DESCRIPTION

TITLE OF THE INVENTION: INFORMATION PROCESSING SYSTEM, INFORMATION PROCESSING METHOD, AND PROGRAM

TECHNICAL FIELD

[0001] The present disclosure relates to an information processing system, an information processing method, and a program.

BACKGROUND ART

[0002] A technique of having a dialogue with a person is known. For example, there is a technique of acquiring an answer provided by a job applicant to a predetermined question, numerically evaluating an interview result of the job applicant based on the answer provided by the job applicant and evaluation criteria information, and displaying score information as information of the interview result.

RELATED ART DOCUMENTS

PATENT DOCUMENTS

[0003] Patent Document 1: Japanese Patent Application Publication No. 2023-31089

SUMMARY OF THE INVENTION

PROBLEMS TO BE SOLVED BY THE INVENTION

[0004] The present disclosure provides a technique of controlling a device configured to have a dialogue with a user.

MEANS FOR SOLVING THE PROBLEMS

[0005] An information processing system according to one aspect of the present disclosure includes: at least one memory; and at least one processor. The at least one processor is configured to repeatedly execute, until a predetermined condition is satisfied: acquisition of information relating to an utterance of a user; generation of information relating to an utterance of a dialogue device based on the information relating to the utterance of the user and a machine learning model; and control of the utterance of the dialogue

device based on the information relating to the utterance of the dialogue device. The generation of the information relating to the utterance of the dialogue device is executed based on information indicating a relationship between the user and the dialogue device. The relationship includes at least one of: a relationship in which the user provides a service or a product to the dialogue device; a relationship in which the user instructs the dialogue device; a relationship in which the dialogue device pays remuneration to the user; or a relationship in which the user having information A provides the information A to a dialogue device that does not have the information A.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] [FIG. 1] FIG. 1 is a block diagram illustrating an example of an overall configuration of a dialogue system according to a first embodiment.

[FIG. 2] FIG. 2 is a block diagram illustrating an example of a functional configuration of a control device according to the first embodiment.

[FIG. 3] FIG. 3 is a process flow illustrating an example of a dialogue method according to the first embodiment.

[FIG. 4] FIG. 4 is a block diagram illustrating an example of an overall configuration of a dialogue system according to a second embodiment.

[FIG. 5] FIG. 5 is a block diagram illustrating an example of a functional configuration of a management device according to the second embodiment.

[FIG. 6] FIG. 6 is a process flow illustrating an example of a dialogue method according to the second embodiment.

[FIG. 7] FIG. 7 is a block diagram illustrating an example of a dialogue system according to an applied example.

[FIG. 8] FIG. 8 is a block diagram illustrating an example of a hardware configuration of a computer.

DETAILED DESCRIPTION OF THE INVENTION

[0007] Hereinafter, embodiments of the present disclosure will be described with reference to the accompanying drawings. In the present specification and drawings, components having substantially the same functional configuration are indicated by the same reference signs, and duplicated descriptions thereof will be omitted.

[0008] [First Embodiment]

A first embodiment of the present disclosure is an example of an information processing system configured to have a dialogue with a user. Hereinafter, the information processing system according to the present embodiment will be referred to as a "dialogue system". The dialogue system may execute the dialogue with the user by controlling a dialogue device, which will be a dialogue partner of the user.

[0009] The dialogue system has a function of outputting an evaluation result relating to a user based on dialogue data recording a dialogue between the user and the dialogue device. As an example, the dialogue system may have a dialogue with a user by using the dialogue device according to a specific scenario, and output an evaluation result acquired by evaluating the user according to predetermined evaluation criteria based on dialogue data recording a dialogue between the user and the dialogue device. As an example, the scenario of the dialogue may be a scenario in which the user is a clerk in a retail store, the dialogue device is a customer, and the user performs a task corresponding to a request of the customer. When this type of scenario is used, the customer-service skills of the user can be evaluated, as an example.

[0010] The present embodiment provides a technique of controlling a device configured to have a dialogue with a user. In the present embodiment, information relating to a first utterance uttered by the user is acquired, information relating to a second utterance is generated based on information relating to the first utterance, input information, and a machine learning model, an utterance of a dialogue device is controlled based on the information relating to the second utterance, and the input information includes information indicating a relationship between the user and the dialogue device.

[0011] In one aspect, according to the present embodiment, since information relating to an utterance of the dialogue device is generated based on the machine learning model, a device configured to have a dialogue with a user can be controlled. In another aspect, according to the present embodiment, information for evaluating the user can be acquired through a dialogue between the user and the dialogue device.

[0012] <Overall Configuration of Dialogue System>

An overall configuration of the dialogue system according to the present embodiment will be described with reference to FIG. 1. FIG. 1 is a block

diagram illustrating an example of the overall configuration of the dialogue system according to the first embodiment.

[0013] As illustrated in FIG. 1, a dialogue system 1000 includes a dialogue device 10, a control device 20, a generation device 30, and a terminal device 40. The dialogue device 10, the control device 20, the generation device 30, and the terminal device 40 may be connected to each other via a communication network to enable data communication. The communication network may be, for example, a LAN (Local Area Network), a WAN (Wide Area Network), a VPN (Virtual Private Network), or any other network, such as the Internet or the like.

[0014] The dialogue device 10 is an example of a device configured to have a dialogue with a dialogue participant S, which is an example of the user of the dialogue system 1000. The dialogue device 10 may receive an utterance of the dialogue participant S. The dialogue device 10 may present, to the dialogue participant S, an utterance of the dialogue device 10. The dialogue device 10 may have a dialogue with the dialogue participant S by repeating reception of the utterance of the dialogue participant S and presentation of the utterance of the dialogue device 10. Hereinafter, the utterance of the dialogue participant S will be also referred to as a “user utterance”. The utterance of the dialogue device 10 will be also referred to as a “system utterance”.

[0015] The dialogue device 10 may be an information processing terminal, such as a personal computer, a smartphone, a tablet terminal, or the like. The dialogue device 10 may execute a dialogue with the dialogue participant S, for example, by displaying a CG (Computer Graphics) character on a display device and controlling the CG character. The CG character may be, for example, a virtual human referred to as an avatar or the like. The CG character is not limited to a human, and may be a real organism, a fictitious organism, a robot, an anthropomorphic object, or the like. In this case, the dialogue participant S has a dialogue with the virtual CG character displayed on the display device.

[0016] The dialogue device 10 may be a robot. The dialogue device 10 may execute a dialogue with the dialogue participant S by operating the robot. As an example, the robot may be a humanoid, which imitates a human. The robot is not limited to a human, and may be a real organism, a fictitious organism, an anthropomorphic object, or the like. In this case, the dialogue participant S has a dialogue with the robot as a physical entity.

[0017] The dialogue device 10 may generate information relating to the user utterance and transmit the generated information to the control device 20. The dialogue device 10 may receive information relating to the system utterance from the control device 20 and present the system utterance to the dialogue participant S. Hereinafter, the information relating to the user utterance will be also referred to as "user utterance information". The information relating to the system utterance will be also referred to as "system utterance information".

[0018] The user utterance information may include at least one of a speech signal acquired by collecting a user utterance, text data indicating the utterance content of the user utterance, or a video signal or image signal acquired by capturing the face or body of the dialogue participant S. The dialogue device 10 may generate a speech signal by collecting a user utterance using a microphone. The dialogue device 10 may generate a video signal or image signal by capturing the face or body of the dialogue participant S using a camera. The dialogue device 10 may generate text data indicating the utterance content by performing speech recognition of a speech signal acquired by collecting a user utterance.

[0019] The system utterance information may include at least one of a speech signal generated by synthesizing a system utterance, text data indicating utterance contents of a system utterance, a video signal or image signal presenting a CG character, or a control signal for controlling the expression of the dialogue device 10. The dialogue device 10 may present the system utterance to the dialogue participant S by outputting, using a speaker, a speech signal indicating the system utterance. The dialogue device 10 may generate a speech signal indicating the system utterance by synthesizing speech from text data indicating the utterance content of the system utterance. The dialogue device 10 may present the system utterance information to the dialogue participant S by displaying, on a display device, a video or image presenting a CG character. When the dialogue device 10 is a robot, the system utterance information may be presented to the dialogue participant S by moving the face or body of the robot.

[0020] The control device 20 is an example of an information processing device, such as a personal computer, a workstation, a server, or the like, which is configured to control a dialogue between the dialogue participant S and the dialogue device 10. The control device 20 may acquire the user utterance

information from the dialogue device 10. The control device 20 may generate the system utterance information based on the user utterance information. The control device 20 may control the dialogue device 10 to utter the system utterance based on the system utterance information.

[0021] The control device 20 may store dialogue data D. The dialogue data D may be electronic data recording a dialogue between the dialogue participant S and the dialogue device 10. The dialogue data D may include at least part of the user utterance information and the system utterance information. The dialogue data D may be stored for each dialogue between the dialogue participant S and the dialogue device 10. The dialogue data D may include identification information capable of specifying a dialogue. An example of the identification information capable of specifying a dialogue may be time information indicating a time when the dialogue is executed. The time information may be, for example, a time when the dialogue starts or a time when the dialogue ends.

[0022] The control device 20 may include an evaluation model M1. The evaluation model M1 is an example of a machine learning model trained to output one or more evaluation values based on the dialogue data D. The evaluation model M1 may use, as an input, at least part of the dialogue data D, thereby outputting one or more evaluation values. The evaluation model M1 may use, as an input, one or more features extracted from the dialogue data D, thereby outputting one or more evaluation values.

[0023] The evaluation model M1 is, for example, a neural network, a decision tree, a random forest, a gradient boosting decision tree, a support vector machine (SVM), a large language model (LLM), a foundation model, or the like. The gradient boosting decision tree is, for example, a Light Gradient Boosting Machine (LightGBM), eXtreme Gradient Boosting (XGBoost), or the like.

[0024] The generation device 30 is an example of an information processing device, such as a personal computer, a workstation, a server, or the like, which is configured to execute a predetermined task based on a machine learning model. The generation device 30 may execute a task of generating the system utterance in response to a generation request from the control device 20. The generation device 30 may transmit, to the control device 20, the generation result of the system utterance.

[0025] The generation device 30 may include a generative model M2. The generative model M2 is an example of a machine learning model configured to

execute a predetermined task. The generative model M2 may be a machine learning model capable of generating various data, such as text, speech, image, video, and the like. The generative model M2 may be, for example, a neural network, a large language model, a generative model, a foundation model, or the like. The generative model M2 may support multimodality.

[0026] The generative model M2 may be realized by a single machine learning model. The generative model M2 may be realized by cooperation of a plurality of machine learning models. The generative model M2 may be configured by a plurality of machine learning models each corresponding to a task to be executed.

[0027] The terminal device 40 is an example of an information processing terminal, such as a personal computer, a smartphone, a tablet terminal, or the like, operated by an evaluator E, which is an example of the user of the dialogue system 1000. The evaluator E is an entity that evaluates the dialogue participant S. As an example, the evaluator E may be a person in charge of evaluating the dialogue participant S to judge whether to hire the dialogue participant S, who is a job applicant. As an example, the evaluator E may be a person in charge of performing personnel evaluation of the dialogue participant S, who is an employee. As an example, the evaluator E may be a training instructor. As an example, the evaluator E may be a grader of an examination.

[0028] The terminal device 40 may acquire, from the control device 20, the evaluation result relating to the dialogue participant S. The terminal device 40 may present, to the evaluator E, the evaluation result relating to the dialogue participant S. As an example, the terminal device 40 may display the evaluation result on the display device of the terminal device 40. As an example, the terminal device 40 may output, from the speaker of the terminal device 40, speech synthesized from the evaluation result.

[0029] Presenting information to a user of the dialogue system 1000 may include a processor executing at least part of processing necessary for displaying the information on the display device. The display device may be provided in the same device in which the processor is provided or may be provided in a device different from that in which the processor is provided. The display device may include a plurality of display devices.

[0030] Note that the overall configuration of the dialogue system 1000

illustrated in FIG. 1 is merely an example, and there can be various system configuration examples according to applications and purposes. The dialogue system 1000 may be configured by one or more devices. Each device included in the dialogue system 1000 may be a system configured by a plurality of devices. Each function included in the dialogue system 1000 may be realized by any device constituting the system. Each component included in the dialogue system 1000 may be included in any device constituting the system.

[0031] The dialogue system 1000 may include multiple instances of one or more of the dialogue device 10, the control device 20, the generation device 30, and the terminal device 40. The control device 20 or the generation device 30 may be realized by a plurality of computers or may be realized as a cloud computing service. The control device 20 and the generation device 30 may be realized by an integrated stand-alone computer. The categorization of the devices illustrated in FIG. 1, i.e., the dialogue device 10, the control device 20, the generation device 30, and the terminal device 40, is merely an example.

[0032] The dialogue device 10 may be any device that has a dialogue with the dialogue participant S, and may be any of the following as a non-limiting example. The dialogue device 10 may be one or more information processing devices. The dialogue device 10 may be an information processing system including a plurality of information processing devices. The dialogue device 10 may include a terminal device including a display device. The dialogue device 10 may be a CG character, such as a virtual human or the like, controlled by one or more hardware components. The dialogue device 10 may be realized by another device included in the dialogue system 1000. The dialogue device 10 may be realized by an information processing device or information processing system external to the dialogue system 1000. Note that “external” means being not included in the dialogue system 1000.

[0033] As an example, the dialogue system 1000 may be configured by one or more server devices and one or more client devices. The one or more server devices may include one or more of the functions of the control device 20 and the generation device 30. The one or more client devices may include one or more of the functions of the dialogue device 10 and the terminal device 40. The server device may be realized by a system including a plurality of information processing devices. The server device may be realized as a cloud computing service.

[0034] As another example, the dialogue system 1000 may be configured by a single information processing device. The single information processing device may include the functions of the dialogue device 10, the control device 20, the generation device 30, and the terminal device 40.

[0035] <Functional Configuration of Control Device>

A functional configuration of the control device 20 will be described with reference to FIG. 2. FIG. 2 is a block diagram illustrating an example of the functional configuration of the control device according to the first embodiment.

[0036] As illustrated in FIG. 2, the control device 20 includes a scenario storage unit 101, a dialogue storage unit 102, a model storage unit 103, an evaluation result storage unit 104, an utterance acquisition unit 110, an utterance generation unit 120, an utterance presentation unit 130, a dialogue evaluation unit 140, and a result output unit 150. When a control program installed in advance is executed, the control device 20 functions as the scenario storage unit 101, the dialogue storage unit 102, the model storage unit 103, the evaluation result storage unit 104, the utterance acquisition unit 110, the utterance generation unit 120, the utterance presentation unit 130, the dialogue evaluation unit 140, and the result output unit 150.

[0037] (Scenario Storage Unit)

The scenario storage unit 101 stores scenario information. The scenario information is information indicating a scenario of a dialogue. The scenario information may include information indicating a situation of a dialogue. The information indicating the situation of the dialogue may include, for example, a relationship between the dialogue participant S and the dialogue device 10, a prerequisite of a dialogue, a constraint for an utterance, and the like. The relationship between the dialogue participant S and the dialogue device 10 may include a relationship between the dialogue participant S and a CG character controlled by the dialogue device 10 or a robot as the dialogue device 10. The relationship between the dialogue participant S and the dialogue device 10 may include a role of the dialogue participant S and a role of the dialogue device 10.

[0038] The relationship between the dialogue participant S and the dialogue device 10 may include: at least one of a relationship in which the dialogue participant S provides a service or product to the dialogue device 10, a relationship in which the dialogue participant S instructs the dialogue device

10, a relationship in which the dialogue device 10 pays remuneration to the dialogue participant S, or a relationship in which the dialogue participant S provides information to the dialogue device 10; or a combination of at least two of the foregoing. Specifically, the relationship between the dialogue participant S and the dialogue device 10 may include at least one of a relationship between a clerk in a retail store and a customer, a relationship between a superior and a subordinate, a relationship between a teacher and a student, a relationship between a speaker and an audience, a relationship between a doctor and a patient, or a relationship between a service professional and a client. As an example, the service professional may include an expert, such as a lawyer, a consultant, or the like.

