

Extrinsic Evaluation

- 1- Ease of use / friendliness: Did you find the system easy to use? Did it take a small effort to complete the objective?
- 2- Understanding/Accuracy/Error Handling: Was the chatbot able to understand what you meant and act accordingly? If it made any mistake, was it able to recover successfully?
- 3- Naturality: Did the conversation feel natural as if you were having a conversation with a human?
- 4- Clarity/Intuitiveness: Was it easy to understand what the bot was asking and what information was trying to obtain from you?

Tasks

1. Order a large margherita and a coke and make a reservation for tomorrow at 8 pm for 1 person. Remember the order number the bot will give you.
2. Cancel the order you just made.
3. Make an order for more than 30€ with at least one lactose-free item (you can ask for options), then choose to do takeaway at 9 pm.

Evaluation

	1	2	3	4	5
Did you find the system easy to use? Did it take a small effort to complete the objective?					
Was the chatbot able to understand what you meant and act accordingly? If it made any mistake, was it able to recover successfully?					
Did the conversation feel natural as if you were having a conversation with a human?					
Was it easy to understand what the bot was asking and what information was trying to obtain from you?					