruction query D10 with reference to the complementary information D5, the summary sentence D6, and the context language, for the material specified by the query D3. Specifically, the prompt generation unit 133 decodes, among the query D3, the complementary information D5, the summary sentence D6, and the context language, those having high-dimensional features into text information, and generates the generation query D4 by textually combining the text information. The query D3, the complementary information D5, the summary sentence D6, and the context language are converted into language information and textually combined, thereby enabling input to the dialogue-based document generation model of the document generation unit 160. For example, when the image recognition unit 110 converts an image into high-dimensional features by using a transformer encoder such as a vision

transformer (ViT), the high-dimensional features are decoded here and restored into language information, thereby being aligned with language information that can be understood by the dialogue-based document generation model.

[0035]

FIG. 4 is a schematic block diagram showing an example configuration of the information refining unit 140 according to the present embodiment. The information refining unit 140 includes an information control unit 141, dialogue-based document generation inference units 142 and 143, an information complementing unit 144, and an information summarization unit 145. The information control unit 141 causes the dialogue-based document generation inference units 142 and 143 to mutually repeat input and output to hold a dialogue. In this case, the information control unit 141 sets the query D3 as the initial input, requests the dialogue-based document generation inference unit 142 to infer the detailed document, and requests the dialogue-based document generation inference unit 143 to infer the concise document. The information control unit 141 may use, for example, language (text) such as “Please provide detailed information regarding the following sentence”, “Please provide specific details regarding the following text”, “Please provide the following information in this context”, “Please explain this in detail”, or “Please provide information supplementing the above content” when requesting the dialogue-based document generation inference unit 142 to infer the detailed document. The information control unit 141 may use language such as “Please concisely summarize the following sentence”, “Please provide the main points of the following text”, “Please provide a summary of this context”, “Please briefly explain this”, or “Please provide the main points of the above content” when requesting the dialogue-based document generation inference unit 142 to infer the concise document.

[0036]

The information control unit 141 may end the repetition of the dialogue after a predetermined number of times, or may end the repetition when it is determined that there is no change in the inference result. The information control unit 141 may extract preference information from the dialogue content by the dialogue-based document generation inference unit 142 or the dialogue-based document generation inference unit 143 and from the storage unit 150. Specifically, when requesting the dialogue-based document generation inference unit 143 or 142 to infer the document, the information control unit 141 adds language (text) such as the following A to I to text output by the dialogue-based document generation inference unit 142 or 143 so that preference information including a preference or a thought pattern of a specific individual is included in the inference result: A. “In order to extract user preference information, please provide the following elements from posts on an SNS: [topics of interest, favorite activities, favorite music or movie genres, and the like]”; B. “Please provide user preference information on an SNS. Specifically, please provide favorite things, topics of interest, and the like”; C. “In order to extract user preference information from an SNS, please provide favorite things, topics of interest, frequently posted content, and the like”; D. “Please describe [hobbies and interests] and explain how they affect daily life”; E. “Please discuss favorite books, movies, and music, and explain why they resonate emotionally”; F. “Please describe the decision-making process when selecting products or services and what factors are most important”; G. “Please describe a subject recently faced and explain how it is addressed based on personal values”; H. “Please imagine an ideal day or a perfect experience, and explain what it would involve and why it is important to you”; and I. “Please discuss a recent achievement or accomplishment and explain what is learned from the experience”.

[0037]

In addition, when requesting the dialogue-based document generation inference unit 143 or 142 to infer a document, the information control unit 141 may add pairs of questions such as J to N below and user answers to the questions so that preference information including a thought pattern of a specific individual is included in the inference result: J. “Please describe a recent experience or thought related to [a specific topic]”; K. “Please describe your opinion or emotion regarding [recent news or an event]”; L. “Please describe a method for addressing a subject in [a specific area or domain]”; M. “What factor most increases your motivation when using ChatGPT (registered trademark) or another AI system?”; and N. “Please imagine an ideal conversation with ChatGPT (registered trademark). What topics would you discuss, and what would you expect from the interaction?”. The wording in the brackets may be stored in advance, or may be extracted by the information control unit 141, the dialogue-based document generation inference unit 143, or the dialogue-based document generation inference unit 142 from the content posted to the SNS stored in the storage unit 150.

[0038]

The preference pattern of the specific individual can also be acquired by the back-end history analysis, and, for example, the acquisition of the thought pattern by the following method is considered.

1) Analysis of chat history: Past chat history of a specific individual can be analyzed to grasp interests and preferences regarding specific topics, characteristic expressions and wording, and tendencies.

2) Natural language processing technology: By utilizing the natural language processing (NLP) technology, it is possible to extract emotions, attitudes, interests, and the like from the sentence of the specific individual.

3) Personalization: By providing personalized responses to specific topics or information based on answers from a specific individual, it becomes easier to grasp a thought pattern of the specific individual.

4) Clustering: Common characteristics and tendencies can be found by using clustering technology that groups individuals having similar answers.

5) Understanding of context: By understanding the context of a conversation of a specific individual and generating a response based thereon, more appropriate dialogue is enabled.