[0039] As an example, the scenario information may be a scenario relating to a dialogue between a clerk and a customer in a retail store or a restaurant. In this case, the role of the dialogue participant S may be any of “client”, “guest”, “customer”, “valued customer”, “visitor”, and the like. The role of the dialogue device 10 may be any of “clerk”, “staff”, “store person”, and the like.

[0040] The scenario information may include one or more system prompts and one or more pieces of guidance information. The system prompts may be information used for causing the generative model M2 to generate the system utterance in accordance with the scenario. The guidance information may be information used for explaining the scenario to the dialogue participant S at the start of the dialogue. The guidance information may be associated with a corresponding system prompt. The guidance information may include information indicating a relationship between the dialogue participant S and the dialogue device 10.

[0041] The system prompt may include information indicating the situation of a dialogue. The system prompt may be generated in advance based on the scenario of a dialogue. The system prompt may include information indicating a relationship between the dialogue participant S and the dialogue device 10. The system prompt may include a prerequisite of the dialogue. The system prompt may include a constraint for the utterance.

[0042] When the system prompt includes information indicating a relationship between the dialogue participant S and the dialogue device 10, the system prompt may include at least information relating to the role of the dialogue device 10, and need not explicitly include information relating to the

role of the dialogue participant S. The system prompt may include information indirectly indicating the role of the dialogue participant S. The role of the dialogue device 10 may include at least one of a customer, a subordinate, a student, an audience, a patient, or a client.

[0043] The system prompt may be, as an example, a character string shown below:

-"You are a male customer who came to a store handling clothes or general goods. You will buy clothes while being asked by a store clerk. You have not decided what clothes you want to buy. You prefer a simple design, blue cotton material, and everyday use, and your budget is 50,000 yen. Please share the information in small doses. You can tell that you are looking for clothes. Please output only the utterance of the customer in 50 characters or less".

[0044] Here, "customer" and "store clerk" are examples of information indicating the relationship between the dialogue device 10 and the dialogue participant S. Note that "customer" is the role of the dialogue device 10, and "store clerk" is the role of the dialogue participant S. "Please output only the utterance of the customer in 50 characters or less" is an example of the constraint for the utterance. The other parts are examples of the prerequisite of the dialogue.

[0045] As another example, the system prompt may be a character string shown below. The below character string is a non-limiting example of the system prompt including information indirectly indicating the role of the dialogue participant S.

-"You are a male customer who came to a store handling clothes or general goods. You will buy some clothes. You have not decided what clothes you want to buy. You prefer a simple design, blue cotton material, and everyday use, and your budget is 50,000 yen. Please share the information in small doses. You can tell that you are looking for clothes. Please output only the utterance of the customer in 50 characters or less".

[0046] Here, "customer" is an example of information relating to the role of the dialogue device 10. Furthermore, "customer" is an example of information indirectly indicating the role of the dialogue participant S. Because the role of the dialogue device 10 is indicated to be "customer", it is indirectly indicated that the role of the dialogue participant S is "store clerk".

[0047] The scenario information may include a plurality of system prompts.

The plurality of system prompts may be, for example, system prompts corresponding to the degree of progress of the dialogue. The degree of progress of the dialogue may be, for example, the number of turns in the dialogue or the elapsed time from the start of the dialogue.

[0048] The guidance information may include content relating to the system prompt. The guidance information may include information indicating a relationship between the dialogue participant S and the dialogue device 10. The information indicating the relationship between the dialogue participant S and the dialogue device 10 may be content that is the same as the content of the system prompt. The amount of information of the guidance information may be less than the amount of information of the system prompt. The guidance information may be information in which at least part of the prerequisite relating to the dialogue device 10 is omitted from the system prompt. The guidance information may include information instructing the start of a dialogue. The information instructing the start of a dialogue may be positioned at the end of the guidance information.

[0049] The guidance information may be, for example, a character string shown below:

-“You are a store clerk at a clothing store. I am standing as a customer, so please talk to me as if you were a clerk. Please begin.”

[0050] Here, “store clerk” and “customer” are an example of information indicating the relationship between the dialogue participant S and the dialogue device 10. Note that “customer” is the role of the dialogue device 10, and “store clerk” is the role of the dialogue participant S. “Please begin” is an example of the information indicating the start of a dialogue.

[0051] The scenario storage unit 101 may store a plurality of pieces of scenario information. As an example, the plurality of pieces of scenario information may include scenario information relating to a scenario according to a business type or a job type. The plurality of pieces of scenario information may be stored in the scenario storage unit 101 in association with identification information indicating a scenario. As a method for associating the identification information and the scenario information with each other, for example, the identification information and the scenario information may be stored as one set. Alternatively, information capable of acquiring, from one piece of information, the other piece of information may be stored.

Alternatively, one piece of information and the identification information of the other piece of information may be stored as one set.

[0052] The scenario information may be created by the evaluator E. The scenario information may be created based on a template created in advance. As an example, the scenario information may be created by changing a proper name (e.g., store name, product name, or service name) in the template. The scenario information selected by the evaluator E among a plurality of pieces of scenario information created in advance may be stored in the scenario storage unit 101.

[0053] (Dialogue Storage Unit)

The dialogue data D is stored in the dialogue storage unit 102. The dialogue data D may be stored for each dialogue between the dialogue participant S and the dialogue device 10. The dialogue data D may include the user utterance information and the system utterance information acquired by the control device 20 from the start to end of a dialogue between the dialogue participant S and the dialogue device 10. The utterance acquisition unit 110 may acquire the user utterance information included in the dialogue data D. The utterance generation unit 120 may generate the system utterance information included in the dialogue data D.

[0054] As an example, the dialogue data D may include at least one of time information, attribute information of the dialogue participant S, speech information of the dialogue participant S, image information of the dialogue participant S, speech information of the dialogue device 10, image information of the dialogue device 10, control information of the dialogue device 10, video data of the overall dialogue, and log information of the dialogue system 1000.

[0055] The attribute information of the dialogue participant S may include a profile of the dialogue participant S. As an example, the profile may include information contained in a resume or a work-history document. The attribute information of the dialogue participant S may be set in advance or may be input by the dialogue participant S at the start of a dialogue.

[0056] The speech information of the dialogue participant S may include a speech signal acquired by collecting speech of the dialogue participant S. The speech information of the dialogue participant S may include text data acquired by recognizing the speech signal. The image information of the dialogue participant S may include a still image or moving image acquired by capturing

at least part of the face or body of the dialogue participant S. The image information of the dialogue participant S may include information recognized from the image of the dialogue participant S. As an example, the image information of the dialogue participant S may include a facial expression label, posture information, or the like. The facial expression label is a label indicating a class into which the facial expression of a person is classified. The posture information is information indicating the position of a characteristic point of a person. For example, the posture may be estimated by acquiring each joint position using a posture estimation technique.

[0057] The speech information of the dialogue device 10 may include a speech signal generated by synthesizing the system utterance. The speech information of the dialogue device 10 may include text data used for synthesizing the speech signal. The image information of the dialogue device 10 may include a still image or moving image including at least part of the face or body of the CG character. The image information of the dialogue device 10 may include information used for generating an image of the CG character. As an example, the image information of the dialogue device 10 may include a facial expression label, posture information, or the like. When the dialogue device 10 is a robot, the control information of the dialogue device 10 may include body control information including a facial expression, gesture, and the like of the robot.

[0058] The video data of the overall dialogue may be moving image data generated by synthesizing the speech and video signals of the dialogue participant S and the speech and video signals of the dialogue device 10. The video data of the overall dialogue may be synthesized by the dialogue device 10 or synthesized by the control device 20.

[0059] The log information of the dialogue system 1000 may be history information recording a result of processing performed by the dialogue system 1000 in the dialogue between the dialogue participant S and the dialogue device 10. The log information of the dialogue system 1000 may include at least one of a processing result of the dialogue device 10, a processing result of the control device 20, or a processing result of the generation device 30. As an example, the log information of the dialogue system 1000 may include a recognition result of the user utterance, input information input into the generative model M2, output information output from the generative model

M2, an utterance content of the system utterance, and the like.

[0060] Each piece of information included in the dialogue data D may be stored in association with time information. As a method for associating each piece of information with the time information, for example, each piece of information and the time information may be stored as one set. Alternatively, information for acquiring, from one piece of information, the other piece of information may be stored. Alternatively, identification information of one piece of information and the other piece of information may be stored as one set.

[0061] (Model Storage Unit)

The model storage unit 103 stores the evaluation model M1. The evaluation model M1 may be stored in advance in the model storage unit 103. The evaluation model M1 may be generated by the control device 20 through training of training data provided in advance. The evaluation model M1 may be generated by another information processing device or information processing system different from the control device 20, and stored in the model storage unit 103.

[0062] The model storage unit 103 may store a plurality of evaluation models M1. The plurality of evaluation models M1 may include evaluation models constructed for each evaluation value. The evaluation value may include an index for evaluating at least one of a facial expression, body movement (e.g., head, hands, arms, or posture), a voice, smoothness of a dialogue, or dialogue content, with respect to the dialogue participant S.

[0063] The evaluation model M1 may include an evaluation model configured to evaluate the facial expression of the dialogue participant S. The evaluation model configured to evaluate the facial expression may include a machine learning model (e.g., a neural network) configured to receive, as an input, image information acquired by capturing an image of the face of the dialogue participant S, and output an evaluation value with respect to the facial expression of the dialogue participant S. The evaluation model M1 configured to evaluate the facial expression may include a machine learning model configured to receive, as an input, a feature extracted from image information, such as facial feature points, facial landmarks, FAU (Facial Action Units), and the like, and output an evaluation value with respect to the facial expression of the dialogue participant S (e.g., a decision tree, a random forest, a support

vector machine, or a neural network).

[0064] The evaluation model M1 may include an evaluation model configured to evaluate the body movement of the dialogue participant S. The evaluation model configured to evaluate the body movement may include a machine learning model configured to receive, as an input, a video signal acquired by capturing an image of the dialogue participant S, and output an evaluation value relating to the body movement of the dialogue participant S (e.g., a foundation model, a posture estimation model, an action recognition model, a gesture recognition model, and a model configured to output an evaluation value from features acquired by these models).

[0065] The evaluation model M1 may include an evaluation model configured to evaluate the voice of the dialogue participant S. The evaluation model configured to evaluate the voice may include a machine learning model (e.g., a neural network) configured to receive, as an input, a speech signal acquired by collecting an utterance of the dialogue participant S, and output an evaluation value relating to non-verbal information of the speech of the dialogue participant S. The non-verbal information may include, for example, tone, inflection, speaking speed, pitch, clarity, speech duration, and the number of fillers.

[0066] The evaluation model M1 may include an evaluation model configured to evaluate smoothness of a dialogue. The evaluation model configured to evaluate the smoothness of the dialogue may include a machine learning model configured to receive, as an input, a speech signal of the overall dialogue, and output an evaluation value relating to the smoothness of the dialogue (e.g., a foundation model, a speech recognition model, a speech interval detection model, a speaking speed estimation model, a filler detection model, or a model that has undergone training of evaluation values using features acquired from these). The evaluation value relating to the smoothness of the dialogue may include, for example, the length of an exchange in the dialogue, the timing of an utterance of the dialogue participant S, the rate at which the dialogue participant S is speaking, the silent time, and the like. At least part of the evaluation value relating to the smoothness of the dialogue may be analyzed on a rule basis.

[0067] The evaluation model M1 may include an evaluation model configured to evaluate the dialogue content. The evaluation model M1 configured to

evaluate the dialogue content may include a machine learning model configured to receive, as an input, at least part of text data indicating utterance content, and output an evaluation value relating to the dialogue content (e.g., a large language model given a system prompt instructing evaluation of dialogue content).

[0068] The evaluation model M1 may be constructed based on the evaluation criteria of the evaluator E. The evaluation model M1 may be constructed for each evaluation criterion of each of the plurality of evaluators E. The evaluation model M1 may be constructed based on video data of a dialogue to which a Ground Truth value of the evaluation value is assigned based on the evaluation criteria of the evaluator E.

[0069] (Evaluation Result Storage Unit)

The evaluation result storage unit 104 stores an evaluation result relating to the dialogue participant S. The dialogue evaluation unit 140 may generate the evaluation result relating to the dialogue participant S. The evaluation result storage unit 104 may store a plurality of evaluation results relating to the dialogue participant S. The plurality of evaluation results may be generated based on each piece of the dialogue data D relating to a plurality of different dialogues. The plurality of dialogues may differ in at least one of time, scenario, evaluation criteria, evaluation value, or the like. The evaluation result storage unit 104 may store a transition or statistical information of the evaluation values indicated in the plurality of evaluation results. The statistical information may be, for example, a maximum, a minimum, an average, or the like.

[0070] The evaluation result storage unit 104 may store an external evaluation value. The external evaluation value is an evaluation value acquired by evaluating the dialogue participant S based on information other than the dialogue data D. The external evaluation value may be generated by an information processing device or an information processing system external to the dialogue system 1000, and stored in the evaluation result storage unit 104. As an example, the external evaluation value may include the dialogue participant S's result of an aptitude test, information indicating language ability, information indicating work experiences, information indicating expertise in a specific field, or the like.

[0071] (Utterance Acquisition Unit)

The utterance acquisition unit 110 is configured to acquire the user utterance information. The utterance acquisition unit 110 may generate the user utterance information based on a speech signal acquired by collecting the user utterance. The utterance acquisition unit 110 may acquire, from the dialogue device 10, a speech signal acquired by collecting the user utterance.

[0072] The utterance acquisition unit 110 may generate the user utterance information based on a video signal acquired by capturing an image of the dialogue participant S. The utterance acquisition unit 110 may acquire, from the dialogue device 10, a video signal acquired by capturing an image of the dialogue participant S.

[0073] The utterance acquisition unit 110 may generate the user utterance information based on text data indicating the utterance content of the user utterance. The utterance acquisition unit 110 may generate text data indicating the utterance content of the user utterance by performing speech recognition of a speech signal acquired by collecting the user utterance.

[0074] The utterance acquisition unit 110 may determine whether the video signal acquired from the dialogue device 10 is a video signal acquired by appropriately capturing an image of the dialogue participant S. The utterance acquisition unit 110 may determine whether the face of the dialogue participant S is captured in the video signal. The utterance acquisition unit 110 may determine whether the size of the face of the dialogue participant S captured in the video signal is within a predetermined range. The utterance acquisition unit 110 may detect, using a face detector, the face of the dialogue participant S captured in the video signal. The face detector may be constructed, for example, using a machine learning model.

[0075] When computing an evaluation value relating to the body movement of the dialogue participant S, the utterance acquisition unit 110 may determine whether the body of the dialogue participant S is captured in the video signal. The utterance acquisition unit 110 may determine whether the size of the body of the dialogue participant S captured in the video signal is within a predetermined range. The utterance acquisition unit 110 may detect, using a body detector, the body of the dialogue participant S captured in the video signal. The body detector may be constructed, for example, using a machine learning model.

[0076] When the dialogue participant S is not appropriately captured in the

video signal, the utterance acquisition unit 110 may notify the dialogue participant S to that effect. As an example, the utterance acquisition unit 110 may transmit, to the dialogue device 10, a notification indicating that the dialogue participant S is not appropriately captured. The dialogue device 10 may display, on a display device, a message indicating that the dialogue participant S is not appropriately captured. The dialogue device 10 may utter that the dialogue participant S is not appropriately captured. When the dialogue participant S is not appropriately captured in the video signal, the utterance acquisition unit 110 may stop a dialogue with the dialogue participant S.

[0077] The utterance acquisition unit 110 may store the acquired user utterance information in the dialogue storage unit 102. The utterance acquisition unit 110 may store the user utterance information in real time or may store the user utterance information collectively in a predetermined processing unit. The predetermined processing unit may be, for example, the data amount of the user utterance information, the number of turns in the dialogue, the elapsed time of the dialogue, or the like.

