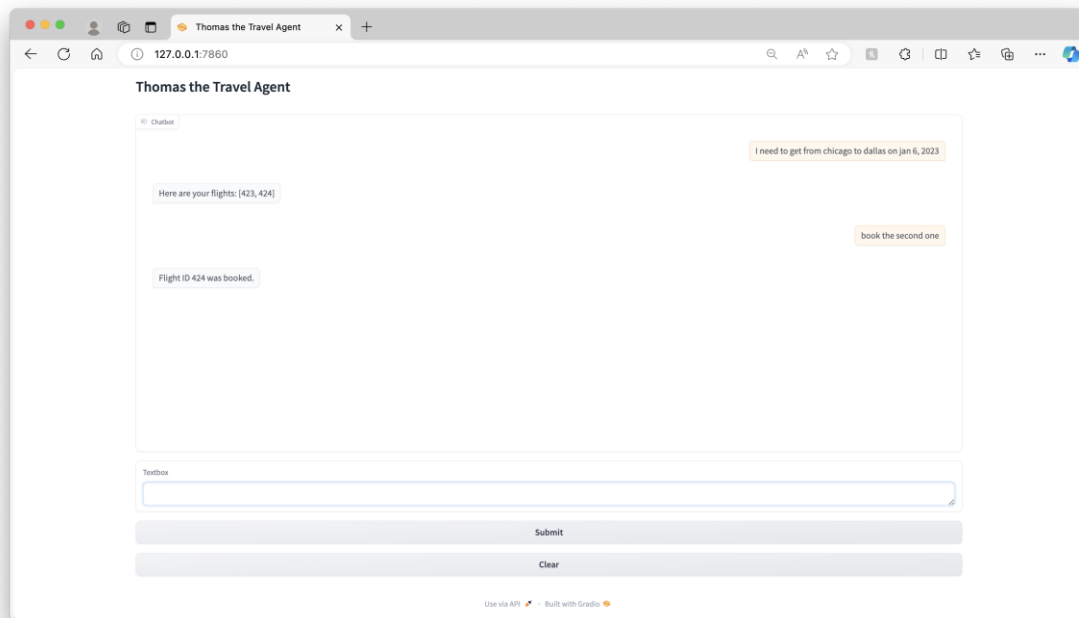
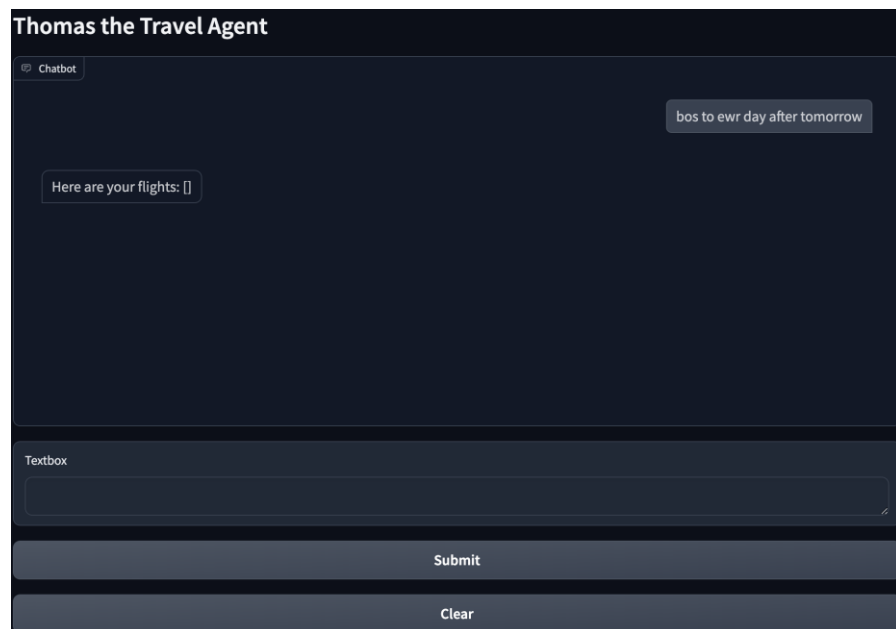