[0039]

Further, the preference information may be extracted based on the language extracted from the query D3. For example, the language included in the parentheses of A or D may be the language extracted from the query D3. When an image or the like is input to the image input unit 100, the information control unit 141 may predict a proper noun of an object in the image by using object recognition or visual question answering (VQA), add a description of the name and context to the query D3 as a document, and cause the name to be written in the parentheses in A above. As a result, the information control unit 141 can automatically designate the preference target. When the query D3 includes a proper noun, a proper noun including preference information or a sentence context may be written in the parentheses in A above. As a result, the preference target is given, and the preference information can be automatically acquired.

[0040]

Here, the preference is, for example, a topic of interest, a favorite activity, a favorite music genre, a favorite movie or drama genre, a food preference, a hobby, a special skill, a preference regarding a tourist spot or travel destination, or a fashion style. In addition, the thought pattern is information indicating a thought tendency likely to be

adopted by a person who is expected to read the generated material, such as a user, and is, for example, information indicating any one of logical thinking, intuitive thinking, concrete thinking, abstract thinking, analytical thinking, and synthetic thinking. A person with logical thinking has characteristics of tending to logically analyze information, make rational judgments, and prefer logical problem solving and argumentation. A person with intuitive thinking has characteristics of tending to process information based on intuition and emotion, make intuitive judgments, and prefer intuitive insight and creative problem solving. A person with concrete thinking has characteristics of thinking based on concrete facts and detailed information, preferring concrete approaches, and placing emphasis on concrete examples and concrete procedures. A person with abstract thinking has characteristics of thinking based on a broad perspective and abstract ideas, and having interest in finding general principles and patterns. A person with analytical thinking has characteristics of analyzing problems, examining information in detail in order to understand fine details and related elements, and placing emphasis on logical reasoning and data analysis. A person with synthetic thinking has characteristics of comprehensively considering a plurality of perspectives and elements, having interest in grasping an overall picture, and seeking integration of information and a comprehensive solution. The SNS may be, for example, any of Facebook (registered trademark), Instagram (registered trademark), Twitter (registered trademark), LINE (registered trademark), or the like.

[0041]

The information control unit 141 may instruct the dialogue-based document generation inference unit 142 or the dialogue-based document generation inference unit 143 to extract the preference information from the storage unit 150. Specifically, the information control unit 141 inputs the text such as A, B, and C to the dialogue-based

document generation inference unit 143 or 142 to extract the preference information of the specific individual from the content posted to the SNS stored in the storage unit 150. When the information control unit 141 gives an instruction to extract the preference information from the storage unit 150, the information control unit 141 may input the content posted to the SNS including the phrase extracted from the query D3 to the dialogue-based document generation inference unit 142 or the dialogue-based document generation inference unit 143. Further, the information control unit 141 may extract the preference information from an external device such as a server of the SNS instead of the storage unit 150.

[0042]

Each of the dialogue-based document generation inference units 142 and 143 inputs the output of the other to the dialogue-based document generation model. The dialogue-based document generation model used by the dialogue-based document generation inference unit 142 and the dialogue-based document generation model used by the dialogue-based document generation inference unit 143 may be different. The information complementing unit 144 sets the preference information as the complementary information D5. In addition, the information complementing unit 144 may combine the preference information extracted from the dialogue content and the preference information extracted from the storage unit 150 to obtain the complementary information D5. The information summarization unit 145 creates the summary of the inference results of the dialogue-based document generation inference units 142 and 143, and sets the summary as the summary sentence D6 and the language answer D9. The summary may be created using a trained model for creating a summary, or the summary may be created by instructing the dialogue-based document generation model to create the summary.

[0043]

FIG. 5 is a schematic block diagram showing an example configuration of the image generation unit 170 according to the present embodiment. The image generation unit 170 includes a trained deep feature value extraction unit 171 and an image decoder unit 172. The trained deep feature value extraction unit 171 inputs the instruction query D10 to the trained model and causes the trained model to infer the feature value. The trained model may be, for example, a hierarchical network such as a text encoder of CLIP. The image decoder unit 172 inputs the feature value of the inference result by the trained deep feature value extraction unit 171 to the trained model and infers the image answer D11, which is an image representing the feature value. The trained model may be, for example, a hierarchical network such as a diffusion decoder corresponding to CLIP.

[0044]

The trained deep feature value extraction unit 171 may identify a feature value having a significant weight from among the feature values that are the inference results, and the image decoder unit 172 may input the identified feature value to the trained model and infer an image representing the identified feature value to obtain the image answer D11.

[0045]

FIG. 6 is a flowchart showing an operation example of the material generation device 10 according to the present embodiment. First, the language input unit 120 receives the input of the query D3 by the user (step S1). Then, the image input unit 100 receives the input of the image D1 by the user (step S2).

[0046]

Next, the information refining unit 140 infers the document requested by the query D3 by using the query D3 input to the language input unit 120 in step S1 as the initial input and causing the two dialogue-based document generation models to mutually repeat input and output to hold a dialogue, and summarizes the inference result to obtain the language answer D9 and the complementary information D5 (step S3). In such a case, the information refining unit 140 may input the preference information of the specific individual read from the storage unit 150 to one of the two dialogue-based document generation models such that the preference information is reflected in the inference result. The information refining unit 140 sets the preference information as the complementary information D5.

