UIUC Chatbot Evaluation Results

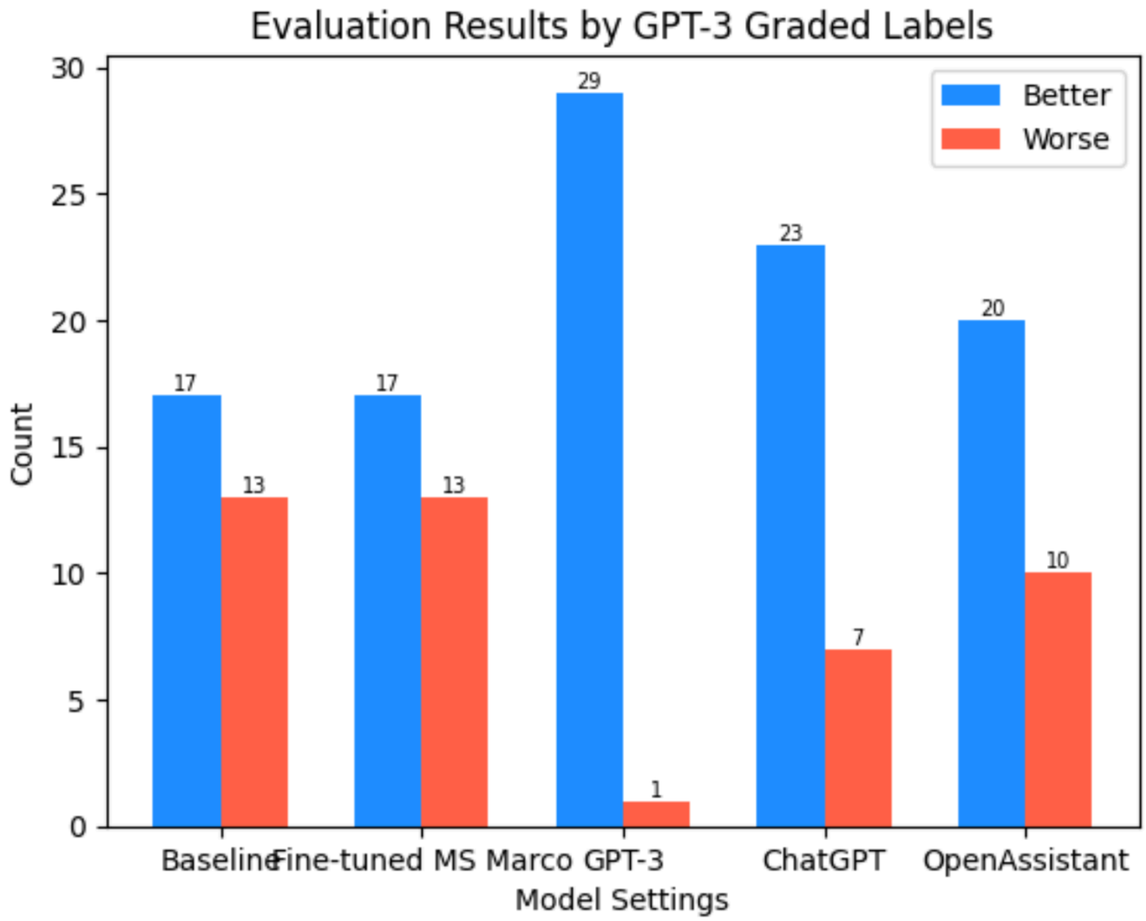

Figure1: GPT-3 evaluation results of different models

The above bar chart shows the evaluation results of chatbot output answers using GPT-3. The overall evaluation data samples are 30 question-answer pairs. The ‘Better’ represents the model’s generated answer is better than the student labeled ground truth, vice versa. All ‘Better’ and ‘Worse’ labels were graded by GPT-3. Specifically, given a question and its ground truth answer, we let GPT-3 to decide whether the new generated answer by our model is ‘Better’ or ‘Worse’ than the ground truth one.

