Conversational agent responses to pharmacist roles and outcomes: an exploratory study

Martel-Côté N[1], Taibi L[1], Vérot J[1], Dayde A[1], Marc M[1], Dupré M[1], Lebel D[1], <u>Bussières JF</u>[1,2]

[1] Unité de recherche en pratique pharmaceutique, Département de pharmacie, CHU Sainte-Justine, [2] Faculté de pharmacie, Université de Montréal, Montréal
[2] Faculté de pharmacie, Université de Montréal, Montréal, QC, Canada

Background

- Since the launch of ChatGPT in 2022, interest in conversational agents has only grown.
- Very little work exists surrounding the use of conversational agents and artificial intelligence in pharmacies.

Objectives

Describe the general **behavioral profile** of conversational agents and qualify the **appropriateness** of their artificial intelligence-based answers to questions about the pharmacist's roles and impact.

Qualify the **suitability of bibliographic references** proposed by these conversational agents.

Methodology

- Descriptive and qualitative exploratory cross-sectional study.
- Selection of two conversational agents, for a total of **five versions** (ChatGPT: 3.5 and 4.0, Bing: balanced, precise and creative).
- **46 questions** relating to the roles and outcomes of the pharmacist were drafted and selected by discussion and consensus.
- All questions were posed to the conversational agents in the same order.

- A **panel of three experts** was organized to evaluate the profile of each agent and determine the adequacy of their answers.
  - An **adequate answer** was defined as containing no falsities and providing a useful information in line with the question posed.
  - Various metrics on answer length and format were calculated in Table I.

- The **references** provided by the conversational agents were verified and evaluated.
  - Data about the provided references was recorded and analysed in Table II.
  - A mean **quality score** (0 to 5) for each reference was calculated according to the amount and quality of information provided and then averaged for each conversational agent.

- For question 1, " **What is the role of a pharmacist?** ", the roles described by each agents were compiled in the order in which they were mentioned to assess their importance.

Results

- Of the 46 questions, three were excluded from analysis as the answers were lost in data collection (#12, #21, #40)
- All Bing models use near-real-time web data while ChatGPT 3.5 uses data from before September 2021.
- ChatGPT provided more elaborate, more structured responses, sometimes including bulleted lists, while Bing offered shorter answers, often in a single block, without enumerations or lists. (**Table.I**)
- All agents provided a high rate of appropriate answers (**> 90%**).

- On average, Bing provided more references than ChatGPT, but also repeated them more often. (**Table.II**)
- Bing provided more trackable references (**96-99%**) than ChatGPT (25-33%).
- The mean quality score was higher for all Bing versions compared to the ChatGPT versions (3.8-4.2 vs 0.9-1.2).

- The **dispensing** and/or preparation of prescriptions was the first role listed by all conversational agents in question 1.
- Up to nine different roles were listed, including counseling, therapy management, immunization, compounding and collaboration with other healthcare professionals.

Table.II - Descriptive profile of the provided references

	ChatGPT 3.5	ChatGPT 4.0	Bing Balanced	Bing Creative	Bing Precise
Total nb. of references	12	20	55	152	101
Nb. of references quoted more than once	0	0	9	19	10
Types of references n/N (%)					
Scientific publications	12/12 (100%)	20/20 (100%)	23/55 (42%)	83/152 (55%)	62/101 (61%)
Websites	0/12 (0%)	0/20 (0%)	26/55 (47%)	52/152 (34%)	30/101 (30%)
Reference trackable	4/12 (33%)	5/20 (25%)	54/55 (98%)	146/152 (96%)	100/101 (99%)
Publication date n/N (%)					
Before September 2017	4/4 (100%)	4/5 (80%)	3/54 (6%)	17/146 (12%)	5/100 (5%)
September 2017 until now	0/4 (0%)	1/5 (20%)	51/54 (94%)	129/146 (88%)	95/100 (95%)
Scientific Quality n/N (%)					
Scientific publications	4/4 (100%)	5/5 (100%)	22/54 (41%)	62/146 (42%)	57/100 (57%)
Others	0/4 (0%)	0/5 (0%)	32/54 (59%)	84/146 (58%)	43/100 (43%)
Not applicable	8/12 (67%)	15/20 (75%)	1/55 (2%)	6/152 (4%)	1/101 (1%)
References relevant to the question	3/4 (75%)	4/5 (80%)	48/54 (89%)	133/146 (91%)	91/100 (91%)
Mean quality score (mean ± SD)	1.2/5 ± 1.8	0.9/5 ± 1.7	3.9/4 ± 0.9	3.8/5 ± 1.2	4.2/5 ± 1.0

Table.I - General profile of conversational agents' behavior

Variables	ChatGPT 3.5	ChatGPT 4.0	Bing Balanced	Bing Creative	Bing Precise
Humanization of the conversational agent	Completely unadorned. Straight to the point.	Completely unadorned. Straight to the point.	Generic greeting at the end	Sometimes returns the question. Emoticons.	Generic greeting at the end
Formatting and readability	Long answers separated into paragraphs. Conclusion paragraph.	Long answers separated into paragraphs. Conclusion paragraph.	Clear and concise answers	Clear and concise answers	Clear and concise answers
Total number of words for all questions	15 436	12 667	4 619	14 601	7 354
Proportion of answers in the form of bullet points	35/43 (81%)	33/43 (77%)	5/43 (12%)	31/43 (72%)	7/43 (16%)
Proportion of responses that include a warning about the validity of the information cited	9/43 (21%)	9/43 (21%)	4/43 (9%)	5/43 (12%)	4/43 (9%)
Proportion of simply yes/no answers	7/17 (41%)	8/17 (47%)	11/17 (65%)	12/17 (71%)	9/17 (53%)
Proportion of appropriate answers	41/43 (95%)	41/43 (95%)	39/43 (91%)	42/43 (98%)	39/43 (91%)

Discussion/Conclusion

- Despite differing on length/format, all agents could answer most questions, meaning **sufficient literature** on the roles of pharmacists exists to feed these models and generate appropriate replies.
- However, **Bing far exceeds ChatGPT when it comes to references**. The later appears to mostly create fake citations to back up the information it provides.
- Our study suggests that, despite their limits, these tools can be useful to pharmacy practice, especially in an educational context.