[0047]

Next, the image recognition unit 110 inputs the image D1 input to the image input unit 100 in step S2 to the trained model, causes the trained model to infer the feature value, and identifies the feature value having a significant weight from among the feature values that are the inference results to obtain the image-derived context feature D2 (step S4).

[0048]

Next, the natural language processing unit 130 inputs the query D3 to the trained model, causes the trained model to infer the feature value, and identifies the feature value (language-derived context feature) having a significant weight from among the feature values that are the inference results. Further, the natural language processing unit 130 generates the generation query D4 by combining the query D3, the language-derived context feature, the image-derived context feature D2, the complementary information D5, and the summary sentence D6 (step S5).

[0049]

Next, the document generation unit 160 inputs the generation query D4 generated in step S5 to the dialogue-based document generation model, infers the document requested by the generation query D4, and sets the inference result as the instruction query D10 (step S6). Next, the image generation unit 170 inputs the instruction query D10 in step S6 to the trained image generation model, causes the trained image generation model to infer the image specified by the instruction query D10, and sets the inference result as the image answer D11 (step S7).

[0050]

Then, the display processing unit 180 acquires the language answer D9 in step S3 from the information refining unit 140 and the image answer D11 in step S7 from the image generation unit 170, and displays the language answer D9 and the image answer D11 on the display unit 190 (step S8). The display processing unit 180 may output the language answer D9 and the image answer D11 as an electronic file, or may transmit the language answer D9 and the image answer D11 to another device.

[0051]

FIG. 7 is a schematic diagram showing a use case of the material generation device 10 according to the present embodiment. In FIG. 7, text E3, “I want to get a rabbit image with holding a carrot”, is input by the user to the language input unit 120 as the query D3. Further, the user inputs three images E1 of a rabbit holding a carrot to the image input unit 100 as the image D1. Here, the rabbit in the image E1 is gray, pink, or brown.

[0052]

The information refining unit 140 extracts text E8, “She loves white rabbits”, as preference information from a posting history of the user to a social networking service (SNS) stored in the storage unit 150, and inputs the text E8 to either of the dialogue-

based document generation inference units 142 and 143. The dialogue-based document generation inference units 142 and 143 conduct a dialogue E12, such as “What does look like a rabbit?” and “Common rabbits are Lepus Timidus”. Based on these, the information summarization unit 145 generates text E9, “A white rabbit like Lepus Timidus with ears up, holding a tip up carrot and facing forward”, as the language answer D9 or the summary sentence D6.

[0053]

Since the instruction query D10 generated by the document generation unit 160 includes the content of the text E9, the image generation unit 170 generates an image E11 in which a white rabbit holds a carrot, as the image answer D11.

[0054]

<Second embodiment>

Hereinafter, a second embodiment of the present disclosure will be described with reference to the accompanying drawings. In the first embodiment, the preference information including the preference or the thought pattern of the person is used as the information indicating the characteristics of the person. In the second embodiment, information indicating a role of the person is used as the information indicating the characteristics of the person. FIG. 8 is a schematic block diagram showing a configuration of a material generation device 20 according to the second embodiment of the present disclosure. The material generation device 10 is a device that generates a multimodal material in response to a request of the user. The material generation device 10 includes a user interface unit 201, a language analysis unit 202, an image analysis unit 203, a multimodal information integration unit 204, a user information database (DB) unit 205, a provider information DB unit 206, an external information DB unit 207, a hypothesis generation unit 208, an artificial intelligence unit 209, and a display

processing unit 212. The material generation device 10 may be implemented by one or more computers reading and executing a program, and at least a part thereof may be arranged on a so-called cloud.

[0055]

The user interface unit 201 includes input/output devices such as a keyboard, a mouse, a touch panel, a microphone, a display, and a speaker, and provides a user interface for inputting data to the material generation device 10. The data to be input is multimodal data including text data, audio data, image data, document data such as a PDF file, and reference destination data such as a URL. Further, the data to be input includes a query that specifies the generation of the material, user information and provider information that are stored in advance and referred to in a case of generating the material. The user information is information possessed by the user of the material generation service provided by the material generation device 10, and is, for example, know-how possessed by the user. The provider information is information possessed by a provider that provides the material generation service by the material generation device 10, and is, for example, know-how possessed by the provider.

[0056]

The language analysis unit 202 processes the text data input to the user interface unit 201. The text data may be text data included in document data such as a PDF file, or may be text data included in content (Web page content or the like) indicated by the reference destination data. Further, the language analysis unit 202 may perform speech recognition processing on voice data input to the user interface unit 201 and convert the voice data into text data.

[0057]

The image analysis unit 203 analyzes the image data input to the user interface unit 201 to extract the feature value. The image data may be image data included in document data such as a PDF file, or may be image data included in content indicated by the reference destination data.

[0058]

The multimodal information integration unit 204 integrates the feature values of the text data by the language analysis unit 202 and the feature values of the image data by the image analysis unit 203, and converts the integrated feature values into vector representations. The multimodal information integration unit 204 inputs the query converted into the vector representation to the hypothesis generation unit 208. Further, the multimodal information integration unit 204 stores the user information converted into the vector representation in the user information DB unit 205, and stores the provider information converted into the vector representation in the provider information DB unit 206.