Here are some examples of each model’s output:

Table1: Baseline 'Better' examples

Question	What does the uninterrupted counting cycle label mean?
Ground Truth Answer	The uninterrupted counting cycle is an infinite loop that continues until the H button is pressed.
Baseline (Fine-tuned Flan T5)	The uninterrupted counting cycle label means that the computer will continue processing instructions, carrying out the instruction cycle again and again, ad nauseum.
Question	What do the buttons 'halt' and 'go' do?
Ground Truth Answer	The 'halt' button makes the machine stop and the 'go' button makes it start again.
Baseline (Fine-tuned Flan T5)	A user can stop the counter in any state by pressing the "halt" but- ton, causing the counter to retain its current value until the user presses the "go" button.
Question	What is the purpose of an FSM?
Ground Truth Answer	The purpose of an FSM is to model the behavior of a system.
Baseline (Fine-tuned Flan T5)	An FSM comprises five parts: a finite set of states, a set of possible inputs, a set of possible outputs, a set of transition rules, and methods for calculating outputs.
Question	How does an FSM work?
Ground Truth Answer	A finite state machine (FSM) is a mathematical model of computation used to design both computer programs and sequential logic circuits. It is conceived as an abstract machine that can be in one of a finite number of user-specified states. An FSM is defined by a list of its states, its initial state, and the conditions for each transition.
Baseline (Fine-tuned Flan T5)	An FSM comprises five parts: a finite set of states, a set of possible inputs, a set of possible outputs, a set of transition rules, and methods for calculating outputs.

Table2: Baseline 'Worse' examples

Question	What does a two-bit Gray code represent?
Ground Truth Answer	A two-bit Gray code is a binary code used to represent numbers in which adjacent values differ in only one bit. For example, the numbers 0, 1, 2, and 3 can be represented as 00, 01, 11, and 10.
Baseline (Fine-tuned Flan T5)	0 or 1
Question	What is the purpose of an abstract model?
Ground Truth Answer	The purpose of an abstract model is to provide a simplified representation of a system. Abstract models are often used to design and analyze complex systems..
Baseline (Fine-tuned Flan T5)	To better understand their job.
Question	What is the purpose of using a Gray code basis for the state sequence?
Ground Truth Answer	The Gray code basis for the state sequence is used to simplify the design complexity by explicitly identifying and making use of common algebraic terms and sub-expressions for different outputs.
Baseline (Fine-tuned Flan T5)	Gray code basis for the state sequence.
Question	Given that we need three flip-flops, can the counter's internal state directly be used as our output values?
Ground Truth Answer	Yes, the counter's internal state can be used as our output values because the colors we need to produce are all unique bit patterns.
Baseline (Fine-tuned Flan T5)	No

Table3: GPT-3 ‘Better’ examples

Question	What does a two-bit Gray code represent?
Ground Truth Answer	A two-bit Gray code is a binary code used to represent numbers in which adjacent values differ in only one bit. For example, the numbers 0, 1, 2, and 3 can be represented as 00, 01, 11, and 10.
GPT-3	A two-bit Gray code is a sequence of binary numbers where each successive number has only one bit different from the previous number. It is commonly used in digital circuits to represent the state of a finite state machine.
Question	What does the uninterrupted counting cycle label mean?
Ground Truth Answer	The uninterrupted counting cycle is an infinite loop that continues until the H button is pressed.
GPT-3	The uninterrupted counting cycle label indicates that the clock circuit will continuously repeat each instruction cycle until it is interrupted. The clock circuit consists of a clock generator and a RUN latch, which control the cycle of instructions. This is done by providing a steady pulse from the crystal oscillator to create regular cycles for each instruction.
Question	What do the buttons 'halt' and 'go' do?
Ground Truth Answer	The 'halt' button makes the machine stop and the 'go' button makes it start again.
GPT-3	The 'halt' button stops the counter and retains its current value, while the 'go' button resumes counting from where it left off. For example, pressing 'halt' in state COUNT B will cause the counter to remain in state HALT C, and pressing 'go' in state HALT C will take the system to state COUNT C.
Question	How does an FSM work?
Ground Truth Answer	A finite state machine (FSM) is a mathematical model of computation used to design both computer programs and sequential logic circuits. It is conceived as an abstract machine that can be in one of a finite number of user-specified states. An FSM is defined by a list of its states, its initial state, and the conditions for each transition.
GPT-3	A finite state machine (FSM) is a model for understanding the behavior of a system by describing the system as occupying one of a finite set of states, moving between these states in response to external inputs, and producing external outputs. The FSM includes five components: a finite set of states, a set of possible inputs, a set of possible outputs, transition rules from one state to another (or self-loops), and methods for calculating outputs. When an FSM is implemented as digital logic, each state must be represented using bits; all inputs must be translated into bits; and all outputs must be translated into bits. Transition rules are complete so that given any state and any pattern or input bits there is always an assigned transition rule leading to another state. Output calculation can then be reduced to Boolean logic expressions. As an example, consider the abstract keyless entry system given in the context: There are four states - Locked (vehicle locked); Driver (driver's door unlocked); Unlocked (all doors unlocked); Alarm (alarm sounding). Inputs include push \"unlock\", push \"lock\", push \"panic\". In this case when starting from Locked pushing unlock once leads to Driver while twice leads to Unlocked while pressing lock will lead back to Locked regardless; pressing panic will lead to Alarm sounding regardless. This behavior can be outlined using next-state tables which map current input/state combination into the next state for the FSM.