[0078] The utterance acquisition unit 110 may include the user utterance information in the dialogue data D stored in the dialogue storage unit 102. The utterance acquisition unit 110 may store the user utterance information in the dialogue storage unit 102 in association with the time information included in the dialogue data D.

[0079] (Utterance Generation Unit)

The utterance generation unit 120 is configured to generate the system utterance information. The utterance generation unit 120 may generate the system utterance information based on the user utterance information acquired by the utterance acquisition unit 110. The utterance generation unit 120 may generate the system utterance information based on the generative model M2. The utterance generation unit 120 may generate the system utterance information based on the user utterance information, the input information, and the generative model M2.

[0080] Generating another piece of information based on a piece of information and a machine learning model may include executing one or more of the following processes:

- Inputting the piece of information into the machine learning model, thereby

generating the other piece of information;
- Inputting information generated based on the piece of information into the machine learning model, thereby generating the other piece of information; and
- Executing a generation process using the machine learning model based on information extracted from the piece of information, thereby generating the other piece of information.

[0081] The input information may be generated based on the scenario information. The scenario information may be set in advance. The scenario information may be set based on predetermined criteria. The scenario information may be set based on instructions from the evaluator E.

[0082] The utterance generation unit 120 may generate input information input into the generative model M2 based on the user utterance information, and transmit the generated input information to the generation device 30. Hereinafter, the input information input into the generative model M2 will be also referred to as “generative model input information”.

[0083] The generative model input information may include at least a system prompt and dialogue history information as examples of the input information. The utterance generation unit 120 may read the system prompt from the scenario information stored in the scenario storage unit 101. The utterance generation unit 120 may select a system prompt to be read according to the degree of progress of the dialogue. The utterance generation unit 120 may select the system prompt to be read according to a business type or a job type. The utterance generation unit 120 may read a plurality of system prompts and combine the plurality of system prompts with each other. As an example, the utterance generation unit 120 may read the system prompt according to a business type and the system prompt according to a job type, thereby generating a system prompt in which they are combined with each other.

[0084] The utterance generation unit 120 may select the system prompt based on a machine learning model. The machine learning model may be the generative model M2 or another machine learning model. The other machine learning model may be, for example, a large language model or a foundation model. The utterance generation unit 120 may input, into the machine learning model, information relating to a dialogue between the dialogue participant S and the dialogue device 10 and a plurality of system prompts, and acquire information for selecting one or more system prompts. The information

relating to the dialogue may be dialogue history information. The utterance generation unit 120 may also generate the system prompt based on a machine learning model.

[0085] The dialogue history information may be information acquired by connecting, in time series, the utterance content of the dialogue participant S and the utterance content of the dialogue device 10. The dialogue history information may include utterance contents from the start of a dialogue to the immediately preceding utterance. The dialogue history information may include a predetermined number of utterance contents obtained by going back in time from the immediately preceding utterance. Information indicating an utterance source (the dialogue participant S or the dialogue device 10) may be associated with each utterance content. The information indicating the utterance source may be added before or after the utterance content. The information indicating the utterance source may be added to be identifiable using a predetermined symbol. The boundary between the utterance contents may be indicated by a predetermined symbol. The dialogue history information may include information indicating a target utterance source whose utterance is to be generated. The utterance content need not be associated with the information indicating the utterance source whose utterance is to be generated.

[0086] As an example, the dialogue history information may be a character string shown below:

-“Store clerk: Welcome ¥n customer:”.

[0087] Here, “Welcome” is an example of the utterance content of the immediately preceding utterance. “Store clerk” and “customer” are examples of the information indicating an utterance source. “:” is an example of a symbol separating the utterance source and its utterance content. “¥n” is an example of a symbol indicating the boundary between the utterance contents (in this case, the end of the immediately preceding utterance content). In the above example, since “customer” is not associated with any utterance content, an utterance of the customer (i.e., the dialogue device 10) is a generation target. In the above example, the dialogue history information includes only the immediately preceding utterance content. However, if one or more turns are performed beforehand, utterance contents of the previous turns may be included before “store clerk”.

[0088] The dialogue history information is not limited to text information,

and may include speech information or image information or may include text information and at least one of speech information or image information. Specifically, the dialogue history information may include information acquired by connecting, in time series, speech information of the dialogue participant S, image information of the dialogue participant S, speech information of the dialogue device 10, and image information of the dialogue device 10.

[0089] The utterance generation unit 120 may generate generative model input information by embedding, in a predetermined template, the system prompt and the dialogue history information. The utterance generation unit 120 may generate the generative model input information by processing the template, the system prompt, and the dialogue history information based on a predetermined rule. The utterance generation unit 120 may generate the generative model input information by inputting, into a machine learning model, the template, the system prompt, and the dialogue history information. The machine learning model may be the generative model M2 or another machine learning model. The other machine learning model may be, for example, a large language model or a foundation model.

[0090] The template for generating the generative model input information may include one or more placeholders into which the system prompt or the dialogue history information is inserted. The template may include instruction information that instructs generation of an utterance. The instruction information may include information that instructs an output format of output information. The output format of the output information may include, for example, at least one of text data, a speech signal, a video signal, or a control signal. The template may include constraint information relating to output information. The constraint information may include a length of an utterance, content that may be spoken, content that may be spoken if asked, content that may not be spoken even if asked, a manner of speaking, and the like. The instruction information and the constraint information may be fixed sentences defined in advance.

[0091] The template may be optimized to obtain good output information by any prompt tuning method. As an example, the template may be optimized by searching for an optimal template by a genetic algorithm or the like based on a benchmark evaluating a prompt generated using the template.

[0092] The generative model input information may include text data, image

data, or acoustic data. As an example, the text data may be a natural language sentence referred to as a prompt. As an example, the image data may be a still image or a moving image. The text data may be text data acquired by performing speech recognition on audio data or audio included in moving-image data. The text data may be text acquired by performing character recognition on image data. As an example, the image data may include an image of a user. As an example, the acoustic data may include speech uttered by a user. The acoustic data may be speech data acquired by synthesizing speech from text data.

[0093] The utterance generation unit 120 may receive, from the generation device 30, output information generated by the generation device 30 inputting the generative model input information into the generative model M2. As the generative model input information, the system prompt and the dialogue history information may be separately input into the generative model M2. Alternatively, as the generative model input information, the system prompt and the dialogue history information combined with each other may be input into the generative model M2.

[0094] The utterance generation unit 120 may generate the system utterance information based on output information of the generative model M2. When the output information of the generative model M2 includes a speech signal indicating the system utterance or a video signal indicating a CG character, the utterance generation unit 120 may use at least one of the speech signal or the video signal as the system utterance information. When the output information of the generative model M2 includes text data indicating the utterance content, the utterance generation unit 120 may generate a speech signal by synthesizing speech from the text data. When the output information of the generative model M2 includes information for controlling the expression of the dialogue device 10 (e.g., a facial expression label, information relating to the shape of the mouth at the time of speaking, or posture information), the utterance generation unit 120 may synthesize a video signal or the like indicating a CG character based on the above information.

[0095] The utterance generation unit 120 may generate the system utterance information based on part of the output information of the generative model M2. When the output information of the generative model M2 is equal to or greater than a predetermined data length or a randomly determined data

length, the utterance generation unit 120 may reduce the output information of the generative model M2 such that the output information is equal to or less than the data length. When the generative model M2 outputs the output information in a stream, the generative model M2 may stop subsequent output when the specified data length has been reached.

[0096] As an example, the utterance generation unit 120 may acquire N sentences sequentially from the beginning of the output information of the generative model M2. Note that N is a positive integer and may be randomly determined. As an example, the utterance generation unit 120 may acquire N' sentences sequentially from the beginning of the output information of the generative model M2 such that the maximum number of characters is M or less. Note that M and N' are positive integers. The boundary between the sentences may be specified by a predetermined symbol. As an example, a punctuation mark (e.g., "。" or ".") and a symbol (e.g., "?" or "!") may be used as a delimiter.

[0097] The utterance generation unit 120 may store the generated system utterance information in the dialogue storage unit 102. The utterance generation unit 120 may store the system utterance information in real time or may store the system utterance information collectively in predetermined processing units. The predetermined processing units may be, for example, the data amount of the system utterance information, the number of turns in the dialogue, the elapsed time of the dialogue, or the like.

[0098] The utterance generation unit 120 may include the system utterance information in the dialogue data D stored in the dialogue storage unit 102. The utterance acquisition unit 110 may store the system utterance information in the dialogue storage unit 102 in association with the time information included in the dialogue data D.

[0099] (Utterance Presentation Unit)

The utterance presentation unit 130 controls the utterance of the dialogue device 10. The utterance presentation unit 130 may transmit, to the dialogue device 10, a speech signal generated by synthesizing the system utterance. The dialogue device 10 may output, from a speaker, a speech signal indicating the system utterance based on the speech signal received from the control device 20.

[0100] The utterance presentation unit 130 may transmit, to the dialogue device 10, text data indicating an utterance content. The dialogue device 10

may generate a speech signal by synthesizing the text data received from the control device 20, and output the generated speech signal from a speaker.

[0101] The utterance presentation unit 130 may transmit, to the dialogue device 10, a video signal indicating a CG character. The dialogue device 10 may display, on a display device, a video indicating a CG character based on the video signal received from the control device 20.

[0102] The utterance presentation unit 130 may transmit, to the dialogue device 10, information used for synthesizing the video signal (e.g., a facial expression label, information relating to the shape of the mouth at the time of speaking, or posture information). The dialogue device 10 may synthesize a video signal based on the information received from the control device 20, and display, on a display device, a video indicating the CG character.

[0103] The utterance presentation unit 130 may transmit, to the dialogue device 10, a control signal for controlling the expression of the CG character. The dialogue device 10 may control the expression of the CG character based on the control signal received from the control device 20.

[0104] The utterance presentation unit 130 may control the dialogue device 10 to utter the system utterance for explaining the situation of a dialogue (hereinafter also referred to as an “explanation utterance”). The utterance presentation unit 130 may cause the dialogue device 10 to utter the explanation utterance based on the guidance information stored in the scenario storage unit 101. The utterance presentation unit 130 may transmit, to the dialogue device 10, the guidance information stored in the scenario storage unit 101. The utterance presentation unit 130 may transmit, to the dialogue device 10, a speech signal generated by synthesizing the guidance information.

[0105] The utterance presentation unit 130 may control the dialogue device 10 to utter the system utterance in response to the user utterance. The utterance presentation unit 130 may cause the dialogue device 10 to utter the system utterance in response to the user utterance based on the system utterance information generated by the utterance generation unit 120. The utterance presentation unit 130 may transmit, to the dialogue device 10, at least one of the text data, speech signal, video signal, or control signal included in the system utterance information.

[0106] (Dialogue Evaluation Unit)

The dialogue evaluation unit 140 generates an evaluation result relating to the

dialogue participant S. The dialogue evaluation unit 140 may generate an evaluation result relating to the dialogue participant S based on the dialogue data D read from the dialogue storage unit 102. The dialogue evaluation unit 140 may generate an evaluation result relating to the dialogue participant S based on the evaluation model M1 read from the model storage unit 103. The dialogue evaluation unit 140 may generate input information for the evaluation model M1 based on the dialogue data D and input the generated input information into the evaluation model M1. The input information for the evaluation model M1 may include a feature extracted from the dialogue data D. Hereinafter, the input information for the evaluation model M1 is also referred to as “evaluation model input information”.

[0107] The evaluation result relating to the dialogue participant S may include one or more evaluation values. The dialogue evaluation unit 140 may select an evaluation value to be included in the evaluation result according to the evaluation criteria of the evaluator E. The dialogue evaluation unit 140 may generate an evaluation result relating to the dialogue participant S based on the evaluation model M1 according to the evaluation criteria of the evaluator E. The dialogue evaluation unit 140 may extract, from the dialogue data D, a feature to be included in the evaluation model input information, according to the evaluation value to be included in the evaluation result.

[0108] When a plurality of evaluation models M1 are stored in the model storage unit 103, the dialogue evaluation unit 140 may generate the evaluation results relating to the dialogue participant S according to the plurality of evaluation models M1. The evaluation result relating to the dialogue participant S may include a plurality of evaluation results based on the plurality of evaluation models M1. The evaluation result relating to the dialogue participant S may include one evaluation result including a plurality of evaluation values output by the plurality of evaluation models M1.

[0109] The evaluation result relating to the dialogue participant S may include a reliability indicating the certainty of the evaluation value. The reliability of the evaluation value may be output together with the evaluation value by the evaluation model M1. For example, when the evaluation model M1 is a classification model, the evaluation model M1 may calculate a probability for each class to be classified, set the class having the highest probability as the evaluation value, and output the probability for that class as the reliability of

the evaluation value.

[0110] As a non-limiting example, the dialogue evaluation unit 140 may generate the evaluation result relating to the dialogue participant S by inputting the evaluation model input information into the generative model M2. The evaluation model M1 may be generated based on the generative model M2. As an example, the evaluation model M1 may be generated by performing transfer learning on the generative model M2. As an example, the evaluation model M1 may be generated by fine-tuning the generative model M2.

[0111] The dialogue evaluation unit 140 may generate, at any timing, an evaluation result relating to the dialogue participant S. The dialogue evaluation unit 140 may generate the evaluation result by batch processing or ad hoc processing. As an example, the dialogue evaluation unit 140 may generate, at predetermined time intervals, an evaluation result relating to each piece of the accumulated dialogue data D. The dialogue evaluation unit 140 may generate, at the end of the dialogue, an evaluation result relating to the dialogue data D relating to the dialogue.

[0112] The dialogue evaluation unit 140 may store, in the evaluation result storage unit 104, an evaluation result relating to the dialogue participant S. The dialogue evaluation unit 140 may store, in the evaluation result storage unit 104, the evaluation result in association with identification information by which the dialogue is identifiable. As an example, the dialogue evaluation unit 140 may store, in the evaluation result storage unit 104, the time information and the evaluation result in association.

[0113] (Result Output Unit)

The result output unit 150 outputs the evaluation result relating to the dialogue participant S. The result output unit 150 may read the evaluation result relating to the dialogue participant S stored in the evaluation result storage unit 104.

[0114] The result output unit 150 may output the evaluation result relating to the dialogue participant S including the external evaluation value relating to the dialogue participant S. The result output unit 150 may read, from the evaluation result storage unit 104, the external evaluation value relating to the dialogue participant S along with the evaluation result relating to the dialogue participant S, and include the external evaluation value in the evaluation result relating to the dialogue participant S.

[0115] The result output unit 150 may include the dialogue data D used for the evaluation of the dialogue participant S in the evaluation result relating to the dialogue participant S. As an example, the result output unit 150 may include, in the evaluation result relating to the dialogue participant S, video data included in the dialogue data D or a feature extracted from the video data. The feature extracted from the video data may include the feature input into the evaluation model M1.

[0116] The result output unit 150 may transmit, to the terminal device 40, the evaluation result relating to the dialogue participant S. The result output unit 150 may transmit, to the terminal device 40, the evaluation result relating to the dialogue participant S in response to an output request from the terminal device 40. When there are a plurality of evaluation results relating to the dialogue participant S, the result output unit 150 may transmit, to the terminal device 40, a transition or statistical information of the evaluation value indicated in the evaluation result. The terminal device 40 may present, to the evaluator E, the evaluation result relating to the dialogue participant S.

[0117] The result output unit 150 may transmit, to the dialogue device 10, the evaluation result relating to the dialogue participant S. The dialogue device 10 may present, to the dialogue participant S, the evaluation result relating to the dialogue participant S. The result output unit 150 may transmit the evaluation result relating to the dialogue participant S to an external information processing device or information processing system. The external information processing device or information processing system may present the evaluation result relating to the dialogue participant S to a user other than the dialogue participant S. The user other than the dialogue participant S may include the evaluator E.