[0059]

The user information DB unit 205, the provider information DB unit 206, and the external information DB unit 207 store information related to the material. The user information DB unit 205 is a database that stores the user information converted into the vector representation. The provider information DB unit 206 is a database that stores the provider information converted into the vector representation. The external information DB unit 207 is a database arranged externally, such as on the Internet, or a database that stores data acquired from the outside.

[0060]

The hypothesis generation unit 208 generates the material specified by the query converted into the vector representation input from the multimodal information

integration unit 204. The hypothesis generation unit 208 infers the material specified by the query by causing the plurality of AI agents to hold a dialogue. The hypothesis generation unit 208 includes information indicating the characteristics of a person corresponding to at least one AI agent among the plurality of AI agents, in the dialogue among the plurality of AI agents. The dialogue may be a dialogue between a hypothesis generation side and a criticism (discrimination) side. For example, the hypothesis generation unit 208 may cause one AI agent to generate a hypothesis and cause the other AI agent to criticize or critique the hypothesis. Therefore, the hypothesis generation unit 208 can improve the accuracy of the hypothesis and improve the quality of the generated material. Further, the hypothesis generation unit 208 may cause the AI agent that has generated the hypothesis to correct the hypothesis based on the result of the criticism or critique.

[0061]

When the hypothesis generation unit 208 generates the material, the hypothesis generation unit 208 refers to the information stored in the user information DB unit 205, the provider information DB unit 206, and the external information DB unit 207. Further, when generating the material, the hypothesis generation unit 208 uses an artificial intelligence unit 209 having an artificial intelligence function.

[0062]

The artificial intelligence unit 209 refers to artificial intelligence having intelligent functions such as inference and judgment, and an operating environment thereof. The artificial intelligence unit 209 is configured with a model control unit 210 and a trained model storage unit 211. The artificial intelligence unit 209 is a model configured to output an answer to a prompt when the prompt is input, and an operating environment thereof. When the prompt is input from the hypothesis generation unit

208, the artificial intelligence unit 209 outputs an answer based on the prompt and a trained model described later.

[0063]

The trained model storage unit 211 stores the trained model.

[0064]

The trained model includes information on a model described later. The trained model may include, for example, model parameters that are information defining the behavior of the model, such as a constraint condition, a weighting variable, and an evaluation function.

[0065]

The model may be, for example, a model referred to as a neural network (NN), a convolutional neural network (CNN), a recurrent neural network (RNN), a variational autoencoder (VAE), generative adversarial networks (GAN), a diffusion model, a transformer, a large language model (LLM), a visual language model (VLM), bidirectional encoder representations from transformers (BERT), a generative pre-trained transformer (GPT), or contrastive language-image pre-training (CLIP). The above-described model is not exclusive, and, for example, the LLM, the VLM, the BERT, and the GPT are included in the transformer. Further, for example, the transformer is included in the NN. Furthermore, the learning algorithm and the model may be a combination of a plurality of types. The model also includes a multimodal model that has been trained by combining a plurality of different types of data.

[0066]

When the prompt is acquired, the model control unit 210 outputs an answer corresponding to the prompt based on the prompt and the trained model. That is, when

the prompt is acquired, the model control unit 210 generates and outputs an answer corresponding to the prompt using a model indicated by the trained model.

[0067]

The trained model and other information used by the artificial intelligence unit 209 may be prepared in advance, or may be acquired via the network as necessary.

[0068]

The display processing unit 212 includes a display such as a liquid-crystal display or an organic electro-luminescence (EL) display, and displays the material generated by the hypothesis generation unit 208. In addition, the display included in the display processing unit 212 may be the same as the display included in the user interface unit 201.

[0069]

FIG. 9 is a schematic block diagram showing a configuration of the hypothesis generation unit 208 according to the present embodiment. The hypothesis generation unit 208 includes three AI agent units 301a, 301b, and 301c (AI agents), an AI agent dialogue management unit 302 (dialogue management unit), a dialogue history knowledge DB unit 303, a dialogue strategy DB unit 304 (dialogue strategy storage unit), an output generation unit 305, and an evaluation/verification unit 306. In FIG. 9, the three AI agent units 301a, 301b, and 301c are shown; however, the number of AI agent units included in the hypothesis generation unit 208 may be two or four or more.

[0070]

The three AI agent units 301a, 301b, and 301c correspond to three persons on a one-to-one basis. Hereinafter, any AI agent unit, such as the AI agent unit selected from among the three AI agent units 301a, 301b, and 301c, is also referred to as the AI agent unit 301. The three AI agent units 301a, 301b, and 301c each include, in the dialogue,

person information indicating the characteristics of the corresponding person among the three persons. At least one of the three persons may be a virtual person. The person information of the virtual person may include information indicating a role of the virtual person. The role of the virtual person may be, for example, an excellent business analyst, an excellent consultant of a provider who promotes digitalization, or an executive of the X business department of the user. In a case where the three persons include a specific person who is not virtual, the person information of the specific person may include information indicating a role of the specific person, or may include a work of the specific person or a work related to the specific person.