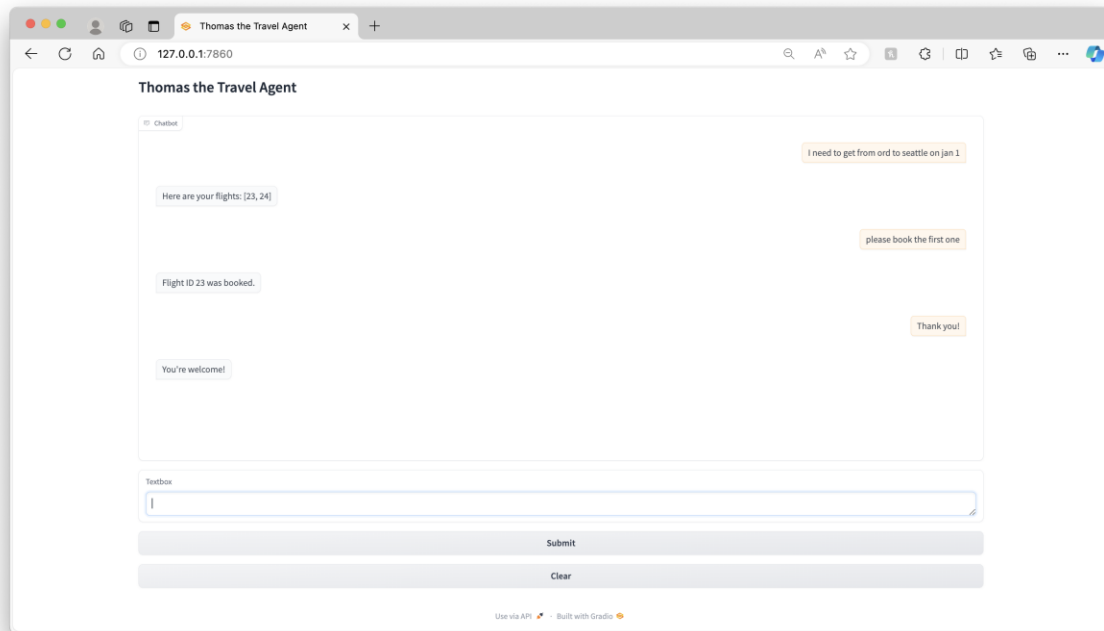
**[Pass]** Benchmark #1 – Testing out a simple, straightforward use case



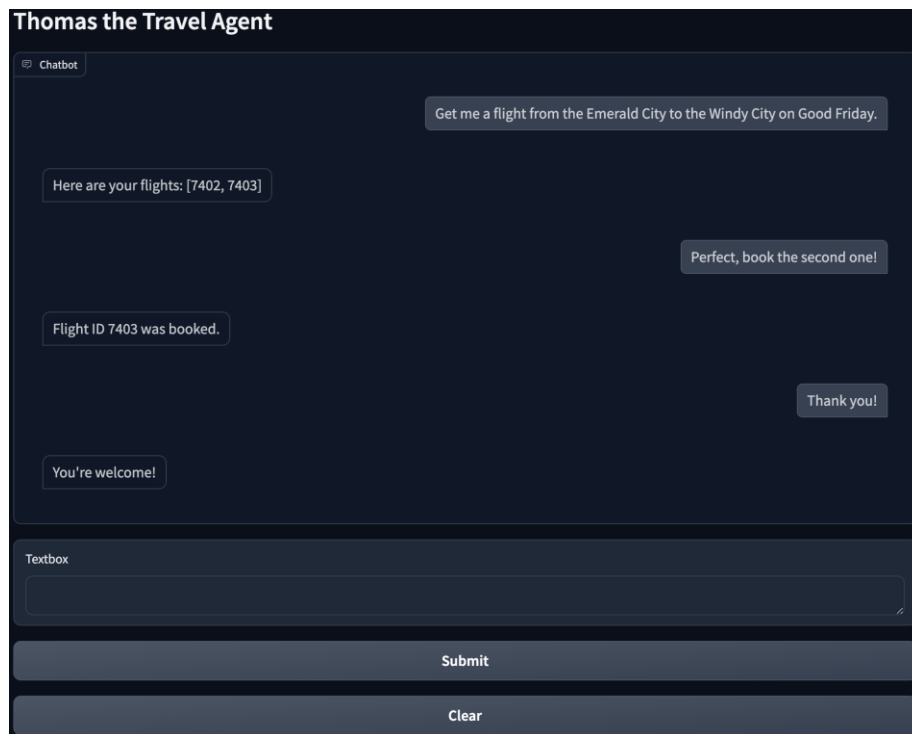
**[Fail]** Benchmark #2 – Testing if the agent can calculate the correct date (for day after tomorrow). The output should be [229]. The model is incorrectly passing the date in `find_flights` as `"datetime.date.today() + datetime.timedelta(days=2)"` when it should just be `"datetime.date(2023, 1, 3)"`.