Table4: GPT-3 ‘Worse’ examples

Question	Given that we need three flip-flops, can the counter's internal state directly be used as our output values?
Ground Truth Answer	Yes, the counter's internal state can be used as our output values because the colors we need to produce are all unique bit patterns.
GPT-3	No, the counter's internal state cannot be used directly as output values because it depends on the prices and other conditions. We need to design with components and follow our rules to get the desired output.

Table5: ChatGPT ‘Better’ examples

Question	What does a two-bit Gray code represent?
Ground Truth Answer	A two-bit Gray code is a binary code used to represent numbers in which adjacent values differ in only one bit. For example, the numbers 0, 1, 2, and 3 can be represented as 00, 01, 11, and 10.
ChatGPT	A two-bit Gray code represents a sequence of binary numbers in which only one bit changes at any given time. For example, the sequence 00, 01, 11, 10 is a two-bit Gray code because only one bit changes between adjacent numbers. This type of code is often used in digital electronics to minimize errors caused by switching multiple bits simultaneously.
Question	What do the buttons 'halt' and 'go' do?
Ground Truth Answer	The 'halt' button makes the machine stop and the 'go' button makes it start again.
ChatGPT	Pressing the \"halt\" button stops the counter in its current state, while pressing the \"go\" button resumes counting from the state where it was halted.
Question	What is the purpose of an FSM?
Ground Truth Answer	The purpose of an FSM is to model the behavior of a system.
ChatGPT	The purpose of an FSM (Finite State Machine) is to represent a system that can be in one of a finite number of states and can transition from one state to another in response to inputs. It is used as a model for designing digital circuits and software systems, where the behavior of the system can be defined in terms of states, inputs, outputs, and transitions between states. The abstraction used in FSM design corresponds almost directly to the implementation using digital logic.
Question	How does an FSM work?
Ground Truth Answer	A finite state machine (FSM) is a mathematical model of computation used to design both computer programs and sequential logic circuits. It is conceived as an abstract machine that can be in one of a finite number of user-specified states. An FSM is defined by a list of its states, its initial state, and the conditions for each transition.
ChatGPT	A Finite State Machine (FSM) is a model that describes the behavior of a system by defining the system as being in one of a finite number of states, moving between these states in response to external inputs, and producing outputs. In other words, an FSM is like a black box that takes inputs and produces outputs based on its current state. \n\nAn FSM has five parts: \n1. A finite set of states\n2. A set of possible inputs\n3. A set of possible outputs\n4. A set of transition rules\n5. Methods for calculating outputs\n\nIn an FSM, when it receives an input while in a particular state, it follows a transition rule to move to another state and produce an output based on its new state.\n\nFor example, consider a keyless entry system for a car with three abstract states: \"locked,\" \"driver unlocked,\" and \"fully unlocked.\" When the user presses the \"unlock\" button while the car is locked (\"locked\" state), the FSM transitions to the \"driver unlocked\" state and unlocks only the driver's door (output). If they press it again, it transitions to the \"fully unlocked\" state and unlocks all doors (output).\n\nFSMs can be implemented using digital systems where all inputs, outputs, and states are represented as patterns using bits or Boolean logic expressions.\n\nIn conclusion, FSMS are useful models for understanding how systems behave by breaking them down into different states based on their inputs/outputs and transition rules.