[0071]

Each of the AI agent units 301a, 301b, and 301c acquires reference information related to the dialogue from the user information DB unit 205, the provider information DB unit 206, the external information DB unit 207, and the dialogue history knowledge DB unit 303, and includes the acquired reference information in the dialogue. For example, the AI agent dialogue management unit 302 designates, for each of the AI agent units 301a, 301b, and 301c, reference information to be acquired from the user information DB unit 205, the provider information DB unit 206, the external information DB unit 207, and the dialogue history knowledge DB unit 303, and a task. Each of the AI agent units 301a, 301b, and 301c acquires the designated reference information from the user information DB unit 205, the provider information DB unit 206, the external information DB unit 207, and the dialogue history knowledge DB unit 303. Each of the AI agent units 301a, 301b, and 301c generates the prompt including the acquired reference information, the designated task, and the corresponding person information, and inputs the generated prompt to the model control unit 210 of the artificial intelligence unit 209. That is, each of the AI agent units 301a, 301b, and 301c requests the model

control unit 210 for an answer in a case where the person indicated by the person information performs the designated task with reference to the acquired reference information. Each of the AI agent units 301a, 301b, and 301c stores the answer from the model control unit 210 in the dialogue history knowledge DB unit 303. The task includes generation of a hypothesis, criticism or critique of the generated hypothesis, and correction of the hypothesis based on the criticism or critique, and includes, for example, “Analyze the reference information and create a detailed report”, “Evaluate feasibility of the report output in the preceding step”, and “Correct the report based on the evaluation of feasibility”. The task may also include designation of a specific format of the answer.

[0072]

The AI agent dialogue management unit 302 causes the AI agent units 301 selected from among the three AI agent units 301a, 301b, and 301c to hold the dialogue in accordance with the dialogue process specified by the dialogue strategy information acquired from the dialogue strategy DB unit 304. The AI agent dialogue management unit 302 inputs the result of the dialogue to the output generation unit 305. The dialogue and the result of the dialogue may be multimodal data including, in addition to text, an image, a diagram, a table, a graph, audio, a moving image, and the like. The dialogue process may include a plurality of steps, and may include the task in each step, the designation of any of the AI agent units 301a, 301b, and 301c, and the designation of the reference information. The AI agent dialogue management unit 302 may cause the three AI agent units 301a, 301b, and 301c to hold the dialogue based on the dialogue strategy information by using the model control unit 210 of the artificial intelligence unit 209. For example, the AI agent dialogue management unit 302 may input, to the model control unit 210, a prompt inquiring about content of a next step to be performed together

with an implementation history of previous steps and the dialogue strategy information, and cause the AI agent unit 301 to hold a dialogue in accordance with an answer corresponding to the prompt.

[0073]

The AI agent dialogue management unit 302 may acquire the dialogue strategy information from the dialogue strategy DB unit 304 based on a query (hereinafter, also referred to as query information) converted into a vector representation input from the multimodal information integration unit 204. For example, the AI agent dialogue management unit 302 may search the dialogue strategy DB unit 304 using the query information and acquire dialogue strategy information as a search result. Further, each step of the dialogue strategy may include acquisition of the dialogue strategy information by searching the dialogue strategy DB unit 304 again and designation of information to be used for the search. The information used for the search is, for example, the query information, the reference information, or the like.

[0074]

The dialogue history knowledge DB unit 303 stores the answer acquired by the AI agent unit 301 from the model control unit 210, as the dialogue history. The dialogue history knowledge DB unit 303 may also store the prompt input to the model control unit 210.

[0075]

The dialogue strategy DB unit 304 stores the dialogue strategy information. The dialogue strategy information may include the purpose of the dialogue strategy and the dialogue process. The dialogue strategy information stored in the dialogue strategy DB unit 304 may be, for example, information obtained by converting, into a vector representation, a sentence in a natural language such as: “Create a proposal for improving

a business process. The dialogue process includes three steps. The first step is to cause an AI agent of a business analyst to collect and analyze the user's business process and submit a report of the result. The format of the report is ...".

[0076]

The output generation unit 305 generates the material based on the result of the dialogue input from the AI agent dialogue management unit 302. The format of the material may be designated by the AI agent dialogue management unit 302 based on the dialogue strategy information. For example, the output generation unit 305 may generate the material by inputting, to the model control unit 210, a prompt for giving an instruction to generate the material together with the result of the dialogue input from the AI agent dialogue management unit 302 and the format of the material, and acquiring the material as an answer corresponding to the prompt. The material may be multimodal data including, in addition to text, an image, a diagram, a table, a graph, and the like.

[0077]

In addition, the output generation unit 305 may correct the material based on a result of evaluation and verification by the evaluation/verification unit 306 for the generated material. For example, the output generation unit 305 may correct the material by inputting, to the model control unit 210, a prompt for giving an instruction to correct the material together with a result of evaluation and verification by the evaluation/verification unit 306 for the generated material, and acquiring the material as an answer corresponding to the prompt. The output generation unit 305 inputs the corrected material to the display processing unit 212. The output generation unit 305 may input the corrected material to the display processing unit 212 after repeating the evaluation and verification by the evaluation/verification unit 306 and the correction a plurality of times.

