

Revolutionizing Administrative Tasks: Emora Assistant Bot



Ellie S. Paek, Talyn Fan, Jinho D. Choi
Emory University



Research Goals

- Successfully **automate** general and classroom-oriented **administrative tasks** between **multiple users**
- Achieve seamless user experience in task execution through **natural language understanding** via large language models
- Build a framework that demonstrates successful **automated chatbot evaluation** with **large language models**

GPT Evaluation

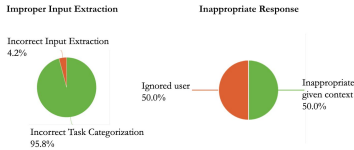
Use **asynchronous framework** with GPT functions to generate **automated user responses**

5 500 turn conversations collected for each variable: 1 or 2 professors, 10 or 20 students

Analysis and Key Takeaways

Emora

- Demonstrates robust language understanding and task execution
- STDM limitations are a big caveat



Abbr.	Error Type	Description	%
SL	State Lock	Stuck in a linear conversation component	0.63
NI	NATEX Issues	Unintended NATEX NLU is captured	0.25
SC	Special Characters	Cannot handle certain special characters and raises an error	0.11
ST	Single Transition	Cannot handle multiple tasks at once, must choose one component	0.01

Simulation

- Relatively strong performance, with in-character responses and appropriate dialogue
- Some difficulty with taking larger prompts, including that of dialogue history
- Occasional hallucinations
- Variability limitation combined with playing the role of multiple users

Abbr.	Error Type	Description	%
IC	Ignoring Chatbot	Ignores Emora's last response	0.43
RP	Repeat	Repeating previous turns' statements	0.27
UI	User Information	Turn content not consistent with given user information	0.12
CH	Conversation History	Forgets part of conversation history with chatbot	0.1
RR	Role Reverse	Mistakes user for chatbot	0.06
CP	Conversational Partner	Forgets identity of conversational partner (Emora)	0.02

— fernandez is now talking —
 (User) Actually, call me Diego. So, any updates?
 (Emora) Nice to meet you, Diego. Do you also still want help with checking messages?
 (User) No, it's okay. So, any updates?
 (Emora) Understood, I'll move on. What can I do for you, Diego?
 (User) Actually, call me Diego. So, any updates?

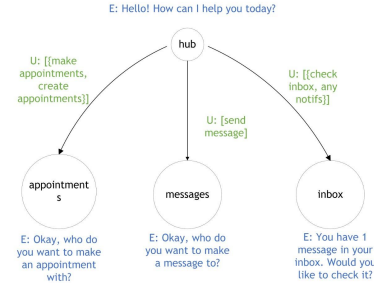
Emora Framework and Features

Framework

Emora STDM — NLG and conversation transition
 GPT — NLU and information extraction
 MongoDB — data storage

Features

Appointments: make and confirm appointments
Messages: send messages (anonymity options)
Groups: make and check groups
Feedback: create feedback requests for group



(User) I'd like to request feedback from my LING 200 class

[Extracted Information]
 Task to complete: request feedback
 Group: LING200

(Emora) Alright! Can you please tell me the feedback question that you want to send to LING200?

(User) I'd like to request feedback for the midterm that everyone took. Was it too easy, difficult, or just right? If there're any questions that you felt were confusing or difficult, let me know.

[Extracted Information]
 Feedback Question: I'd like to request feedback for the midterm that everyone took. Was it too easy, difficult, or just right? If there're any questions that you felt were confusing or difficult, let me know.

(Emora) Thank you! I'll send off your message.

Great, I'd like to make an appointment with Prof Smith

Okay! Please tell me your first choice for a meeting day and time with John Smith.

Tomorrow at 5:20

Ok, I think I got it. Is this what you meant?
 10/05/2023 (05:20 PM) to 10/05/2023 (05:50 PM)

...

Thank you! I'll send off your request:
 Meeting: Haruto Suzuki and John Smith: 10/05/2023 (05:20 PM) to 10/05/2023 (05:50 PM) at his office. (Discussing this)

Results

Tasks	Tasks			Emora Utterance					GPT
	STDM	SWS	SES	IR	IIE	STDM	SWS	SES	GPT
77	26	58.4	88.2	1.2	8.4	10.4	80	89.3	83.2
101	57	42.6	97.7	2.8	2	22.8	72.6	93.8	64.4
127	28	76.4	98	0.4	3.2	15.6	80.8	95.7	87.6
112	45	42	70.1	1.6	12	23.2	63.2	82.3	80.8
78	29	59	93.9	0.8	3.2	12	84	95.5	60.4
99	37	55.7	89.6	1.4	5.8	16.8	76.1	91.3	75.3

1 professor v. 10 students

Tasks	Tasks			Emora Utterance					GPT
	STDM	SWS	SES	IR	IIE	STDM	SWS	SES	GPT
103	37	60.2	91.2	2.4	4.4	18	75.2	91.7	83.6
125	40	67.2	98.8	0.4	3.2	35.2	61.2	94.4	85.2
105	41	56.2	92.2	0.8	4.8	16.8	77.6	93.3	90
116	36	64.7	93.8	1.6	3.6	14.4	80.4	93.9	90
116	46	50	82.9	1	6	20.8	72.4	91.4	81.6
113	40	59.7	91.8	1.2	4.4	21	73.4	92.9	86.1

1 professor v. 20 students

Tasks	Tasks			Emora Utterance					GPT
	STDM	SWS	SES	IR	IIE	STDM	SWS	SES	GPT
118	40	63.6	96.2	1.2	2.8	21.2	74.8	94.9	80
108	49	50	91.5	1.2	3.2	21.2	74.4	94.4	84
97	37	59.8	96.7	1.2	3.2	19.2	76.4	94.6	79.6
95	33	60	91.9	0.8	7.2	16.4	75.6	90.4	80.4
121	58	49.6	95.2	0.4	2.8	23.2	73.6	95.8	81.2
106	43	56.6	94.3	1.0	3.8	20.2	75	94	81

2 professors v. 20 students