Table6: ChatGPT 'Worse' examples

Question	What does the uninterrupted counting cycle label mean?
Ground Truth Answer	The uninterrupted counting cycle is an infinite loop that continues until the H button is pressed.
ChatGPT	The given context does not provide any information about an \"uninterrupted counting cycle label\". Therefore, it is impossible to generate an objective and logical answer to this question based on the given context.
Question	What is the purpose of the ``halt'' button?
Ground Truth Answer	The ``halt'' button is used to stop the counter. If the counter is halted, the ``halt'' button has no additional effect. The table implies that if the user presses the ``halt'' button, the counter will retain its current value until the user presses the ``go'' button.
ChatGPT	The context does not provide any information about the \"halt\" button. Therefore, it is not possible to generate an objective and logical answer.
Question	What does the 'x' mean in the K-map?
Ground Truth Answer	The 'x' in the K-map indicates a \"don't care\" condition. This means that the output will be the same regardless of whether the input is a 0 or a 1.
ChatGPT	The context mentions that the 'x' in the K-map is not discussed yet and will be covered next week. Therefore, it is unclear what the 'x' means in the K-map at this point.
Question	Given that we need three flip-flops, can the counter's internal state directly be used as our output values?
Ground Truth Answer	Yes, the counter's internal state can be used as our output values because the colors we need to produce are all unique bit patterns.
ChatGPT	It's not clear from the context what the counter is being used for or what the output values should represent. Therefore, it's difficult to give a definitive answer to this question.

Table7: OpenAssistant 'Better' examples

Question	What does the uninterrupted counting cycle label mean?
Ground Truth Answer	The uninterrupted counting cycle is an infinite loop that continues until the H button is pressed.
OpenAssistant	An uninterrupted counting cycle is a sequence of numbers where each number is obtained by adding the previous number in the sequence to a constant value. For example, in the sequence 1, 2, 3, 4, 5, the constant value is 0, and each number is obtained by adding the previous number to 0. The uninterrupted counting cycle label is a way of representing this sequence of numbers in a mathematical formula.\n\nThe uninterrupted counting cycle label is written as a formula that includes the constant value and the starting number of the sequence. For example, the label for the sequence 1, 2, 3, 4, 5 is given by the formula:\n\nc_n = n + 0\n\nwhere c_n is the nth uninterrupted counting cycle label and n is the current number in the sequence.\n\nThe uninterrupted counting cycle label is useful in mathematical problems that involve sequences of numbers, such as finding the nth term in a sequence, determining the length of a sequence, and solving problems involving addition and subtraction of sequences.
Question	What do the buttons 'halt' and 'go' do?
Ground Truth Answer	The 'halt' button makes the machine stop and the 'go' button makes it start again.
OpenAssistant	Halt and Go are buttons found on many electronic devices, including computers, smartphones, and video game consoles. They are used to pause or resume a game or application.\n\nHalt is a button that pauses the current activity in a game or application. When you press the Halt button, the game or application will stop running and you will be returned to the main menu or game selection screen.\n\nGo is a button that resumes the activity that was paused by the Halt button. When you press the Go button, the game or application will continue from where it was stopped.\n\nIt's important to note that the exact behavior of the Halt and Go buttons may vary depending on the device and the game or application being played.
Question	What is the purpose of an FSM?
Ground Truth Answer	The purpose of an FSM is to model the behavior of a system.
OpenAssistant	An FSM (Finite State Machine) is a model used to represent and solve problems in a finite number of states. It is a graphical representation of a system that shows how the system can transition from one state to another based on certain events or conditions.\n\nAn FSM can be used to model a variety of systems, such as automatons, control systems, and communication networks. It is a useful tool for analyzing and solving problems that involve the flow of data or the control of a system.\n\nFor example, consider a traffic light system. An FSM could be used to model the behavior of the traffic light, showing how it changes from green to yellow to red based on the presence of vehicles, pedestrians, or other factors. The FSM could also be used to determine the best course of action for the traffic light in different situations, such as when there is a lot of traffic or when a vehicle approaches the intersection at a high speed.\n\nIn general, an FSM is a useful tool for analyzing and solving problems that involve the flow of data or the control of a system. By using an FSM, you can gain a better understanding of the system and identify potential problems or issues that may arise.
Question	How does an FSM work?
Ground Truth Answer	A finite state machine (FSM) is a mathematical model of computation used to design both computer programs and sequential logic circuits. It is conceived as an abstract machine that can be in one of a finite number of user-specified states. An FSM is defined by a list of its states, its initial state, and the conditions for each transition.
OpenAssistant	An FSM (Finite State Machine) is a model used to represent a system that can change between different states over time. It is a way of breaking down a complex system into smaller, more manageable parts, and it is used to model how a system will respond to different inputs.\n\nIn an FSM, there are a finite number of states and a set of transitions between those states. Each state represents a possible state of the system, and each transition represents a way in which the system can change from one state to another.\n\nFor example, consider a thermostat that controls the temperature in a room. The thermostat can be modeled as an FSM, with a set of states representing the different temperatures that the room can reach, and a set of transitions representing the actions the thermostat can take to change between those states.\n\nIn the thermostat example, the states could be \"cold\", \"medium\", and \"hot\", and the transitions could be \"heat up to medium\", \"heat to hot\", and \"cool down to cold\". By modeling the thermostat this way, we can predict how the system will respond to different inputs and make decisions based on those predictions.\n\nFSMs can be used in a variety of applications, from computer systems to industrial processes. They can be used to model the behavior of complex systems, to predict how they will respond to different inputs, and to make decisions based on those predictions.