[0078]

The evaluation/verification unit 306 evaluates and verifies the material input from the output generation unit 305. For example, the evaluation/verification unit 306 may input, to the model control unit 210, the prompt for giving an instruction to evaluate and verify the material together with the material input from the output generation unit 305, and acquire the result of the evaluation and verification as the answer corresponding to the prompt. Further, the evaluation/verification unit 306 may include, for example, person information indicating a proofreader or the like in the prompt. The person information included in the prompt may be changed each time the evaluation and verification are performed.

[0079]

As described above, in the material generation device 20, the plurality of AI agent units 301 correspond to a plurality of persons on a one-to-one basis, and each of the plurality of AI agent units 301 includes, in the dialogue, person information indicating characteristics of a corresponding person among the plurality of persons. Therefore, the material generation device 20 can generate the material corresponding to the characteristics of the plurality of persons.

[0080]

FIG. 10 is an explanatory diagram showing a hardware configuration of the material generation devices 10 and 20 according to the present embodiment.

The material generation devices 10 and 20 include an input/output module I, a storage module M, and a control module P. The input/output module I is implemented by including a part or all of a communication module H11, a connection module H12, a pointing device H21, a keyboard H22, a display H23, a button H3, a microphone H41, a speaker H42, a camera H51, and a sensor H52. The storage module M is implemented

by including a drive H7. The storage module M may further include a part or all of a memory H8. The control module P is implemented by including the memory H8 and a processor H9. These hardware components are connected to each other via the bus to be communicable with each other, and are supplied with power from a power supply H6.

[0081]

The connection module H12 is a digital input/output port such as a Universal Serial Bus (USB). In a case of a portable device, the pointing device H21, the keyboard H22, and the display H23 are touch panels. The sensor H52 is an acceleration sensor, a gyro sensor, a GPS reception module, a proximity sensor, or the like. The power supply H6 is a power supply unit that supplies electricity required to operate each device. In a case of a portable device, the power supply H6 is a battery. The drive H7 is an auxiliary storage medium such as a hard disk drive or a solid-state drive. The drive H7 may be a non-volatile memory such as an EEPROM or a flash memory, or a magneto-optical disk drive or a flexible disk drive. In addition, the drive H7 is not limited to being incorporated in each device, and may be an external storage device connected to a connector of the connection module H12, for example. The memory H8 is a main storage medium such as a random-access memory. In addition, the memory H8 may be a cache memory. When instructions are executed by one or a plurality of processors H9, the memory H8 stores the instructions. The processor H9 is a central processing unit (CPU). The processor H9 may be a microprocessing unit (MPU) or a graphics processing unit (GPU). The processor H9 executes instructions stored in one or more memories H8 by reading a program and various types of data from the drive H7 via the memory H8 and performing computation.

[0082]

In the present specification and the like, the description of the material generation devices 10 and 20 may be replaced with the description of the control module P.

[0083]

The following embodiments may be adopted.

(1) According to an embodiment of the present disclosure, there is provided a material generation device including: an input unit configured to receive input of a query that specifies a material to be generated; and a dialogue management unit configured to infer, based on the query, the material specified by the query by causing a plurality of artificial intelligence (AI) agents to hold a dialogue, in which information indicating characteristics of a person corresponding to at least one AI agent among the plurality of AI agents is included in the dialogue among the plurality of AI agents.

[0084]

Therefore, the material generation device can generate the material corresponding to the characteristics of the person.

[0085]

(2) According to the embodiment of the present disclosure, in the material generation device according to (1), the plurality of AI agents are two dialogue-based document generation models, the input unit is a language input unit configured to receive the input of the query, the dialogue management unit is an information refining unit configured to infer a document in the material specified by the query by using the query as initial input and causing the two dialogue-based document generation models to mutually repeat input and output to hold a dialogue, the material generation device further comprises an image generation unit configured to generate an image based on the document that is an inference result by the information refining unit, and preference

information including a preference or a thought pattern of a specific individual is added to the dialogue among the two dialogue-based document generation models.

[0086]

Therefore, the material generation device can generate the material in which the preference or the thought pattern of the specific individual is reflected.

[0087]

(3) According to the embodiment of the present disclosure, the material generation device according to (2) further includes an image input unit configured to receive input of an image serving as an example of the material to be generated, and an image recognition unit configured to input the image input to the image input unit to a trained model, cause the trained model to infer a feature value of the image input to the image input unit, and set the feature value as an image-derived context feature, and the image generation unit is configured to generate an image based on the image-derived context feature, in addition to the document that is the inference result by the information refining unit.

[0088]

Therefore, the material generation device can generate the image in which the image that is the example of the material is reflected.

[0089]

(4) According to the embodiment of the present disclosure, the material generation device according to (3) further includes a natural language processing unit configured to input the query to a trained model, cause the trained model to infer a feature value of the query, set the feature value as a language-derived context feature, and generate a generation query in which at least some of the query, the language-derived context feature, the image-derived context feature, the preference information, and the

document that is the inference result by the information refining unit are combined, and a document generation unit configured to input the generation query to the dialogue-based document generation model, infer a document requested by the generation query, and set an inference result as an instruction query, and the image generation unit is configured to input the instruction query to a trained image generation model and cause the trained image generation model to infer an image specified by the instruction query.