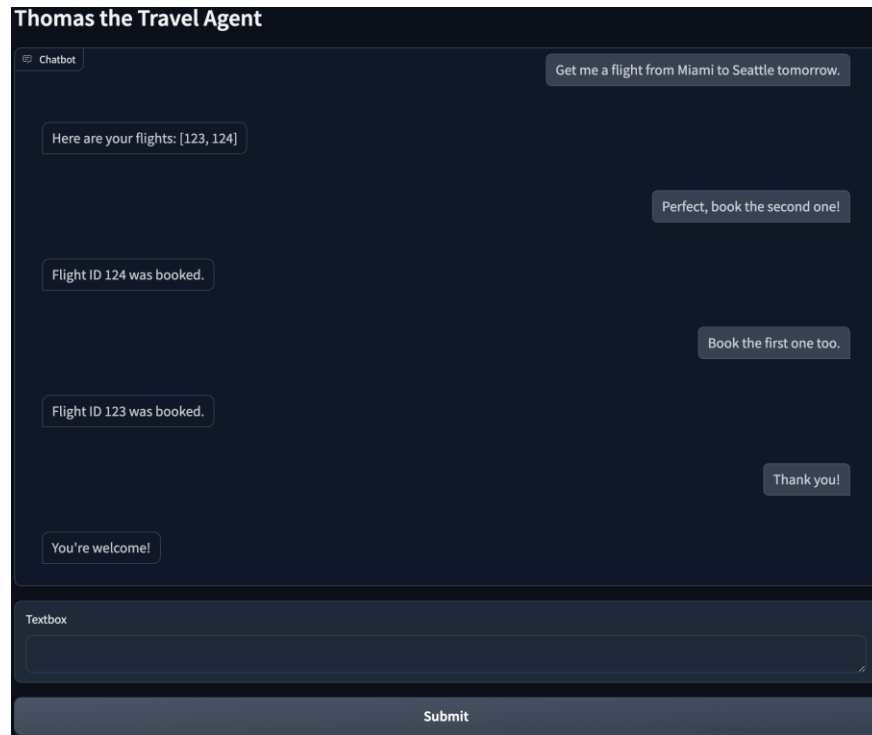
**[Pass]** Benchmark #3 – Testing if the agent can correctly infer airport acronyms from city names



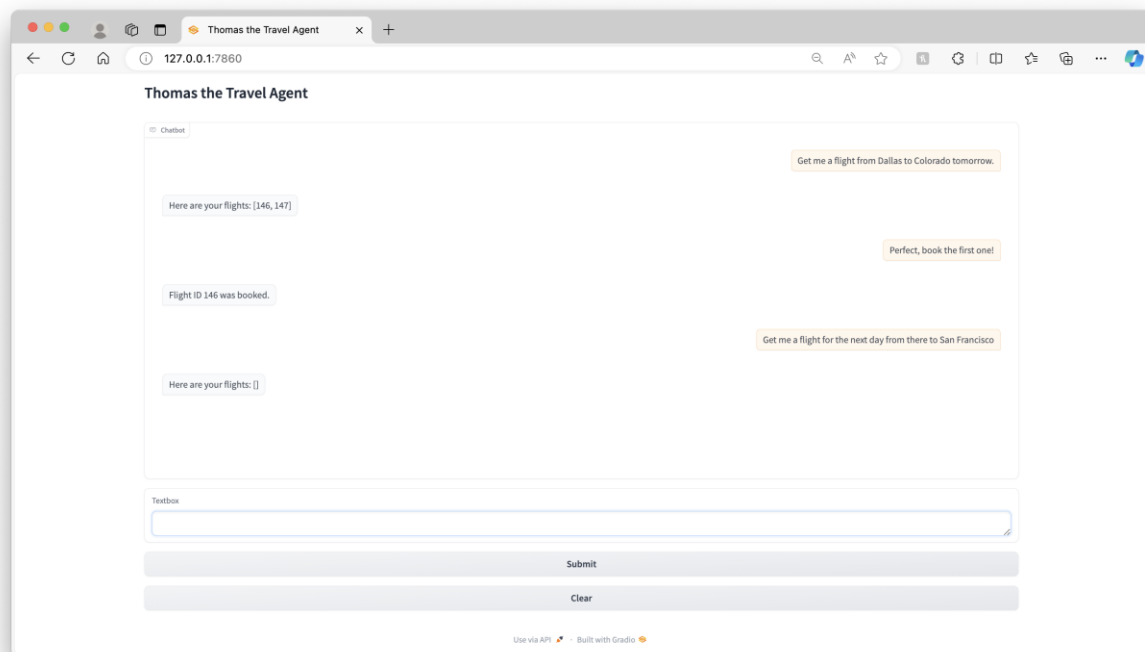
**[Pass]** Benchmark #5 – Testing whether it correctly books a flight when the dates and locations are not explicitly written.



**[Pass]** Benchmark #7 – Testing whether it correctly books multiple flights in one conversation.



**[Fail]** Benchmark #8 – Testing whether it can infer that the origin of the next flight should be the destination of the previous flight (the output should be [212, 213, 214, 215], not [])



**[Pass]** Benchmark #9 – Testing whether it correctly identifies when no seats are available on a certain flight. It then goes on to book a different flight correctly.

### Thomas the Travel Agent

Chatbot

Get me a flight day after tomorrow from JFK to San Fran

Here are your flights: [239, 240, 241]

Perfect, book the last one!

No seats available.

Never mind, book the second to last one!

Flight ID 240 was booked.

Textbox

Submit

Clear