

# Assessing the accuracy of data extracted by ChatGPT 4.0 compared to Impact Pharmacie entries



Juliette MUTIN<sup>1,2</sup>; Marie-Lili Hafizin<sup>1</sup>; Victor El-Jammal<sup>1,2</sup>; Célia Morel<sup>1</sup>; Lisa Leca<sup>1</sup>; Maxime Bergeron<sup>3</sup>; Jean-François Bussières<sup>1,4</sup>

<sup>1</sup> Unité de recherche en pratique pharmaceutique, CHU Sainte-Justine, Montréal, Québec, Canada

<sup>2</sup> Faculté de Pharmacie de Lyon, Univ Lyon, Université Claude Bernard Lyon 1, ISPB, Lyon, France

<sup>3</sup> CISSS des Laurentides, Hôpital de Saint-Jérôme, Saint-Jérôme, Québec, Canada

<sup>4</sup> Faculté de pharmacie, Université de Montréal, Montréal, Québec, Canada

## Background

- Generative Artificial Intelligence (GenAI), particularly conversational agents like ChatGPT, can be used to automate data extraction.
- However, we don't know if these tools can extract scientific data with accuracy comparable to humans, especially in the field of pharmacy.
- Since 2010, the Impact Pharmacie platform has been updated annually by the URPP team at CHU Sainte-Justine.
- Articles are reviewed and data are entered into the platform by pharmacy students.
- Each article requires an average of 1 to 4 hours of work.
- We aimed to explore the use of ChatGPT to support the updating of Impact Pharmacie.

## Objective

To assess the accuracy of data extracted by ChatGPT 4.0 compared to Impact Pharmacie entries.

## Method

- Design:** Descriptive comparative study.
- Sample:** A random selection of 100 articles entered on Impact Pharmacie (<http://Impactpharmacie.org>) by pharmacy students was identified.
- Intervention:**
- ChatGPT 4.0 (paid version) was programmed (in French) with a series of instructions and queried with 18 questions per article to extract targeted data.
  - Five pharmacy students each uploaded 25 articles, one by one, along with the question set.
  - ChatGPT responses were extracted into a Word file and compared to the data entered in Impact Pharmacie platform.
- Evaluation criteria:**
- Concordance was assessed across 23 criteria (e.g., study type description, pharmaceutical intervention description, limitations).
- Analysis plan:**
- Comparison of extracted data with Impact Pharmacie entries
  - Agreement assessed at three levels: concordant, partially concordant, non-concordant
  - Identification of missing data
  - Identification of AI hallucinations
  - Measurement of ChatGPT response generation time

## Results

ChatGPT concordance compared to Impact Pharmacie entries by pharmacy students:

- Criteria: 77% ± 19% on average
- Outcomes: 449 identified by students, 311 by ChatGPT
- Figure 1 shows the concordance profile for study outcomes analyzed