[0090]

Therefore, the material generation device can generate the image in which the preference or the thought pattern of the specific individual and the image that is the example of the material are reflected.

[0091]

(5) According to the embodiment of the present disclosure, in the material generation device according to (1), the dialogue among the plurality of AI agents includes generation of a hypothesis by one AI agent and criticism or critique of the generated hypothesis by another AI agent.

[0092]

Therefore, the material generation device can improve the quality of the material to be generated.

[0093]

(6) According to the embodiment of the present disclosure, in the material generation device according to (1) or (5), the plurality of AI agents correspond to a plurality of persons on a one-to-one basis, and each of the plurality of AI agents includes, in the dialogue, person information indicating characteristics of a corresponding person among the plurality of persons.

[0094]

Therefore, the material generation device can generate the material corresponding to the characteristics of the plurality of persons.

[0095]

(7) According to the embodiment of the present disclosure, in the material generation device according to (6), at least one of the plurality of persons is a virtual person, and the person information of the virtual person includes information indicating a role of the virtual person.

[0096]

Therefore, the material generation device can generate the material corresponding to the role of the virtual person.

[0097]

(8) According to the embodiment of the present disclosure, the material generation device according to any one of (1) to (7) further includes a dialogue strategy storage unit configured to store dialogue strategy information indicating a dialogue process and persons to participate in a dialogue, and the dialogue management unit is configured to cause the AI agents selected from among the plurality of AI agents to hold the dialogue in accordance with the dialogue process specified by the dialogue strategy information acquired from the dialogue strategy storage unit.

[0098]

Therefore, the material generation device can make the dialogue process flexible.

[0099]

(9) According to the embodiment of the present disclosure, the material generation device according to any one of (1) to (8) further includes a storage unit configured to store information related to the material, and each of the plurality of AI

agents is configured to acquire reference information related to the dialogue from the storage unit and include the acquired reference information in the dialogue.

[0100]

Therefore, the material generation device can generate the material based on the reference information related to the dialogue.

[0101]

(10) According to the embodiment of the present disclosure, in the material generation device according to any one of (1) to (9), the input unit is configured to receive input of multimodal data related to the material, the dialogue management unit is configured to cause the plurality of AI agents to hold the dialogue based on the query and the multimodal information, and the dialogue among the plurality of AI agents is multimodal.

[0102]

Therefore, the material generation device can generate the material based on the multimodal information.

[0103]

(11) According to the embodiment of the present disclosure, the material generation device according to any one of (1) to (10) further includes an output generation unit configured to generate the material based on a result of the dialogue among the plurality of AI agents, and an evaluation/verification unit configured to evaluate and verify the material generated by the output generation unit.

[0104]

Therefore, the material generation device can generate the material that is evaluated and verified.

[0105]

(12) According to another embodiment of the present disclosure, there is provided a material generation method including: a first step of receiving input of a query that specifies a material to be generated; and a second step of inferring, based on the query, the material specified by the query by causing a plurality of AI agents to hold a dialogue, in which information indicating characteristics of a person, including a virtual person, corresponding to at least one AI agent among the plurality of AI agents is added to the dialogue among the plurality of AI agents.

[0106]

(13) According to still another embodiment of the present disclosure, there is provided a program causing a computer to function as: an input unit configured to receive input of a query that specifies a material to be generated; and a dialogue management unit configured to infer, based on the query, the material specified by the query by causing a plurality of AI agents to hold a dialogue, in which information indicating characteristics of a person, including a virtual person, corresponding to at least one AI agent among the plurality of AI agents is added to the dialogue among the plurality of AI agents.

[0107]

A program for implementing the functions of the material generation device 10 in FIG. 1 may be recorded on a computer-readable recording medium, and the program recorded on the recording medium may be read into a computer system and executed to implement the material generation device 10. The term “computer system” used herein encompasses an OS and hardware such as a peripheral device.

[0108]

The term “computer-readable recording medium” refers to a portable medium such as a flexible disk, a magneto-optical disk, a ROM, or a CD-ROM, or a storage

device such as a hard disk built into the computer system. Further, the term “computer-readable recording medium” also encompasses a medium that dynamically holds a program for a short period of time, such as a communication line when transmitting a program via the network such as the Internet or a communication line such as a telephone line, and a medium that holds a program for a certain period of time, such as a volatile memory inside a computer system that is a server or a client in such a case. Furthermore, the above program may be a program for implementing a part of the above functions, and the above functions may be implemented in combination with a program already recorded in the computer system.

[0109]

Although the embodiments of the present disclosure have been described in detail with reference to the accompanying drawings, the specific configuration is not limited to the embodiments, and includes design changes and the like within a scope not departing from the gist of the present disclosure.