[0118] As an example, the result output unit 150 may transmit, to the terminal device 40, screen data for displaying the evaluation result. As an example, the screen data may be electronic data written in HTML (HyperText Markup Language) or the like. The result output unit 150 may transmit, to the dialogue device 10, screen data for displaying the evaluation result.

[0119] As an example, the result output unit 150 may transmit the evaluation result to an account of the evaluator E or dialogue participant S by e-mail, messaging service, or the like. The result output unit 150 may transmit link information to a screen for displaying the evaluation result to the account of

the evaluator E or dialogue participant S by e-mail, messaging service, or the like.

[0120] The functional configuration of the control device 20 illustrated in FIG. 2 is merely an example, and there can be various functional configuration examples according to applications and purposes. The categorization of storage units, such as the scenario storage unit 101, the dialogue storage unit 102, the model storage unit 103, and the evaluation result storage unit 104 illustrated in FIG. 2, is merely an example. The categorization of processing units, such as the utterance acquisition unit 110, the utterance generation unit 120, the utterance presentation unit 130, the dialogue evaluation unit 140, and the result output unit 150 illustrated in FIG. 2, is merely an example.

[0121] For example, at least two of the scenario storage unit 101, the dialogue storage unit 102, the model storage unit 103, or the evaluation result storage unit 104 may be integrated into a single storage unit. Furthermore, for example, at least one of the dialogue storage unit 102, the model storage unit 103, or the evaluation result storage unit 104 may be divided into a plurality of storage units.

[0122] For example, at least two of the utterance acquisition unit 110, the utterance generation unit 120, the utterance presentation unit 130, the dialogue evaluation unit 140, or the result output unit 150 may be integrated into a single processing unit. Furthermore, for example, at least one of the utterance acquisition unit 110, the utterance generation unit 120, the utterance presentation unit 130, the dialogue evaluation unit 140, or the result output unit 150 may be divided into a plurality of processing units.

[0123] <Flow of Dialogue Method>

A dialogue method executed by the dialogue system 1000 will be described with reference to FIG. 3. FIG. 3 is a process flow illustrating an example of a dialogue method according to the first embodiment. Hereinafter, an example in which a CG character is used to have a dialogue with the dialogue participant S operating a personal computer, which is an example of the dialogue device 10, will be described.

[0124] In step S101, the dialogue participant S performs an operation to start a dialogue with the dialogue device 10. As an example, the dialogue participant S may perform an operation to select the control device 20 for dialogue from a list of the control devices 20 displayed on the display device of the dialogue

device 10. When the dialogue participant S performs an operation to select the control device 20, the dialogue device 10 may execute a process to connect to the control device 20.

[0125] The dialogue participant S may perform a test of a microphone or camera of the dialogue device 10. The dialogue participant S may turn on a switch of the microphone of the dialogue device 10 and check the volume level of the microphone. Here, the dialogue device 10 and the control device 20 may cooperate to instruct the dialogue participant S to read aloud a specific sentence, and the dialogue participant S may read aloud the specified sentence, followed by performing speech recognition of the sentence read aloud, and determining that the microphone is operating properly if the recognized text is the same as or similar to the instructed sentence. The dialogue participant S may turn on a switch of the camera of the dialogue device 10 and check a camera image displayed on the display device. Here, the dialogue device 10 may use a face or body detector to determine whether an image of the dialogue participant S is properly captured in the camera image. The dialogue device 10 may perform control not to start a dialogue until an image of the dialogue participant S is properly captured in the camera image. The image of the dialogue participant S being properly captured in the camera image may include, for example, that the face appears at a size larger than a predetermined size, the face is directed to the front, and the body appears within a predetermined range.

[0126] In step S102, the utterance presentation unit 130 of the control device 20 reads the guidance information stored in the scenario storage unit 101. The utterance presentation unit 130 synthesizes speech from the guidance information. Accordingly, a speech signal indicating the explanation utterance is generated. The utterance presentation unit 130 transmits, to the dialogue device 10, the speech signal indicating the explanation utterance. The utterance presentation unit 130 may transmit the guidance information to the dialogue device 10 alone or along with the speech signal.

[0127] In step S103, the dialogue device 10 receives, from the control device 20, a speech signal indicating the explanation utterance. The dialogue device 10 outputs the explanation utterance from a speaker based on the speech signal indicating the explanation utterance. When receiving the guidance information from the control device 20, the dialogue device 10 may display the guidance

information on the display device. When receiving the guidance information from the control device 20, the dialogue device 10 may output, from a speaker, a speech signal generated by synthesizing the guidance information.

[0128] In step S104, the control device 20 starts recording of a dialogue between the dialogue participant S and the dialogue device 10. Specifically, the control device 20 requests the utterance acquisition unit 110 to store the user utterance information in the dialogue storage unit 102. The control device 20 also requests the utterance generation unit 120 to store the system utterance information in the dialogue storage unit 102. When acquiring the user utterance information, the utterance acquisition unit 110 enables the function of storing the user utterance information in the dialogue storage unit 102. When generating the system utterance information, the utterance generation unit 120 enables the function of storing the system utterance information in the dialogue storage unit 102.

[0129] In step S105, the dialogue device 10 displays a CG character on the display device. The dialogue device 10 transmits, to the control device 20, a speech signal collected using a microphone and a video signal captured using a camera. Subsequently, the dialogue device 10 continues to transmit the speech signal and the video signal until the end of the dialogue.

[0130] The dialogue participant S utters in response to the explanation speech output in step S103. The dialogue device 10 receives the user utterance made by the dialogue participant S. The speech signal transmitted by the dialogue device 10 to the control device 20 includes speech corresponding to the user utterance uttered by the dialogue participant S. Furthermore, the video signal transmitted by the dialogue device 10 to the control device 20 includes an image of the facial expression or movement when the dialogue participant S uttered the user utterance.

[0131] In step S106, the utterance acquisition unit 110 of the control device 20 receives a speech signal acquired by collecting the user utterance from the dialogue device 10. Furthermore, the utterance acquisition unit 110 receives, from the dialogue device 10, a video signal acquired by capturing an image of the dialogue participant S.

[0132] The utterance acquisition unit 110 performs speech recognition on the speech signal acquired by collecting the user utterance. Accordingly, text data indicating the utterance content of the user utterance is generated. The

utterance acquisition unit 110 transmits, to the utterance generation unit 120, the text data indicating the utterance content of the user utterance.

[0133] The utterance acquisition unit 110 generates the user utterance information based on the speech signal acquired by collecting the user utterance, the video signal acquired by capturing an image of the dialogue participant S, and the text data indicating the utterance content of the user utterance. The utterance acquisition unit 110 stores the user utterance information in the dialogue storage unit 102.

[0134] In step S107, the utterance generation unit 120 of the control device 20 receives, from the utterance acquisition unit 110, the text data indicating the utterance content of the user utterance. The utterance generation unit 120 reads the system prompt stored in the scenario storage unit 101. The utterance generation unit 120 reads the dialogue data D from the dialogue storage unit 102. The utterance generation unit 120 generates dialogue history information based on the text data indicating the utterance content of the user utterance and the dialogue data D read from the dialogue storage unit 102.

[0135] The utterance generation unit 120 generates the generative model input information based on the system prompt and the dialogue history information. The utterance generation unit 120 may generate the generative model input information by embedding, in a predetermined template, the system prompt and the dialogue history information. The utterance generation unit 120 transmits the generative model input information to the generation device 30.

[0136] The generation device 30 receives the generative model input information from the control device 20. The generation device 30 inputs the received generative model input information into the generative model M2. The generative model M2 executes a predetermined task based on the generative model input information, and outputs the output information generated as a result. The output information from the generative model M2 may include at least the text data indicating the utterance content of the system utterance. The generation device 30 transmits, to the control device 20, the output information output from the generative model M2.

[0137] In step S108, the utterance generation unit 120 of the control device 20 receives the output information from the generation device 30. The utterance generation unit 120 generates the system utterance information

based on at least part of the received output information. The utterance generation unit 120 may synthesize speech from the text data indicating the utterance content of the system utterance. Accordingly, a speech signal indicating the system utterance is generated. The speech signal indicating the system utterance may be included in the output information from the generative model M2.

[0138] The utterance generation unit 120 may generate a video signal indicating the CG character. The utterance generation unit 120 may synthesize the video signal indicating the CG character based on the information included in the output information from the generative model M2. The video signal indicating the CG character may be included in the output information from the generative model M2. In other words, the video signal indicating the CG character may be directly generated by the generative model M2.

[0139] The utterance generation unit 120 generates the system utterance information based on the speech signal indicating the system utterance, the video signal indicating the CG character, and the text data indicating the utterance content of the system utterance. The utterance acquisition unit 110 stores the system utterance information in the dialogue storage unit 102. The utterance generation unit 120 transmits the system utterance information to the dialogue device 10.

[0140] In step S109, the dialogue device 10 receives the system utterance information from the control device 20. The dialogue device 10 utters the system utterance based on the system utterance information. The dialogue device 10 may output, from a speaker, speech indicating the system utterance. The dialogue device 10 may display, on the display device, the video indicating the CG character. When the dialogue device 10 is a robot, the robot may be controlled based on the system utterance information. Note that the dialogue device 10 may turn off the switch of the microphone during the utterance of the system utterance. In other words, the dialogue device 10 may collect the utterance of the dialogue participant S by the microphone until the system utterance is uttered.

[0141] In step S110, the control device 20 determines whether an end condition for ending the dialogue is satisfied. As an example, the end condition may be that the number of turns in the dialogue or the elapsed time from the start of the dialogue exceeds a predetermined threshold. In FIG. 3, an end

determination step is performed only immediately after step S108. However, when the determination is made based on the elapsed time, the end determination step may be performed at any timing from immediately before step S106 to immediately after step S108. In this case, the end determination may not be made until a predetermined time elapses after each utterance of the dialogue participant S ends. In this case, the dialogue can be ended at a time close to the predetermined time while preventing the conversation from ending abruptly.

[0142] At any timing from immediately before step S106 to immediately after step S108, confirmation as to whether the face or body appears in the camera can be made using a face or body detector. When the face or body is not detected for a predetermined time, the dialogue is ended and the process is caused to proceed to step S111. Before the end of the dialogue, the dialogue device 10 may output, to the dialogue participant S, a message indicating that the face or body cannot be detected, by displaying the message by text or outputting the message by speech.

[0143] The control device 20 may determine whether to end the dialogue based on the content of the dialogue. As an example, the control device 20 may input the dialogue history information into the machine learning model, and determine whether to end the dialogue based on the output information from the machine learning model. The machine learning model configured to determine the end of the dialogue may be the generative model M2 or another large language model or foundation model.

[0144] When it is determined that the end condition is satisfied (YES), the control device 20 moves the process to step S111. When it is determined that the end condition is satisfied, the control device 20 may transmit, to the dialogue device 10, a speech signal indicating a predetermined utterance. The predetermined utterance may be an utterance indicating the end of the dialogue. The dialogue device 10 may receive, from the control device 20, the speech signal indicating the predetermined utterance, and output the predetermined utterance from a speaker.

[0145] Conversely, when it is determined that the end condition is not satisfied (NO), the control device 20 returns the process to step S106. When the process returns to step S106, the utterance acquisition unit 110 of the control device 20 acquires the user utterance information relating to the next user

utterance. Subsequently, the control device 20 executes the process from step S106 to step S110 again based on the new user utterance information. In this manner, the dialogue system 1000 repeatedly executes the reception of the user utterance and the presentation of the system utterance until it is determined that the end condition is satisfied in step S110. Note that the utterance generation unit 120 may change the system prompt used for generating the system utterance with every repetition.

[0146] In step S111, the dialogue evaluation unit 140 of the control device 20 reads, from the dialogue storage unit 102, the dialogue data D relating to the dialogue with the dialogue participant S. The dialogue evaluation unit 140 also reads, from the model storage unit 103, the evaluation model M1 corresponding to the evaluator E.

[0147] The dialogue evaluation unit 140 generates the evaluation model input information based on the read dialogue data D. The dialogue evaluation unit 140 inputs the evaluation model input information into the evaluation model M1. The evaluation model M1 computes one or more evaluation values based on the evaluation model input information. Accordingly, the evaluation result relating to the dialogue participant S is generated. The dialogue evaluation unit 140 stores, in the evaluation result storage unit 104, the evaluation result relating to the dialogue participant S.

[0148] In step S112, the evaluator E performs an operation to display the evaluation result to the terminal device 40. As an example, the evaluator E may perform an operation to open the evaluation result of the dialogue participant S on the evaluation screen displayed on the display device of the terminal device 40. The terminal device 40 transmits, to the control device 20, an output request of the evaluation result relating to the dialogue participant S.

[0149] The result output unit 150 of the control device 20 receives, from the terminal device 40, the output request of the evaluation result. The result output unit 150 reads, from the evaluation result storage unit 104, the evaluation result relating to the dialogue participant S. The result output unit 150 generates screen data for displaying the evaluation result relating to the dialogue participant S. The result output unit 150 transmits the screen data to the terminal device 40.

[0150] In step S113, the terminal device 40 receives the screen data from the control device 20. The terminal device 40 displays, on the display device, the

evaluation result relating to the dialogue participant S based on the screen data. The evaluator E may evaluate the dialogue participant S by referring to the evaluation result relating to the dialogue participant S displayed on the display device of the terminal device 40. As an example, the evaluator E may determine whether to hire the dialogue participant S by considering the evaluation result relating to the dialogue participant S. As an example, the evaluator E who is a training instructor may determine whether the training is good or bad by considering the evaluation result of the dialogue participant S. As an example, the evaluator E who is a grader of the examination may determine the score of the dialogue participant S and determine pass or fail of the examination, by considering the evaluation result of the dialogue participant S.

[0151] The terminal device 40 may display, on the display device, the video data included in the evaluation result relating to the dialogue participant S. The evaluator E may compare the video data included in the evaluation result with the evaluation value, and determine whether the evaluation result relating to the dialogue participant S is appropriate.

[0152] The terminal device 40 may display, on the display device, a screen for correcting the evaluation result relating to the dialogue participant S. The evaluator E may correct, in the terminal device 40, the evaluation result relating to the dialogue participant S. The terminal device 40 may transmit, to the control device 20, the evaluation result corrected by the evaluator E. The control device 20 may update the evaluation result relating to the dialogue participant S stored in the evaluation result storage unit 104 based on the evaluation result received from the terminal device 40. The control device 20 may additionally train the evaluation model M1 based on the updated evaluation result.

[0153] [Modified Example 1 of First Embodiment]

The dialogue system 1000 may store the evaluation model M1 and the system prompt in association. As an example, the evaluation model M1 may be stored in the model storage unit 103 in association with identification information for identifying the system prompt. The dialogue evaluation unit 140 may read, from the model storage unit 103, the evaluation model M1 associated with the system prompt, and compute one or more evaluation values based on the evaluation model M1.

[0154] The dialogue system 1000 may store the evaluation model M1 and scenario information in association. As an example, scenario information including identification information for identifying the evaluation model M1 may be stored in the scenario storage unit 101. The utterance generation unit 120 may read, from the scenario storage unit 101, the system prompt corresponding to the scenario, and include the read system prompt in the input information input into the generative model M2. The dialogue evaluation unit 140 may read, from the model storage unit 103, the evaluation model M1 corresponding to the scenario, and compute one or more evaluation values based on the evaluation model M1. The dialogue system 1000 can select the system prompt and the evaluation model M1 only by selecting the scenario.

[0155] [Modified Example 2 of First Embodiment]

The control device 20 may store information used for control of the dialogue device 10. The information used for control of the dialogue device 10 may include, for example, text data used to synthesize a speech signal indicating the system utterance, start and end times of the synthesized speech signal, a label used to control the facial expression or body movement (e.g., laughing, waving, or nodding) of the dialogue device 10, and start and end times of a moving image indicating movement based on the label. The time may be Japan Standard Time (JST), Coordinated Universal Time (UTC), or an elapsed time from the time at which recording of the dialogue was started.