Table8: OpenAssistant ‘Worse’ examples

Question	What is the purpose of the ``halt'' button?
Ground Truth Answer	The ``halt'' button is used to stop the counter. If the counter is halted, the ``halt'' button has no additional effect. The table implies that if the user presses the ``halt'' button, the counter will retain its current value until the user presses the ``go'' button.
OpenAssistant	The purpose of the \"halt\" button is to stop the program or task that is currently running. It is often used in programming to pause the execution of a program and allow the user to take action, such as making a change to the code. The \"halt\" button can also be used in other situations, such as when a program is experiencing a problem and needs to be stopped to allow the user to troubleshoot the issue.
Question	What does the 'x' mean in the K-map?
Ground Truth Answer	The 'x' in the K-map indicates a \"don't care\" condition. This means that the output will be the same regardless of whether the input is a 0 or a 1.
OpenAssistant	In the K-map, the 'x' symbol represents a question mark. The K-map is a tool used to help students organize their knowledge and understand complex concepts. When students encounter a new concept or term, they can use the K-map to identify the main ideas, relationships, and concepts that are related to the topic. The 'x' symbol on the K-map indicates that a question has not been answered or is still un understood. The student can then use the information provided by the K-map to research and answer the question more thoroughly.
Question	What does pressing the 'halt' button do when the counter is in a 'halted' state?
Ground Truth Answer	Pressing the 'halt' button when the counter is in a 'halted' state has no effect.
OpenAssistant	Halted: The counter is in a halted state, meaning that it has reached a certain point and cannot be incremented any further. To restart the counter, you would need to press the 'halt' button again.\n\nUnhalted: If the counter is not in a halted state, pressing the 'halt' button will not have any effect. The counter will continue to increment as normal.
Question	What is the difference between 'counting' and 'halted' states?
Ground Truth Answer	The 'counting' state represents uninterrupted counting, in which the counter cycles through these states indefinitely. A user can stop the counter in any state by pressing the ``halt'' button, causing the counter to retain its current value until the user presses the ``go'' button. The 'halted' state represents when a user has pressed the ``halt'' button, and the counter is not counting.
OpenAssistant	In computing, the \"counting\" state is a state in which a processor is executing instructions and performing some type of calculation. In this state, the processor is continuously executing instructions and making progress towards a desired result.\n\nOn the other hand, the \"halted\" state is a state in which the processor has stopped executing instructions and is waiting for some event to occur. In this state, the processor is not performing any calculations and is simply waiting for a specific event to occur.\n\nFor example, in a computer program, a processor may be in the counting state while it is waiting for a user to input a command. Meanwhile, the processor may be in the halted state while it is waiting for a hardware device to complete a task.\n\nIn general, the counting state is a normal state for a processor, while the halted state is a special state that is used for waiting for events or for performing certain tasks.