REFERENCE SIGNS LIST

[0110]

10, 20 Material generation device

100 Image input unit

110 Image recognition unit

111 Trained deep feature value extraction unit

112 Context analysis unit

120 Language input unit

130 Natural language processing unit

131 Trained deep feature value extraction unit

132 Context analysis unit

133 Prompt generation unit

140 Information refining unit

141 Information control unit

142, 143 Dialogue-based document generation inference unit

144 Information complementing unit

145 Information summarization unit

150 Storage unit

160 Document generation unit

170 Image generation unit

171 Trained deep feature value extraction unit

172 Image decoder unit

180 Display processing unit

190 Display unit

201 User interface unit

202 Language analysis unit

203 Image analysis unit

204 Multimodal information integration unit

205 User information DB unit

206 Provider information DB unit

207 External information DB unit

208 Hypothesis generation unit

209 Artificial intelligence unit

210 Model control unit

211 Trained model storage unit

212 Display processing unit

301a, 301b, 301c AI agent unit

302 AI agent dialogue management unit

303 Dialogue history knowledge DB unit

304 Dialogue strategy DB unit

305 Output generation unit

306 Evaluation/verification unit

CLAIMS

[Claim 1] A material generation device comprising:

an input unit configured to receive input of a query that specifies a material to be generated; and

a dialogue management unit configured to infer, based on the query, the material specified by the query by causing a plurality of artificial intelligence (AI) agents to hold a dialogue,

wherein information indicating characteristics of a person corresponding to at least one AI agent among the plurality of AI agents is included in the dialogue among the plurality of AI agents.

[Claim 2] The material generation device according to Claim 1,

wherein the plurality of AI agents are a plurality of dialogue-based document generation models,

the input unit is a language input unit configured to receive the input of the query,

the dialogue management unit is an information refining unit configured to infer a document in the material specified by the query by using the query as initial input and causing the plurality of dialogue-based document generation models to mutually repeat input and output to hold a dialogue,

the material generation device further comprises an image generation unit configured to generate an image based on the document that is an inference result by the information refining unit, and

preference information including a preference or a thought pattern of a specific individual is added to the dialogue among the plurality of dialogue-based document generation models.

[Claim 3] The material generation device according to Claim 2, further comprising:

an image input unit configured to receive input of an image serving as an example of the material to be generated; and

an image recognition unit configured to input the image input to the image input unit to a trained model, cause the trained model to infer a feature value of the image input to the image input unit, and set the feature value as an image-derived context feature,

wherein the image generation unit is configured to generate an image based on the image-derived context feature, in addition to the document that is the inference result by the information refining unit.

[Claim 4] The material generation device according to Claim 3, further comprising:

a natural language processing unit configured to input the query to a trained model, cause the trained model to infer a feature value of the query, set the feature value as a language-derived context feature, and generate a generation query in which at least some of the query, the language-derived context feature, the image-derived context feature, the preference information, and the document that is the inference result by the information refining unit are combined; and

a document generation unit configured to input the generation query to the dialogue-based document generation model, infer a document requested by the generation query, and set an inference result as an instruction query,

wherein the image generation unit is configured to input the instruction query to a trained image generation model and cause the trained image generation model to infer an image specified by the instruction query.

[Claim 5] The material generation device according to Claim 1,

wherein the dialogue among the plurality of AI agents includes generation of a hypothesis by one AI agent and criticism or critique of the generated hypothesis by another AI agent.

[Claim 6] The material generation device according to Claim 1,

wherein the plurality of AI agents correspond to a plurality of persons on a one-to-one basis, and

each of the plurality of AI agents includes, in the dialogue, person information indicating characteristics of a corresponding person among the plurality of persons.

[Claim 7] The material generation device according to Claim 6,

wherein at least one of the plurality of persons is a virtual person, and

the person information of the virtual person includes information indicating a role of the virtual person.

[Claim 8] The material generation device according to Claim 1, further comprising:

a dialogue strategy storage unit configured to store dialogue strategy information indicating a dialogue process and persons to participate in a dialogue,

wherein the dialogue management unit is configured to cause the AI agents selected from among the plurality of AI agents to hold the dialogue in accordance with the dialogue process specified by the dialogue strategy information acquired from the dialogue strategy storage unit.

[Claim 9] The material generation device according to Claim 1, further comprising:

a storage unit configured to store information related to the material,

wherein each of the plurality of AI agents is configured to acquire reference information related to the dialogue from the storage unit and include the acquired reference information in the dialogue.

[Claim 10] The material generation device according to Claim 1,

wherein the input unit is configured to receive input of multimodal data related to the material,

the dialogue management unit is configured to cause the plurality of AI agents to hold the dialogue based on the query and the multimodal information, and

the dialogue among the plurality of AI agents is multimodal.

[Claim 11] The material generation device according to Claim 1, further comprising:

an output generation unit configured to generate the material based on a result of the dialogue among the plurality of AI agents; and

an evaluation/verification unit configured to evaluate and verify the material generated by the output generation unit.

[Claim 12] A material generation method comprising:

a first step of receiving input of a query that specifies a material to be generated; and

a second step of inferring, based on the query, the material specified by the query by causing a plurality of AI agents to hold a dialogue,

wherein information indicating characteristics of a person, including a virtual person, corresponding to at least one AI agent among the plurality of AI agents is added to the dialogue among the plurality of AI agents.

[Claim 13] A program causing a computer to function as:

an input unit configured to receive input of a query that specifies a material to be generated; and

a dialogue management unit configured to infer, based on the query, the material specified by the query by causing a plurality of AI agents to hold a dialogue,

wherein information indicating characteristics of a person, including a virtual person, corresponding to at least one AI agent among the plurality of AI agents is added to the dialogue among the plurality of AI agents.