[0156] The information used for control of the dialogue device 10 may be stored in the storage device of the dialogue device 10. The information used for control of the dialogue device 10 may be stored in the storage device of the generation device 30. The information used for control of the dialogue device 10 may be stored in another information processing device or storage device different from the dialogue device 10, the control device 20, or the generation device 30.

[0157] The information used for control of the dialogue device 10 may be used to compute the evaluation value along with the speech signal or the video signal. As an example, when the information used for synthesis of the speech signal is used, an utterance section of the dialogue device 10 can be specified from the recorded video data of the overall dialogue based on the start and end times of reproduction of the speech signal. Furthermore, when the information used to synthesize the speech signal is used, the utterance content of the

dialogue device 10 can be acquired as text without using speech recognition. As an example, when the information relating to the facial expression or body movement of the dialogue device 10 is used, information such as the content of the dialogue participant S's response to the facial expression or body movement, the facial expression exhibited by the dialogue participant S in response thereto, and the tone of voice in which the dialogue participant S responded thereto can be input into the evaluation model M1 as additional information.

[0158] [Modified Example 3 of First Embodiment]

The utterance generation unit 120 may generate the system utterance by using a technique referred to as retrieval augmented generation (RAG). The retrieval augmented generation is a technique of including reference information, acquired by searching for an external data source, in a prompt for acquiring a good generation result from a machine learning model.

[0159] The control device 20 may store, in advance, reference information including background information relating to the dialogue device 10. The utterance generation unit 120 may generate a query based on the user utterance information and search for reference information based on the query. The utterance generation unit 120 may generate the system prompt based on the searched reference information. The utterance generation unit 120 may include the generated system prompt in input information input into the generative model M2, and transmit, to the generation device 30, the input information input into the generative model M2.

[0160] An example in which a dialogue is executed in a scenario where the dialogue participant S is a doctor and the dialogue device 10 is a patient will be described. The reference information includes a history of the occurrence of a symptom, a current symptom, a medical history, a chronic disease, a medication history, vital signs, a facial expression, and text describing the state of an injury. The utterance generation unit 120 generates a query based on a question made by the dialogue participant S, and acquires, based on the generated query, reference information relating to the question made by the dialogue participant S. The utterance generation unit 120 generates the system prompt including the acquired reference information, and generates the system utterance responding to the question made by the dialogue participant S based on the generative model M2. Accordingly, the dialogue device 10 can

appropriately respond to the question made by the dialogue participant S.

[0161] Although the foregoing description assumes, for simplicity, that the reference information is text information, the reference information is not limited to text information and may include speech information, image information, or the like.

[0162] [Second Embodiment]

In the first embodiment, a configuration of generating the evaluation result relating to the dialogue participant S based on the evaluation model M1 constructed in advance has been described. In a second embodiment, a configuration of constructing the evaluation model M1 corresponding to evaluation criteria of the evaluator E and generating an evaluation result relating to the dialogue participant S based on the evaluation model M1 according to the evaluation criteria of the evaluator E will be described.

[0163] Hereinafter, the dialogue system 1000 according to the second embodiment will be described focusing on differences from the first embodiment.

[0164] <Overall Configuration of Dialogue System>

An overall configuration of the dialogue system according to the present embodiment will be described with reference to FIG. 4. FIG. 4 is a block diagram illustrating an example of the overall configuration of the dialogue system according to the second embodiment.

[0165] As illustrated in FIG. 4, the dialogue system 1000 includes the dialogue device 10, the control device 20, the generation device 30, the terminal device 40, and the management device 50. That is, the dialogue system 1000 according to the present embodiment differs from the dialogue system 1000 according to the first embodiment (see FIG. 1) in that the management device 50 is further included.

[0166] The management device 50 is an example of an information processing device, such as a personal computer, a workstation, a server, or the like, which is configured to manage the evaluation model M1. The management device 50 may generate the evaluation model M1. The management device 50 may generate training data to be used for generating the evaluation model M1. The management device 50 may select the evaluation model M1 to be used for evaluating the dialogue participant S. The management device 50 may transmit, to the control device 20, the evaluation model M1 to be used for evaluating the

dialogue participant S.

[0167] The management device 50 may store video data V of a dialogue. The video data V may be electronic data recording a dialogue between a first dialogue participant and a second dialogue participant. The video data V may include a speech signal acquired by collecting a speech made by the first dialogue participant or the second dialogue participant. The video data V may include a video signal acquired by capturing an image of the first dialogue participant or the second dialogue participant.

[0168] The first dialogue participant may be a real person or the dialogue device 10. The second dialogue participant may be a real person or the dialogue device 10. In other words, the video data V may be video data recording a dialogue between real persons, video data recording a dialogue between a real person and the dialogue device 10, or video data recording a dialogue between the dialogue devices 10.

[0169] The management device 50 may include an evaluation model group M3. The evaluation model group M3 may include a plurality of evaluation models M1. The evaluation model group M3 may include a plurality of evaluation models M1 having different evaluation criteria. The evaluation model M1 may be generated based on the video data V to which a Ground Truth label according to the evaluation criteria is assigned. The Ground Truth label may be a Ground Truth value of the evaluation value.

[0170] The evaluation criteria may be determined for each evaluator E. The evaluation criteria may be determined based on an attribute of the evaluator E. As an example, the evaluation criteria may be determined for each business type of the evaluator E. The evaluation criteria may be determined based on an attribute of the dialogue participant S. As an example, the evaluation criteria may be determined for each job type of the dialogue participant S. As an example, the evaluation criteria may be determined for each scenario. This is because, when there is a scenario where it is better to respond with a smile and a scenario where it is better to respond with a blank facial expression, the evaluation criteria are considered to be different for each scenario.

[0171] The evaluation according to the specific evaluation criteria may be performed as follows. First, one or more features are extracted based on the video data V of a dialogue. One or more features may be features independent of the evaluation criteria. Next, the evaluation according to the evaluation

criteria of the evaluator E is performed based on the extracted one or more features. Here, an evaluation value may be computed by linear transformation for the one or more features. A value acquired by adding the respective features after the linear transformation may be computed as the evaluation value. The evaluation value may be acquired based on one or more machine learning models.

[0172] Note that the overall configuration of the dialogue system 1000 illustrated in FIG. 4 is merely an example, and there can be various system configuration examples according to applications and purposes. As an example, one or more of the dialogue device 10, the control device 20, the generation device 30, the terminal device 40, and the management device 50 may be included in the dialogue system 1000. The management device 50 may be realized by a plurality of computers or may be realized as a cloud computing service. The management device 50 may be realized by a stand-alone computer integrated with the control device 20 or the generation device 30.

[0173] The dialogue system 1000 may include a plurality of control devices 20 provided for each evaluator E. The dialogue system 1000 may include the single control device 20 to be used by a plurality of evaluators E. The dialogue system 1000 may include a plurality of management devices 50 provided for each evaluator E. The dialogue system 1000 may include the single control device 20 to be used by a plurality of evaluators E.

[0174] <Functional Configuration of Management Device>

A functional configuration of the management device 50 will be described with reference to FIG. 5. FIG. 5 is a block diagram illustrating an example of the functional configuration of the management device according to the second embodiment.

[0175] As illustrated in FIG. 5, the management device 50 includes a data storage unit 501, a model storage unit 502, a label assignment unit 510, a feature extraction unit 520, a model training unit 530, a model evaluation unit 540, and a model selection unit 550. By execution of a management program installed in advance, the management device 50 functions as the data storage unit 501, the model storage unit 502, the label assignment unit 510, the feature extraction unit 520, the model training unit 530, the model evaluation unit 540, and the model selection unit 550.

[0176] (Data Storage Unit)

The data storage unit 501 stores the video data V. The video data V may be generated by the control device 20. The video data V may be recorded using the dialogue device 10. The video data V may be generated by another information processing device or information processing system.

[0177] The video data V may include video data included in the dialogue data D stored in the control device 20. As an example, the video data V may be video data extracted from the dialogue data D stored in the control device 20 during a predetermined period of data collection.

[0178] The video data V may include video data collected from unspecified users. As an example, the video data of a dialogue may be collected by a method, such as crowdsourcing or the like. The video data V may be recorded for any purpose. In other words, the video data V does not necessarily need to be recorded for constructing the evaluation model M1.

[0179] The video data V may be video data of a dialogue in which one dialogue participant is the dialogue participant S designated by the evaluator E. The evaluator E may designate the dialogue participant S having a fixed evaluation. The evaluator E may designate the dialogue participant S having a high evaluation. The evaluator E may designate the dialogue participant S having a low evaluation.

[0180] (Model Storage Unit)

The model storage unit 502 stores the evaluation model group M3. The evaluation model group M3 may include a plurality of evaluation models M1 having different evaluation criteria. The evaluation models M1 included in the evaluation model group M3 may be stored in advance in the model storage unit 502. The evaluation model M1 included in the evaluation model group M3 may be acquired from the terminal device 40. The evaluation model M1 acquired from the terminal device 40 may be generated by the evaluator E. The evaluation model M1 included in the evaluation model group M3 may be generated by the model training unit 530 and stored in the model storage unit 502.

[0181] The evaluation model M1 included in the evaluation model group M3 is similar to that in the first embodiment. Accordingly, the evaluation model M1 may include an evaluation model for evaluating the facial expression of the dialogue participant S. The evaluation model M1 may include an evaluation model for evaluating the body movement of the dialogue participant S. The

evaluation model M1 may include an evaluation model for evaluating the voice of the dialogue participant S. The evaluation model M1 may include an evaluation model for evaluating smoothness of a dialogue. The evaluation model M1 may include an evaluation model for evaluating the content of a dialogue.

[0182] (Label Assignment Unit)

The label assignment unit 510 assigns a Ground Truth label to the video data V. The label assignment unit 510 may assign a Ground Truth label to the video data V stored in the data storage unit 501. The label assignment unit 510 may assign a Ground Truth label to the video data V according to the evaluation criteria. The label assignment unit 510 may assign, to the video data V, a Ground Truth label designated by the evaluator E.

[0183] The Ground Truth label is information indicating a Ground Truth value of the evaluation value included in the evaluation result. The Ground Truth label may include, for example, a numerical value (e.g., 1 to 10), a character (e.g., an alphabetical character), or a symbol (e.g., O, X, or Δ).

[0184] The Ground Truth label may include an evaluation value acquired by evaluating the speech included in the video data V. The Ground Truth label may include an evaluation value acquired by evaluating the video included in the video data V. The Ground Truth label may include the evaluation value acquired by evaluating the content of the dialogue included in the video data V.

[0185] The label assignment unit 510 may correct the Ground Truth label assigned to the video data relating to the dialogue participant S designated by the evaluator E. The label assignment unit 510 may correct the Ground Truth label assigned to the video data relating to the dialogue participant S rated highly by the evaluator E to a Ground Truth label indicating a high evaluation. The label assignment unit 510 may correct the Ground Truth label assigned to the video data relating to the dialogue participant S rated poorly by the evaluator E to a Ground Truth label indicating a low evaluation.

[0186] The dialogue participant S rated highly by the evaluator E is considered to be a dialogue participant capable of having a dialogue according to the evaluation criterion of the evaluator E. The dialogue participant S rated poorly by the evaluator is considered to be a dialogue participant that does not have a dialogue according to the evaluation criterion of the evaluator E. When a Ground Truth label according to the evaluation of the evaluator E is assigned to

the video data relating to the dialogue participant S designated by the evaluator E, the evaluation result according to the evaluation criterion of the evaluator E can be generated.

[0187] The label assignment unit 510 may assign a Ground Truth label based on a machine learning model. The machine learning model may be the trained evaluation model M1, the generative model M2, or another machine learning model. The other machine learning model may be, for example, a large language model or a foundation model.

[0188] The label assignment unit 510 may correct the Ground Truth label assigned based on the machine learning model. The label assignment unit 510 may present, to the evaluator E, the Ground Truth label assigned based on the machine learning model, and correct the assigned Ground Truth label to a Ground Truth label designated by the evaluator E.

[0189] (Feature Extraction Unit)

The feature extraction unit 520 extracts a feature vector from the video data V. The feature extraction unit 520 may extract a feature vector input into the evaluation model M1 included in the evaluation model group M3. The feature extraction unit 520 may extract a feature vector including a feature designated by the evaluator E.

[0190] The feature included in the feature vector is similar to that in the first embodiment. Accordingly, the feature vector may include a feature extracted from the speech information of the dialogue participant S. The feature extracted from the speech information may include, for example, tone, inflection, speaking speed, pitch, clarity, speech duration, and the number of fillers.

[0191] The feature vector may include a feature extracted from image information of the dialogue participant S. The feature extracted from the image information may include, for example, the degree of a smile, the degree of a blank facial expression, the degree of a facial expression expressing an emotion, such as anger, sadness, or the like, the direction of a gaze, and the like.

[0192] The feature vector may include a feature extracted from the dialogue content. The feature extracted from the dialogue content may include, for example, the number of positive words, the number of negative words, the rate of polite language, and the like.

[0193] (Model Training Unit)

The model training unit 530 trains the evaluation model M1 included in the evaluation model group M3. The model training unit 530 may train the evaluation model M1 based on training data in which a Ground Truth label is assigned to the video data V.

[0194] Specifically, the model training unit 530 inputs, into the evaluation model M1, a feature vector extracted from the video data V by the feature extraction unit 520. The evaluation model M1 predicts one or more evaluation values based on the input feature vector, and outputs a predicted value of the evaluation value. The model training unit 530 computes difference information between the predicted value output from the evaluation model M1 and the Ground Truth label assigned to the video data V. The model training unit 530 updates the parameter of the evaluation model M1 based on the difference information between the predicted value and the Ground Truth label. The model training unit 530 repeats the prediction of the evaluation value and the update of the parameter until a predetermined convergence condition is satisfied. The convergence condition may be that the number of updates of the parameter is equal to or greater than a threshold, the update amount of the parameter is less than a threshold, or the numerical value computed from the difference information is less than a threshold.

[0195] The model training unit 530 may train the evaluation model M1 according to evaluation criteria. The model training unit 530 may train the evaluation model M1 configured to output the evaluation value according to each of a plurality of evaluation criteria. The model training unit 530 may train the evaluation model M1 according to evaluation criteria based on the training data to which Ground Truth labels according to the evaluation criteria are assigned.

[0196] The model training unit 530 may additionally train the pre-trained evaluation model M1. The additional training may include, for example, transfer learning or fine tuning. The model training unit 530 may additionally train the evaluation model M1 based on training data in which Ground Truth labels are assigned to the video data V. As an example, the model training unit 530 may additionally train the evaluation model M1 pre-trained using general-purpose evaluation criteria based on training data to which Ground Truth labels according to the evaluation criteria of the evaluator E are assigned.

[0197] The model training unit 530 may construct the evaluation model M1

by combining a plurality of features extracted from the video data. As an example, the model training unit 530 may construct the evaluation model M1 configured to convert a plurality of features into a single evaluation value by a weighted linear sum. The model training unit 530 may construct the evaluation model M1 based on a combination operation of the features performed by the evaluator E. In the case of a weighted linear sum, the evaluator E may construct the evaluation model M1 by manually operating the weights using the terminal device 40.

[0198] (Model Evaluation Unit)

The model evaluation unit 540 evaluates the evaluation model M1 trained by the model training unit 530. The model evaluation unit 540 may acquire a predicted value of the evaluation value predicted by the evaluation model M1 using video data provided in advance, and present the predicted value to the evaluator E. The video data provided in advance may be video data not used for training the evaluation model M1.

[0199] The evaluator E may evaluate the performance of the evaluation model M1 based on the predicted value of the evaluation value presented by the model evaluation unit 540. As an example, the evaluator E may determine whether the predicted value of the evaluation value is an expected result. The evaluator E may input, into the terminal device 40, the evaluation result of the evaluation model M1.

[0200] The model evaluation unit 540 may acquire the evaluation result of the evaluation model M1. The model evaluation unit 540 may acquire the evaluation result of the evaluation model M1 input into the terminal device 40 by the evaluator E. The model evaluation unit 540 may store, in the model storage unit 502, the evaluation result of the evaluation model M1 in association with the evaluation model M1.