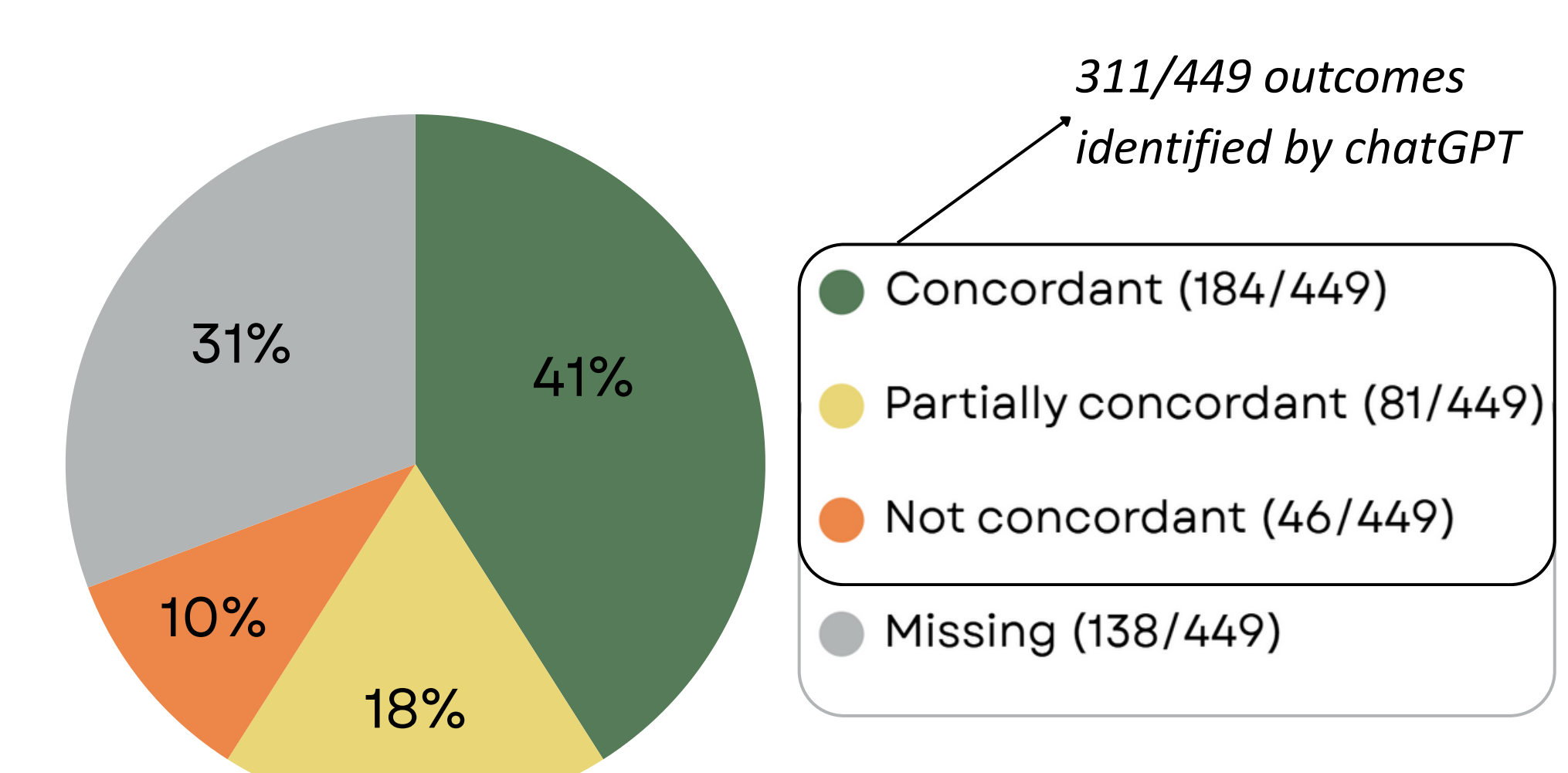


Figure 1

- Hallucinations:
  - 29 hallucinations detected across 23 of the 100 articles (study limitations, duration, stakeholders)
- Average extraction time by ChatGPT:
  - 1 minute 45 seconds ± 28 seconds per article
- Overall inter-rater agreement (Kappa):
  - 0.816, p < 0.001
  - Lower agreement for some criteria, such as intervention description (Kappa = 0.594)

**Table I:** Concordance profile of data extracted by ChatGPT compared to student-extracted data, expressed as a percentage and presented in descending order

Evaluation criteria	Total concordance	Partial + total concordance
Country	99%	99%
Year of publication	99%	99%
Presence or absence of blinding	96%	96%
Number of study groups	95%	96%
Type of healthcare setting where the study takes place	94%	94%
Participant distribution	93%	93%
Study design (retrospective, prospective, or cross-sectional)	89%	89%
Identification of the primary objective	89%	97%
Study duration	85%	89%
Location of the pharmaceutical intervention	84%	96%
Presence of a control group	83%	83%
Study category	80%	81%
Exclusion criteria	80%	93%
Description of outcomes *	77%	88%
Stakeholders involved in the intervention (pharmacists and other healthcare professionals)	73%	94%
Statistical significance of the parameters *	71%	87%
Duration of the pharmaceutical intervention within the study	68%	70%
Results of the parameters*	66%	84%
Description of the pharmaceutical intervention (pharmacist's role in the study)	66%	98%
Inclusion criteria	61%	96%
Classification of parameters as primary or secondary	54%	54%
Study limitations	35%	84%
Identification of secondary objective(s)	32%	37%
<b>Mean ± standard deviation</b>	<b>77 ± 19 %</b>	<b>86 ± 15 %</b>

\* The concordance of criteria marked with an asterisk is assessed only on outcomes that were not missing from ChatGPT's extraction.

## Discussion

- Positive points:**
- ChatGPT can extract data from pharmacy scientific articles
    - High total concordance (>90%) for criteria with uniform vocabulary (e.g., country of origin, publication year).
    - Including partial concordance in the analysis significantly improves results for some criteria (e.g., study limitations, outcomes).
  - ChatGPT extraction time is much shorter compared to a student.
  - High overall inter-rater agreement (Kappa = 0.816, p < 0.001)

- Points for discussion:**
- Human validation by a trained student is still required before integration into Impact Pharmacie.
  - Lower concordance for criteria involving diverse vocabulary (e.g., study type, intervention description).
  - Hallucinations detected in nearly 25% of articles, mostly related to the study limitations.
  - PDF format and the presence of tables/graphs are challenging for ChatGPT and may explain missing outcomes.
  - Asking questions in French about English-language articles may hinder ChatGPT's comprehension and accuracy.

## Conclusion

- This study confirmed the feasibility of data extraction from pharmaceutical scientific articles using ChatGPT 4.0.
- This approach could accelerate article review and data extraction for updating the Impact Pharmacie platform.
- **HOWEVER**, human validation remains necessary to ensure the accuracy and reliability of the extracted data.

## Future perspectives:

- Improve instructions (prompt engineering)
- Test other conversational agents
- Ask questions in the same language as the articles



**Contact:** jean-francois.bussieres.hsj@sss.gouv.qc.ca  
**Conflict of interest:** None  
**Funding:** None  
 Poster presented at the 2025 PPC, June 6-8th, Ottawa, Ontario, Canada (Virtual)