[0201] The evaluator E may use the evaluation result of the evaluation model M1 to select the evaluation model M1 to be used for evaluating the dialogue participant S. The evaluation result of the evaluation model M1 may be shared among a plurality of evaluators E. That is, the evaluation result of the evaluation model M1 input by the first evaluator E may be referred to by the second evaluator E.

[0202] (Model Selection Unit)

The model selection unit 550 selects the evaluation model M1 to be used for

evaluating the dialogue participant S. The model selection unit 550 may select the evaluation model M1 to be used for evaluating the dialogue participant S among a plurality of evaluation models M1 included in the evaluation model group M3. The model selection unit 550 may select a plurality of evaluation models M1 to be used for evaluating the dialogue participant S.

[0203] The model selection unit 550 may select one or more evaluation models M1 designated by the evaluator E. The model selection unit 550 may select a plurality of evaluation models M1 designated by the evaluator E. The evaluator E may select the evaluation model M1 according to scenario information used for evaluating the dialogue participant S. The evaluator E may select one or more evaluation models M1 based on the evaluation result produced by the model evaluation unit 540.

[0204] As an example, when evaluating customer service skills of the dialogue participant S, the evaluator E may select the evaluation model M1 constructed according to evaluation criteria for customer service. When evaluating a product explanation made by the dialogue participant S, the evaluator E may select the evaluation model M1 constructed according to evaluation criteria for product explanation. When evaluating a technical explanation made by the dialogue participant S, the evaluator E may select the evaluation model M1 constructed according to evaluation criteria for technical explanation.

[0205] The model selection unit 550 may transmit the selected evaluation model M1 to the control device 20. The model selection unit 550 may transmit the evaluation model M1 to the control device 20 in response to an instruction from the evaluator E. The control device 20 may store, in the model storage unit 103, the evaluation model M1 received from the management device 50.

[0206] The model selection unit 550 may select the evaluation model M1 associated with the scenario information. When the scenario information is selected by the control device 20, the model selection unit 550 may select the evaluation model M1 associated with the scenario information, and transmit the selected evaluation model M1 to the control device 20.

[0207] The functional configuration of the management device 50 illustrated in FIG. 5 is merely an example, and there can be various functional configuration examples according to applications and purposes. The categorization of storage units, such as the data storage unit 501 and the model

storage unit 502 illustrated in FIG. 5, is merely an example. The categorization of processing units, such as the label assignment unit 510, the feature extraction unit 520, the model training unit 530, the model evaluation unit 540, and the model selection unit 550 illustrated in FIG. 5, is merely an example.

[0208] For example, the data storage unit 501 and the model storage unit 502 may be integrated into a single storage unit. Furthermore, for example, the data storage unit 501 or the model storage unit 502 may be divided into a plurality of storage units.

[0209] For example, at least two of the label assignment unit 510, the feature extraction unit 520, the model training unit 530, the model evaluation unit 540, or the model selection unit 550 may be integrated into a single processing unit. Furthermore, for example, at least one of the utterance acquisition unit 110, the utterance generation unit 120, the utterance presentation unit 130, the dialogue evaluation unit 140, or the result output unit 150 may be divided into a plurality of processing units.

[0210] <Flow of Dialogue Method>

A dialogue method executed by the dialogue system 1000 will be described with reference to FIG. 6. FIG. 6 is a process flow illustrating an example of the dialogue method according to the second embodiment.

[0211] In step S201, the terminal device 40 acquires the video data V from the management device 50. The terminal device 40 displays the video data V on the display device. The evaluator E refers to the video data V displayed on the display device and inputs an evaluation value relating to the video data V. The evaluator E may repeatedly evaluate a plurality of pieces of video data V.

[0212] The terminal device 40 transmits, to the management device 50, the evaluation value input by the evaluator E. Every time the evaluation value relating to the video data V is input, the terminal device 40 may transmit the evaluation value to the management device 50. The terminal device 40 may collectively transmit, to the management device 50, the evaluation values relating to a plurality of pieces of video data V.

[0213] In step S202, the label assignment unit 510 of the management device 50 receives, from the terminal device 40, the evaluation value input by the evaluator E. The label assignment unit 510 assigns a Ground Truth label to the video data V based on the evaluation value input by the evaluator E. Accordingly, training data in which the Ground Truth label is assigned to the

video data V is generated. The label assignment unit 510 transmits the training data to the model training unit 530.

[0214] In step S203, the feature extraction unit 520 of the management device 50 extracts a feature vector from the video data V. The feature extraction unit 520 may extract a feature vector from the video data V to which a Ground Truth label is assigned in step S202. The feature extraction unit 520 transmits the feature vector to the model training unit 530.

[0215] In step S204, the model training unit 530 of the management device 50 receives the training data from the label assignment unit 510. The model training unit 530 receives a feature vector from the feature extraction unit 520.

[0216] The model training unit 530 trains the evaluation model M1 based on the training data and the feature vector. The model training unit 530 stores the trained evaluation model M1 in the model storage unit 502. Accordingly, the trained evaluation model M1 is included in the evaluation model group M3.

[0217] In step S205, the evaluator E performs, on the terminal device 40, an operation to evaluate the evaluation model M1. The terminal device 40 transmits an evaluation request to the management device 50 in response to the operation performed by the evaluator E.

[0218] The model evaluation unit 540 of the management device 50 evaluates the evaluation model M1 in response to the evaluation request from the terminal device 40. Specifically, the model evaluation unit 540 extracts a feature vector from the video data provided in advance, and inputs the extracted feature vector into the evaluation model M1. The evaluation model M1 predicts one or more evaluation values based on the input feature vector, and outputs a predicted value of the evaluation value. The model evaluation unit 540 transmits, to the terminal device 40, the predicted value of the evaluation value. The terminal device 40 displays, on the display device, the predicted value of the evaluation value received from the management device 50.

[0219] The evaluator E may evaluate the performance of the evaluation model M1 by referring to the predicted value of the evaluation value displayed on the display device. The evaluator E may input, into the terminal device 40, the evaluation result of the evaluation model M1. When the evaluation result of the evaluation model M1 is input by the evaluator E, the terminal device 40 transmits, to the management device 50, the input evaluation result of the

evaluation model M1.

[0220] In step S206, the model evaluation unit 540 of the management device 50 receives, from the terminal device 40, the evaluation result of the evaluation model M1. The model evaluation unit 540 stores, in the model storage unit 502, the evaluation result of the evaluation model M1 in association with the evaluation model M1.

[0221] In step S207, the evaluator E performs, on the terminal device 40, an operation to select the evaluation model M1. The terminal device 40 transmits a selection request to the management device 50 in response to the operation performed by the evaluator E.

[0222] In response to the selection request from the terminal device 40, the model selection unit 550 of the management device 50 transmits, to the terminal device 40, a list of the evaluation models M1 included in the evaluation model group M3 and the evaluation result of each of the evaluation models M1. The terminal device 40 displays, on the display device, the list of the evaluation models M1 and the evaluation result, which are received from the management device 50.

[0223] The evaluator E may select, from the list of the evaluation models M1 displayed on the display device, the evaluation model M1 to be used for evaluating the dialogue participant S. When the evaluator E selects the evaluation model M1, the terminal device 40 transmits, to the management device 50, the selection result of the evaluation model M1. The selection result includes information indicating the evaluation model M1 selected by the evaluator E.

[0224] In step S208, the model selection unit 550 of the management device 50 receives, from the terminal device 40, the selection result of the evaluation model M1. Based on the selection result of the evaluation model M1, the model selection unit 550 selects one or more evaluation models M1 to be used for evaluating the dialogue participant S among the evaluation models M1 included in the evaluation model group M3. The model selection unit 550 transmits the selected one or more evaluation models M1 to the control device 20.

[0225] In step S209, the control device 20 receives one or more evaluation models M1 from the management device 50. The control device 20 stores the received one or more evaluation models M1 in the model storage unit 103.

[0226] In step S210, the dialogue participant S executes a dialogue with the

dialogue device 10. The control device 20 controls the dialogue between the dialogue participant S and the dialogue device 10, and stores, in the dialogue storage unit 102, the dialogue data D recording the dialogue between the dialogue participant S and the dialogue device 10. The dialogue system 1000 may execute, as step S210, the process from step S101 to step S110 (see FIG. 3) of the dialogue method according to the first embodiment.

[0227] In step S211, the dialogue evaluation unit 140 of the control device 20 reads, from the dialogue storage unit 102, the dialogue data D relating to the dialogue with the dialogue participant S. The dialogue evaluation unit 140 also reads, from the model storage unit 103, the evaluation model M1 according to the evaluator E.

[0228] The dialogue evaluation unit 140 generates the evaluation model input information based on the read dialogue data D. The dialogue evaluation unit 140 inputs the evaluation model input information into the evaluation model M1. The evaluation model M1 computes one or more evaluation values based on the evaluation model input information. Accordingly, an evaluation result relating to the dialogue participant S is generated. The dialogue evaluation unit 140 stores, in the evaluation result storage unit 104, the evaluation result relating to the dialogue participant S.

[0229] In step S212, the evaluator E performs, on the terminal device 40, an operation to display the evaluation result. As an example, the evaluator E may perform an operation to open the evaluation result of the dialogue participant S on the evaluation screen displayed on the display device of the terminal device 40. The terminal device 40 transmits, to the control device 20, an output request of the evaluation result relating to the dialogue participant S.

[0230] The result output unit 150 of the control device 20 receives, from the terminal device 40, the output request of the evaluation result. The result output unit 150 reads, from the evaluation result storage unit 104, the evaluation result relating to the dialogue participant S. The result output unit 150 generates screen data for displaying the evaluation result relating to the dialogue participant S. The result output unit 150 transmits the screen data to the terminal device 40.

[0231] In step S213, the terminal device 40 receives the screen data from the control device 20. Based on the screen data, the terminal device 40 displays, on the display device, the evaluation result relating to the dialogue participant S.

The evaluator E may evaluate the dialogue participant S by referring to the evaluation result relating to the dialogue participant S displayed on the display device of the terminal device 40. As an example, the evaluator E may determine whether to hire the dialogue participant S by considering the evaluation result of the dialogue participant S. As an example, the evaluator E who is a training instructor may determine whether the training is good or bad by considering the evaluation result of the dialogue participant S. As an example, the evaluator E who is a grader of the examination may determine the score of the dialogue participant S and determine pass or fail of the examination, by considering the evaluation result of the dialogue participant S.

[0232] [Modified Example 1 of Second Embodiment]

The model storage unit 103 of the control device 20 may store a plurality of evaluation models M1 according to a plurality of evaluation criteria. The model storage unit 103 of the control device 20 may store all evaluation models M1 included in the evaluation model group M3 stored in the model storage unit 502 of the management device 50. The plurality of evaluation criteria may include general-purpose evaluation criteria independent of the evaluator E. The plurality of evaluation criteria may include a plurality of general-purpose evaluation criteria.

[0233] The dialogue evaluation unit 140 of the control device 20 may generate, for a single piece of the dialogue data D, an evaluation result according to each of the plurality of evaluation criteria. The dialogue evaluation unit 140 of the control device 20 may generate, for a single piece of the dialogue data D, evaluation results according to all the evaluation criteria by using all the evaluation models M1 stored in the model storage unit 103.

[0234] When the control device 20 generates the evaluation results according to the plurality of evaluation criteria, the evaluator E may refer only to the evaluation results according to his/her own evaluation criteria. When the control device 20 generates evaluation results according to the plurality of evaluation criteria, the evaluator E may refer to the evaluation results according to the plurality of evaluation criteria. In other words, the evaluator E may refer to the evaluation results according to the evaluation criteria of another evaluator E. Whether the evaluator E can refer to the evaluation result according to the evaluation criteria of another evaluator E may be set by a usage plan of the evaluator E. The evaluation criteria capable of being referred

to by the evaluator E may be selected based on usage conditions, input information, or the like of the evaluator E or the other evaluator E.

[0235] The scenario information stored in the control device 20 may be associated with the evaluator E. The scenario information may be generated by the evaluator E. The scenario information may be associated with the evaluator E who generated the scenario information. When generating the evaluation result according to the evaluation criteria of the evaluator E, the dialogue evaluation unit 140 of the control device 20 may generate the evaluation result relating to the dialogue participant S using only the dialogue data D relating to the dialogue executed using the scenario information associated with the evaluator E.

[0236] [Applied Examples]

The dialogue system 1000 according to the second embodiment can provide, for a plurality of evaluators E, a service that provides evaluation results relating to the dialogue participant S. In the present applied example, a configuration in which a service provider collects the dialogue data D relating to the dialogue between the dialogue participant S and the dialogue device 10 and provides evaluation results according to evaluation criteria of the evaluators E to each of the plurality of evaluators E will be described.

[0237] An overall configuration of the dialogue system according to the present applied example will be described with reference to FIG. 7. FIG. 7 is a block diagram illustrating an example of the overall configuration of the dialogue system according to the applied example. Although an example of a service for providing evaluation results relating to the dialogue participant S to two evaluators E (E-1, E-2) will be described in the present applied example, the number of evaluators E is not limited and may be three or more.

[0238] As illustrated in FIG. 7, the dialogue system 1000 includes the dialogue device 10, the control device 20, two evaluation devices 25 (25-1, 25-2), the generation device 30, two terminal devices 40 (40-1, 40-2), and the management device 50.

[0239] The control device 20 and the management device 50 are installed by a service provider. As an example, the control device 20 and the management device 50 may be installed at a facility managed by the service provider or at a remote location, such as a data center. The generation device 30 may be installed by the service provider or an entity other than the service provider.

[0240] The dialogue device 10 is installed for each dialogue participant S. The dialogue device 10 may be installed by the dialogue participant S or the service provider. As an example, the dialogue device 10 may be installed in a space where the dialogue participant S is present.

[0241] The evaluation device 25 and the terminal device 40 are installed for each evaluator E. The evaluation device 25 and the terminal device 40 may be installed by the evaluator E or the service provider. As an example, the evaluation device 25 may be installed in a facility managed by the service provider, may be installed in a remote location, such as a data center or the like, or may be installed in a space where the evaluator E is present. The terminal device 40 may be installed in a space where the evaluator E is present.

[0242] The plurality of evaluators E may be entities having different evaluation criteria. As an example, the evaluator E-1 may be a first company or an employee thereof, and the evaluator E-2 may be a second company or an employee thereof. As an example, the evaluator E-1 may be a company of a first business type or an employee thereof, and the evaluator E-2 may be a company of a second business type or an employee thereof.

[0243] The management device 50 stores the evaluation model group M3. The evaluation model group M3 includes an evaluation model M1-1 and an evaluation model M1-2. The evaluation model M1-1 is constructed based on training data to which the evaluator E-1 assigned Ground Truth labels according to evaluation criteria of the evaluator E-1. The evaluation model M1-2 is constructed based on training data to which the evaluator E-2 assigned Ground Truth labels according to evaluation criteria of the evaluator E-2.

[0244] In the present applied example, although the evaluation model M1 is constructed for each evaluator E, the single evaluation model M1 may be constructed for a plurality of evaluators E. As an example, the same evaluation model M1 may be constructed for the evaluators E having the same evaluation criteria. As an example, the evaluation model M1 may be constructed such that identification information for identifying the evaluator E is an input and an evaluation value according to the evaluation criteria of the evaluator E identified by the input identification information is an output.

[0245] The management device 50 transmits the evaluation model M1-1 to the evaluation device 25-1. The evaluation device 25-1 stores the evaluation model M1-1 received from the management device 50. The management device

50 transmits the evaluation model M1-2 to the evaluation device 25-2. The evaluation device 25-2 stores the evaluation model M1-2 received from the management device 50.

[0246] The dialogue device 10 has a dialogue with the dialogue participant S. The control device 20 controls the dialogue between the dialogue participant S and the dialogue device 10, and stores the dialogue data D recording the dialogue between the dialogue participant S and the dialogue device 10. The control device 20 transmits the dialogue data D to each of the evaluation devices 25-1 and 25-2. Each of the evaluation devices 25-1 and 25-2 stores the dialogue data D.

[0247] The evaluation device 25-1 generates, based on the evaluation model M1-1, an evaluation result relating to the dialogue participant S. The evaluation device 25-1 transmits, to a terminal device 40-1, the evaluation result relating to the dialogue participant S. The terminal device 40-1 presents, to the evaluator E-1, the evaluation result relating to the dialogue participant S. The evaluator E-1 can acquire the evaluation result according to the evaluation criteria of the evaluator E-1.

[0248] The evaluation device 25-2 generates, based on the evaluation model M1-2, the evaluation result relating to the dialogue participant S. The evaluation device 25-2 transmits, to a terminal device 40-2, the evaluation result relating to the dialogue participant S. The terminal device 40-2 presents, to the evaluator E-2, the evaluation result relating to the dialogue participant S. The evaluator E-2 can acquire the evaluation result according to the evaluation criteria of the evaluator E-2.

[0249] The evaluators E-1 and E-2 may prepare respective evaluation criteria by customizing a template of evaluation criteria prepared in advance. The template of the evaluation criteria may be provided by a service provider or the like, or may be provided by another evaluator E. The evaluators E-1 and E-2 may select evaluation criteria to be used for evaluating the dialogue participant S from a plurality of evaluation criteria prepared in advance. The service provider may propose evaluation criteria based on the attributes of the evaluators E-1 and E-2 or the attributes of the dialogue participant S. As an example, the attributes of the evaluators E-1 and E-2 may be business types. As an example, the attributes of the dialogue participant S may be job types.

[0250] The control device 20 may switch scenario information based on the

evaluation criteria of the evaluators E-1 and E-2. The control device 20 may switch scenario information according to the attributes of the evaluators E-1 and E-2, the attributes of the dialogue participant S, or the like.

[0251] For configuring the dialogue system 1000 according to the present applied example, the control device 20 may include the scenario storage unit 101, the dialogue storage unit 102, the utterance acquisition unit 110, the utterance generation unit 120, and the utterance presentation unit 130 included in the control device 20 according to the first embodiment. In addition, the evaluation device 25 may include the dialogue storage unit 102, the model storage unit 103, the evaluation result storage unit 104, the dialogue evaluation unit 140, and the result output unit 150 included in the control device 20 according to the first embodiment.

[0252] [Other Embodiments]

In the above embodiments, modified examples, and applied examples, an example in which a dialogue using speech information and image information is executed between the dialogue participant S and the dialogue device 10 has been described. However, the above embodiments, modified examples, and applied examples are not limited to the dialogue using speech information and image information. As an example, they are applicable to a dialogue using text information. The dialogue using text information may include, for example, text chat, e-mail, messaging service, and the like.

[0253] <Conclusion>

As shown from the above description, the dialogue system 1000 according to the embodiment of the present disclosure acquires information relating to a first utterance uttered by a user, generates information relating to a second utterance based on the information relating to the first utterance, input information, and a machine learning model, and controls an utterance of a dialogue device based on the information relating to the second utterance. The input information includes information indicating a relationship between the user and the dialogue device.

[0254] The relationship between the user and the dialogue device may include at least one of: a relationship in which the user provides a service or a product to the dialogue device; a relationship in which the user instructs the dialogue device; a relationship in which the dialogue device pays remuneration to the user; or a relationship in which the user having information A provides

the information A to a dialogue device that does not have the information A. The relationship between the user and the dialogue device may include at least one of: a relationship between a store clerk and a customer; a relationship between a superior and a subordinate; a relationship between a teacher and a student; a relationship between a speaker and an audience; a relationship between a doctor and a patient; or a relationship between a service professional and a client.

[0255] The information indicating the relationship between the user and the dialogue device may include at least information relating to a role of the dialogue device. The role of the dialogue device may include at least one of: a customer; a subordinate; a student; an audience; a patient; or a client. The information indicating the relationship between the user and the dialogue device may include information relating to the role of the user and the role of the dialogue device.

[0256] The dialogue system 1000 may generate an evaluation result relating to the user based on at least part of the information relating to the first utterance. The dialogue system 1000 may generate, based on dialogue data including the information relating to the first utterance and the information relating to the second utterance, second input information input into a second machine learning model, and input the generated second input information into the second machine learning model, thereby generating an evaluation result.

[0257] The second input information may include one or more features extracted from the dialogue data. The features may include information relating to at least one of: a facial expression of the user; body movement of the user; a voice of the user; smoothness of a dialogue between the user and the dialogue device; or content of a dialogue.

[0258] The dialogue system 1000 may acquire an evaluation value assigned by an evaluator to dialogue data and generate the second machine learning model based on training data in which the evaluation value is assigned to the dialogue data. The dialogue system 1000 may acquire the evaluation value for each of a plurality of evaluation criteria and generate the second machine learning model for each evaluation criterion.

[0259] The dialogue system 1000 may select the second machine learning model to be used for generating the evaluation result from the second machine learning models for each evaluation criterion. The dialogue system 1000 may

select the second machine learning model for the evaluation criterion corresponding to the evaluator.

[0260] The dialogue system 1000 may generate, based on one or more second machine learning models, a plurality of evaluation results for a single piece of dialogue data. The dialogue system 1000 may output an evaluation result relating to the user and information used for the evaluation of the user. The dialogue system 1000 may present, to the user or another user, at least part of the information relating to the first utterance.

[0261] The dialogue system 1000 may generate the input information based on scenario information set in advance. The dialogue system 1000 may set scenario information based on predetermined criteria. The dialogue system 1000 may set scenario information based on an instruction of an evaluator.

[0262] The dialogue system 1000 may repeatedly execute**,** until a predetermined condition is satisfied: acquisition of information relating to an utterance uttered by a user; generation of information relating to an utterance of a dialogue device based on the information relating to the utterance of the user and a machine learning model; and control of the utterance of the dialogue device based on the information relating to the utterance of the dialogue device. The dialogue system 1000 may generate an evaluation result of the user based on dialogue data including information relating to a plurality of utterances of the user and information relating to a plurality of utterances of the dialogue device. The dialogue system 1000 may generate an evaluation result based on the dialogue data and a second machine learning model.

[0263] According to the embodiment of the present disclosure, since the information relating to the utterance of the dialogue device is generated based on the machine learning model, a device having a dialogue with the user can be controlled. In one aspect, according to the embodiment, information for evaluating the user can be acquired through dialogue between the user and the dialogue device. In one aspect, according to the embodiment, since the utterance of the user is responded to based on the machine learning model, the user can have a dialogue with the dialogue device in a natural flow, and the evaluator can evaluate the user appropriately. In another aspect, according to the embodiment, an appropriate evaluation result according to evaluation criteria defined by the evaluator can be acquired.

[0264] [Hardware Configuration of Information Processing Device]

A part or all of the respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) according to the above-described embodiment may be configured by hardware, or by information processing of software (program) executed by a CPU (Central Processing Unit), a GPU (Graphics Processing Unit), or the like. When the respective devices are configured by the information processing of the software, the information processing of the software may be executed by storing the software, which realizes at least a portion of the functions of the respective devices in the above-described embodiment, in a non-transitory storage medium (a non-transitory computer-readable medium, e.g., a CD-ROM (Compact Disc-Read Only Memory), a USB (Universal Serial Bus) memory, or the like) and by reading the non-transitory storage medium in a computer. Also, the software may be downloaded via a communication network. Further, the information processing by software may be executed by hardware with a part or all of the software being mounted in circuits such as an ASIC (Application Specific Integrated Circuit), an FPGA (Field Programmable Gate Array), or the like.

[0265] The storage medium storing the software may be a removable storage medium, such as an optical disk or the like, and may be a fixed storage medium, such as a hard disk, a memory, or the like. Also, the storage medium may be provided internally of the computer (a main storage device, an auxiliary storage device, or the like), or provided externally to the computer.

[0266] FIG. 8 is a block diagram illustrating a hardware configuration of the respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) in the above-described embodiment. As an example, the respective devices include a processor 71, a main storage device 72 (memory), an auxiliary storage device 73 (memory), a network interface 74, and a device interface 75. These are connected to each other via a bus 76 and may be realized as a computer 7.

[0267] The computer 7 of FIG. 8 includes one of each of the constituting elements, but may include multiple constituting elements that are the same. Also, FIG. 8 illustrates one computer 7, but the software may be installed in multiple computers, which may each execute the same or different partial processes of the software. In this case, the computers may be in the form of

distributed computing in which the computers communicate with each other via the network interface 74 or the like, thereby executing the process. That is, the respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) in the above-described embodiment may be configured as a system in which one or more computers execute instructions stored in one or more storage devices, thereby realizing the functions. Also, the respective devices in the above-described embodiment may be configured such that information transmitted from a terminal is processed by one or more computers provided on the cloud and the result of the processing is transmitted to the terminal.

[0268] Various computations of the respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) in the above-described embodiment may be executed in parallel using one or more processors or multiple computers connected via a network. Also, various computations may be distributed among multiple computation cores inside the processor, and executed through parallel processing. Also, a part or all of the processing, means, or the like of the present disclosure may be realized by at least one of a processor and a storage device provided on the cloud communicable with the computer 7 via a network. In this way, the devices in the above-described embodiment may be in the form of parallel computing by one or more computers.

[0269] The processor 71 may be an electronic circuit (e.g., a process circuit, a processing circuit, processing circuitry, a CPU, a GPU, an FPGA, an ASIC, or the like) configured to perform control of a computer, an arithmetic process, or both. Also, the processor 71 may be a general-purpose processor or may be a semiconductor device including a dedicated processing circuit designed to perform specific computation or both of the general-purpose processor and the dedicated processing circuit. Further, the processor 71 may include an optical circuit or computing functions based on quantum computing.

[0270] The processor 71 may perform computation processing based on data and software input from the devices provided internally of the computer 7, and output a computation result and a control signal to the devices. The processor 71 may control the constituting elements of the computer 7 by executing an OS

(Operating System), an application, or the like of the computer 7.

[0271] The respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) in the above-described embodiment may be realized by one or more processors 71. Here, the processor 71 may refer to one or more electronic circuits disposed on a single chip, or may refer to one or more electronic circuits disposed on two or more chips or on two or more devices. When using two or more electronic circuits, each of the electronic circuits may communicate by wire or wirelessly.

[0272] The main storage device 72 may store instructions to be executed by the processor 71, various data, and the like, and the information stored in the main storage device 72 may be read out by the processor 71. The auxiliary storage device 73 is a storage device other than the main storage device 72. These storage devices refer to given electronic components capable of storing electronic information, and may be semiconductor memories. The semiconductor memory may be a volatile memory or a non-volatile memory. The storage device for storing the various data and the like in the respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) in the above-described embodiment may be realized by the main storage device 72 or the auxiliary storage device 73, or may be realized by an internal memory provided internally of the processor 71. For example, the respective storage units in the above-described embodiment may be realized by the main storage device 72 or the auxiliary storage device 73.

[0273] When the respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) in the above-described embodiment are configured with at least one storage device (memory) and at least one processor connected (coupled) to the at least one storage device, at least one processor may be connected to a single storage device. Alternatively, at least one storage device may be connected to a single processor. Also, at least one processor of the multiple processors may be connected to the at least one storage device of the multiple storage devices. Also, such a configuration may be realized by a storage device and a processor included in multiple computers. Further, the configuration may include the storage device integrated with the

processor (e.g., a cache memory including a L1 cache, a L2 cache, or the like).

[0274] The network interface 74 is an interface for connecting to a communication network 8, by wire or wirelessly. The network interface 74 may use an appropriate interface such as an interface conforming to existing communication standards. Exchange of information with an external device 9A connected via the communication network 8 may be performed via the network interface 74. Note that the communication network 8 may be a WAN (Wide Area Network), a LAN (Local Area Network), a PAN (Personal Area Network), or the like, or may be a combination thereof, as long as exchange of information is performed between the computer 7 and the external device 9A. Examples of the WAN include the Internet and the like. Examples of the LAN include the IEEE 802.11, ETHERNET (registered trademark), and the like. Examples of the PAN include Bluetooth (registered trademark), NFC (Near Field Communication), and the like.

[0275] The device interface 75 may be an interface, such as a USB or the like, that directly connects to an external device 9B.

[0276] The external device 9A is a device that is connected to the computer 7 via a network. The external device 9B is a device that is directly connected to the computer 7.

[0277] The external device 9A or the external device 9B may be an input device, as an example. The input device may be a device, such as a camera, a microphone, a motion capture device, various sensors, a keyboard, a mouse, a touch panel, or the like, that provides obtained information to the computer 7. Also, the external device 9A or the external device 9B may be a device including an input unit, a memory, and a processor, such as a personal computer, a tablet terminal, a smartphone, or the like.

[0278] Also, the external device 9A or the external device 9B may be an output device, as an example. The output device may be a display device, such as an LCD (Liquid Crystal Display), an organic EL (Electro Luminescence) panel, or the like, or may be a speaker or the like that outputs sound or the like. Also, the external device 9A or the external device 9B may be a device including an output unit, a memory, and a processor, such as a personal computer, a tablet terminal, a smartphone, or the like.

[0279] Also, the external device 9A or the external device 9B may be a storage device (memory). For example, the external device 9A may be a

network storage or the like, and the external device 9B may be a storage, such as an HDD or the like.

[0280] Also, the external device 9A or the external device 9B may be a device having the functions of a part of the constituting elements of the respective devices (the dialogue device 10, the control device 20, the evaluation device 25, the generation device 30, the terminal device 40, and the management device 50) in the above-described embodiment. That is, the computer 7 may transmit part or all of its processing results to the external device 9A or the external device 9B, or may receive part or all of the processing results from the external device 9A or the external device 9B.

[0281] In the present specification (including the claims), if the expression "at least one of a, b, and c" or "at least one of a, b, or c" is used (including similar expressions), any one of a, b, c, a-b, a-c, b-c, or a-b-c is included. Multiple instances may also be included in any of the elements, such as a-a, a-b-b, and a-a-b-b-c-c. Further, the addition of another element other than the listed elements (i.e., a, b, and c), such as adding d as a-b-c-d, is included.

[0282] In the present specification (including the claims), in a case where an expression, such as "data as an input", "using data", "based on data", "according to data", or "in accordance with data" (including similar expressions) is used, such a case may, unless otherwise noted, encompass a case in which data themselves are used and a case in which data obtained by processing data (e.g., data obtained by adding noise, normalized data, feature extracted from data, and intermediate representation of data) are used. If it is described that any result can be obtained "based on data as an input", "using data", "based on data", "according to data", or "in accordance with data" (including similar expressions), unless otherwise noted, a case in which the result is obtained based on only the data is included, and a case in which the result is obtained affected by another data other than the data, factors, conditions, and/or states is included. If it is described that "data are output" (including similar expressions), unless otherwise noted, a case in which data themselves are used as an output is included, and a case in which data obtained by processing data in some way (e.g., data obtained by adding noise, normalized data, feature extracted from data, and intermediate representation of various data) are used as an output is included.

[0283] In the present specification (including the claims), if the terms

"connected" and "coupled" are used, the terms are intended as non-limiting terms that include any of direct, indirect, electrically, communicatively, operatively, and physically connected/coupled. Such terms should be interpreted according to a context in which the terms are used, but a connected/coupled form that is not intentionally or naturally excluded should be interpreted as being included in the terms without being limited.

[0284] In the present specification (including the claims), if the expression "A configured to B" is used, a case in which a physical structure of the element A has a configuration that can perform the operation B, and a permanent or temporary setting/configuration of the element A is configured/set to actually perform the operation B may be included. For example, if the element A is a general-purpose processor, the processor may have a hardware configuration that can perform the operation B and be configured to actually perform the operation B by setting a permanent or temporary program (i.e., an instruction). If the element A is a dedicated processor or a dedicated arithmetic circuit, a circuit structure or the like of the processor may be implemented so as to actually perform the operation B irrespective of whether the control instruction and the data are actually attached.

[0285] In the present specification (including the claims), if a term indicating containing or possessing (e.g., "comprising/including" and "having") is used, the term is intended as an open-ended term, including an inclusion or possession of an object other than a target object indicated by the object of the term. If the object of the term indicating an inclusion or possession is an expression that does not specify a quantity or that suggests a singular number (i.e., an expression using "a" or "an" as an article), the expression should be interpreted as being not limited to a specified number.

[0286] In the present specification (including the claims), even if an expression such as "one or more" or "at least one" is used in a certain description, and an expression that does not specify a quantity or that suggests a singular number (i.e., an expression using "a" or "an" as an article) is used in another description, it is not intended that the latter expression indicates "one". Generally, an expression that does not specify a quantity or that suggests a singular number (i.e., an expression using "a" or "an" as an article) should be interpreted as being not necessarily limited to a particular number.

[0287] In the present specification, if it is described that a particular

advantage/result is obtained in a particular configuration included in an embodiment, unless there is a particular reason, it should be understood that the advantage/result may be obtained in another embodiment or other embodiments including the configuration. It should be understood, however, that the presence or absence of the advantage/result generally depends on various factors, conditions, states, and/or the like, and that the advantage/result is not necessarily obtained by the configuration. The advantage/result is merely an advantage/result that results from the configuration described in the embodiment when various factors, conditions, and/or states are satisfied, and is not necessarily obtained in the claimed invention that defines the configuration or a similar configuration.

[0288] In the present specification (including the claims), if a plurality of hardware components perform predetermined processes, each hardware component may cooperate to perform the predetermined processes, or some of the hardware components may perform all of the predetermined processes. Additionally, some of the hardware components may perform some of the predetermined processes while other hardware components may perform the remainder of the predetermined processes. In the present specification (including the claims), if an expression such as “one or more hardware components perform a first process and the one or more hardware components perform a second process” (including similar expressions) is used, the hardware that performs the first process may be the same as or different from the hardware that performs the second process. That is, the hardware that performs the first process and the hardware that performs the second process may be included in the one or more hardware components. The hardware may include an electronic circuit, a device including an electronic circuit, or the like.

[0289] In the present specification (including the claims), if multiple storage devices (memories) store data, each of the multiple storage devices may store only a portion of the data or may store an entirety of the data. Also, a configuration in which some of the multiple storage devices store data may be included.

[0290] In the present specification (including the claims), such terms as “first”, “second”, and the like are used merely as a way of distinguishing two or more elements, and are not necessarily intended to impose any technical meaning, such as a temporal aspect, a spatial aspect, an order, a quantity, or the

like, on the target. Therefore, for example, references to a first element and a second element do not necessarily mean, for example, that only two elements can be employed in the context, the first element needs to precede the second element, and the first element must exist in order for the second element to exist.

[0291] Although the embodiments of the present disclosure have been described in detail above, the present disclosure is not limited to the individual embodiments described above. Various additions, modifications, substitutions, partial deletions, and the like may be made without departing from the conceptual idea and spirit of the invention derived from the contents defined in the claims and the equivalents thereof. For example, in the embodiments described above, when numerical values or mathematical formulae are used for description, these are used for illustrative purposes and do not limit the scope of the present disclosure. Additionally, the orders of operations described in the embodiments are illustrative and do not limit the scope of the present disclosure.

[0292] In the disclosed technique, embodiments according to the following clauses are conceivable.

[0293] (Clause 1)

A dialogue system, including:

at least one memory; and

at least one processor, wherein

the at least one processor is configured to:

acquire information relating to a first utterance uttered by a user;

generate information relating to a second utterance based on the information relating to the first utterance, input information, and a machine learning model; and

control an utterance of a dialogue device based on the information relating to the second utterance, and

the input information includes information indicating a relationship between the user and the dialogue device.

[0294] (Clause 2)

The dialogue system according to clause 1, wherein

the relationship includes at least one of: a relationship in which

the user provides a service or a product to the dialogue device; a relationship in which the user instructs the dialogue device; a relationship in which the dialogue device pays remuneration to the user; or a relationship in which the user having information A provides the information A to a dialogue device that does not have the information A.

[0295] (Clause 3)

The dialogue system according to clause 2, wherein

the relationship includes at least one of: a relationship between a store clerk and a customer; a relationship between a superior and a subordinate; a relationship between a teacher and a student; a relationship between a speaker and an audience; a relationship between a doctor and a patient; or a relationship between a service professional and a client.

[0296] (Clause 4)

The dialogue system according to any one of clauses 1 to 3, wherein

the information indicating the relationship between the user and the dialogue device includes, at least, information relating to a role of the dialogue device.

[0297] (Clause 5)

The dialogue system according to clause 4, wherein

the role of the dialogue device includes at least one of: a customer; a subordinate; a student; an audience; a patient; or a client.

[0298] (Clause 6)

The dialogue system according to any one of clauses 1 to 5, wherein

the at least one processor is further configured to:

generate an evaluation result relating to the user based on at least part of the information relating to the first utterance.

[0299] (Clause 7)

The dialogue system according to clause 6, wherein

the at least one processor is further configured to:

generate, based on dialogue data including the information relating to the first utterance and the information relating to the second utterance, second input information input into a second machine learning model; and

generate the evaluation result by inputting the second input information into the second machine learning model.

[0300] (Clause 8)

The dialogue system according to clause 7, wherein

the second input information includes one or more features extracted from the dialogue data.

[0301] (Clause 9)

The dialogue system according to clause 8, wherein

the feature includes information relating to at least one of: a facial expression of the user; body movement of the user; a voice of the user; smoothness of a dialogue between the user and the dialogue device; or content of a dialogue.

[0302] (Clause 10)

The dialogue system according to any one of clauses 6 to 9, wherein

the at least one processor is further configured to:

acquire an evaluation value provided by an evaluator for dialogue data; and

generate a second machine learning model based on training data in which the evaluation value is assigned to the dialogue data.

[0303] (Clause 11)

The dialogue system according to clause 10, wherein

the at least one processor is further configured to:

acquire the evaluation value for each of a plurality of evaluation criteria; and

generate the second machine learning model for each of the evaluation criteria.

[0304] (Clause 12)

The dialogue system according to any one of clauses 7 to 11, wherein

the at least one processor is further configured to:

select, from among the second machine learning models for the respective evaluation criteria, the second machine learning model to be used for generating the evaluation result.

[0305] (Clause 13)

The dialogue system according to clause 12, wherein

the at least one processor is further configured to:

select the second machine learning model for the evaluation criterion corresponding to an evaluator.

[0306] (Clause 14)

The dialogue system according to any one of clauses 7 to 13, wherein

the at least one processor is further configured to:

generate a plurality of the evaluation results for a single piece of the dialogue data based on the one or more second machine learning models.

[0307] (Clause 15)

The dialogue system according to any one of clauses 6 to 14, wherein

the at least one processor is further configured to:

output the evaluation result and information used for evaluating the user.

[0308] (Clause 16)

The dialogue system according to any one of clauses 1 to 15, wherein

the at least one processor is further configured to:

present, to the user or another user, at least part of the information relating to the first utterance.

[0309] (Clause 17)

The dialogue system according to any one of clauses 1 to 16, wherein

the at least one processor is further configured to:

generate the input information based on scenario information set in advance.

[0310] (Clause 18)

The dialogue system according to clause 17, wherein

the at least one processor is further configured to:

set the scenario information based on predetermined criteria.

[0311] (Clause 19)

The dialogue system according to clause 18, wherein

the at least one processor is further configured to:

set the scenario information based on an instruction of an evaluator.

[0312] (Clause 20)

The dialogue system according to any one of clauses 1 to 19, wherein

the at least one processor is further configured to repeatedly execute, until a predetermined condition is satisfied:

the acquisition of the information relating to the utterance uttered by the user;

the generation of information relating to an utterance of the dialogue device based on information relating to an utterance of the user, the input information, and the machine learning model; and

the control of the utterance of the dialogue device based on the information relating to the utterance of the dialogue device.

[0313] (Clause 21)

The dialogue system according to clause 20, wherein

the at least one processor is further configured to:

generate an evaluation result of the user based on dialogue data including information relating to a plurality of utterances of the user and information relating to a plurality of utterances of the dialogue device.

[0314] (Clause 22)

The dialogue system according to clause 21, wherein

the at least one processor is further configured to:

generate the evaluation result based on the dialogue data and a second machine learning model.

[0315] (Clause 23)

A dialogue method executed by at least one processor, the dialogue method including:

acquiring information relating to a first utterance uttered by a user;

generating information relating to a second utterance based on the information relating to the first utterance, input information, and a machine learning model; and

controlling an utterance of a dialogue device based on the information relating to the second utterance, wherein

the input information includes information indicating a relationship between the user and the dialogue device.

[0316] (Clause 24)

A program causing at least one processor to execute a process including:

acquiring information relating to a first utterance uttered by a user;

generating information relating to a second utterance based on the information relating to the first utterance, input information, and a machine learning model; and

controlling an utterance of a dialogue device based on the information relating to the second utterance, wherein

the input information includes information indicating a relationship between the user and the dialogue device.

[0317] This application claims priority to Provisional Patent Application No. 63/564,537, filed with the United States Patent and Trademark Office, on March 13, 2024, Provisional Patent Application No. 63/648,341, filed with the United States Patent and Trademark Office, on May 16, 2024, and Japanese Patent Application No. 2024-122898, filed with the Japan Patent Office, on July 29, 2024, the entire contents of each of which are incorporated herein by reference.

REFERENCE SIGNS LIST

[0318]

10: Dialogue device
20: Control device
25: Evaluation device
30: Generation device
40: Terminal device
50: Management device
101: Scenario storage unit
102: Dialogue storage unit
103: Model storage unit

104: Evaluation result storage unit
110: Utterance acquisition unit
120: Utterance generation unit
130: Utterance presentation unit
140: Dialogue evaluation unit
150: Result output unit
501: Data storage unit
502: Model storage unit
510: Label assignment unit
520: Feature extraction unit
530: Model training unit
540: Model evaluation unit
550: Model selection unit
1000: Dialogue system
D: Dialogue data
M1: Evaluation model
M2: Generative model

CLAIMS

[Claim 1]

An information processing system, comprising:

at least one memory; and

at least one processor, wherein

the at least one processor is configured to repeatedly execute, until a predetermined condition is satisfied:

acquisition of information relating to an utterance of a user;

generation of information relating to an utterance of a dialogue device based on the information relating to the utterance of the user and a machine learning model; and

control of the utterance of the dialogue device based on the information relating to the utterance of the dialogue device,

the generation of the information relating to the utterance of the dialogue device is executed based on information indicating a relationship between the user and the dialogue device, and

the relationship includes at least one of: a relationship in which the user provides a service or a product to the dialogue device; a relationship in which the user instructs the dialogue device; a relationship in which the dialogue device pays remuneration to the user; or a relationship in which the user having information A provides the information A to a dialogue device that does not have the information A.

[Claim 2]

The information processing system as claimed in claim 1, wherein

the relationship includes at least one of: a relationship in which the user is a store clerk and the dialogue device is a customer; a relationship in which the user is a superior and the dialogue device is a subordinate; a relationship in which the user is a teacher and the dialogue device is a student; a relationship in which the user is a speaker and the dialogue device is an audience; a relationship in which the user is a doctor and the dialogue device is a patient; or a relationship in which the user is a service professional and the dialogue device is a client.

[Claim 3]

The information processing system as claimed in claim 1, wherein the information indicating the relationship between the user and the dialogue device includes, at least, information relating to a role of the dialogue device.

[Claim 4]

The information processing system as claimed in claim 3, wherein the role of the dialogue device includes at least one of: a customer; a subordinate; a student; an audience; a patient; or a client.

[Claim 5]

The information processing system as claimed in any one of claims 1 to 4, wherein

input information input into the machine learning model includes a system prompt, and

the system prompt includes:

the information indicating the relationship between the user and the dialogue device; and

at least, information indicating a situation of a dialogue, a prerequisite of a dialogue, or a constraint for an utterance.

[Claim 6]

The information processing system as claimed in claim 5, wherein the system prompt includes, at least, the constraint for the utterance, and

the constraint for the utterance includes a character-count limit of the utterance of the dialogue device.

[Claim 7]

The information processing system as claimed in claim 5 or 6, wherein

the at least one processor is further configured to:

change the system prompt used for the generation of the

information relating to the utterance of the dialogue device based on a degree of progress of the dialogue.

[Claim 8]

The information processing system as claimed in claim 7, wherein

the degree of progress of the dialogue includes, at least, a number of turns in the dialogue or an elapsed time from start of the dialogue.

[Claim 9]

The information processing system as claimed in any one of claims 1 to 8, wherein

after completion of the repeated execution, one or more evaluation results relating to the user are generated using dialogue data stored in the at least one memory and including at least the information relating to the utterance of the user.

[Claim 10]

The information processing system as claimed in claim 9, wherein

the at least one processor is further configured to:

generate the one or more evaluation results based on one or more second machine learning models.

[Claim 11]

The information processing system as claimed in claim 10, wherein

the at least one processor is further configured to:

generate, based on at least part of the dialogue data, input information input into each of the one or more second machine learning models; and

generate the one or more evaluation results by inputting the input information into each of the one or more second machine learning models.

[Claim 12]

The information processing system as claimed in claim 11,

wherein

the input information includes one or more features extracted from the at least part of the dialogue data, and

the one or more features include, at least, information relating to at least one of: a facial expression of the user; body movement of the user; a voice of the user; smoothness of a dialogue between the user and the dialogue device; or content of a dialogue.

[Claim 13]

The information processing system as claimed in claim 11, wherein

the input information includes one or more features extracted from at least part of the dialogue data, and

the one or more features include, at least:

a feature extracted from speech information of the user included in the dialogue data;

a feature extracted from image information of the user included in the dialogue data; or

a feature extracted from dialogue content included in the dialogue data.

[Claim 14]

The information processing system as claimed in any one of claims 10 to 13, wherein

the at least one processor is further configured to:

generate the second machine learning model using training data including other dialogue data and an evaluation value provided by an evaluator for the other dialogue data; and

generate the evaluation result using the generated second machine learning model.

[Claim 15]

The information processing system as claimed in any one of claims 10 to 14, wherein

the second machine learning model is the same model as the

machine learning model or a model generated based on the machine learning model.

[Claim 16]

The information processing system as claimed in any one of claims 1 to 15, wherein

the at least one processor is further configured to:

acquire the information indicating the relationship between the user and the dialogue device based on scenario information set based on predetermined criteria.

[Claim 17]

The information processing system as claimed in claim 16, wherein

the at least one processor is further configured to:

set the scenario information based on at least one of a business type or a job type.

[Claim 18]

The information processing system as claimed in claim 17, wherein

the at least one processor is further configured to:

set the scenario information based on an attribute of the user.

[Claim 19]

The information processing system as claimed in claim 18, wherein

the attribute of the user includes, at least, a resume or a work-history document of the user.

[Claim 20]

The information processing system as claimed in any one of claims 16 to 19, wherein

the scenario information includes guidance information, and

the at least one processor is further configured to:

cause the dialogue device to output the guidance information before start of the repeated processing.

[Claim 21]

The information processing system as claimed in any one of claims 16 to 19 depending on claim 10, wherein

the at least one processor is further configured to:

select, based on the set scenario information, the one or more second machine learning models to be used for generating the one or more evaluation results; and

generate the one or more evaluation results using the selected one or more second machine learning models.

[Claim 22]

An information processing method using the information processing system of any one of claims 1 to 21, the information processing method comprising:

repeatedly executing, until the predetermined condition is satisfied:

the acquisition of the information relating to the utterance of the user;

the generation of the information relating to the utterance of the dialogue device based on the information relating to the utterance of the user and the machine learning model; and

the control of the utterance of the dialogue device based on the information relating to the utterance of the dialogue device.

[Claim 23]

A program causing at least one processor to execute the information processing method of claim 22.