EVALUATING THE PERFORMANCE OF TWO OPENIA MODELS FOR CLINICAL DECISION-MAKING IN HOSPITAL PHARMACY

Authors: F. Lago Ballester[1], A. Lázaro Cebas[1], Á. Varea Béjar[1], M.E. Martínez Madrid[1], E.J. Sevilla Alarcón[1], J.M. Moya Flores[1].
[1]Hospital General Universitario Santa Lucía, Hospital Pharmacy, Cartagena, Spain.

Background and Importance

Artificial intelligence (AI) is increasingly used to support clinical decision-making in hospital pharmacy. However, different AI models may affect reliability and safety in high-risk settings. This study compares two OpenAI models—ChatGPT-Instant (GPTI) and ChatGPT-Thinking (GPTT)—in solving clinical queries in hospital pharmacy.

Aim and Objectives

To compare two AI models, GPTI and GPTT, evaluating accuracy and response time (RT) to determine their suitability for pharmaceutical practice.
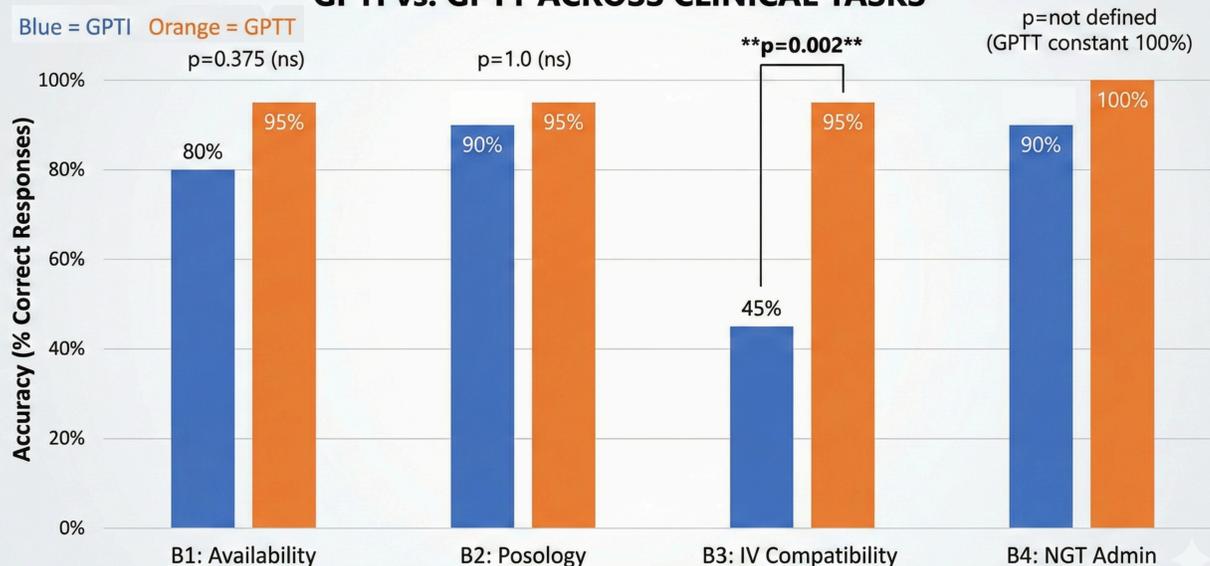
Materials and Methods

METHODOLOGY FLOWCHART: EXPERIMENTAL DESIGN & EVALUATION PROCESS
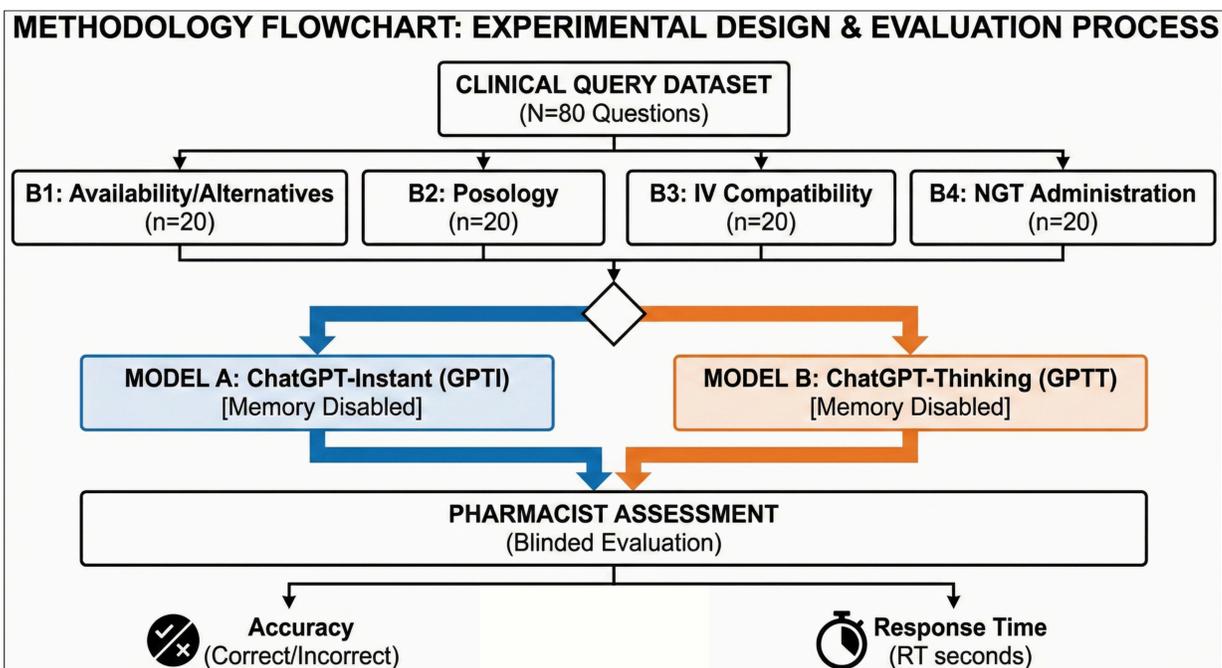
A standardized prompt was used to present identical queries to both models, with the memory function disabled to ensure independent responses.
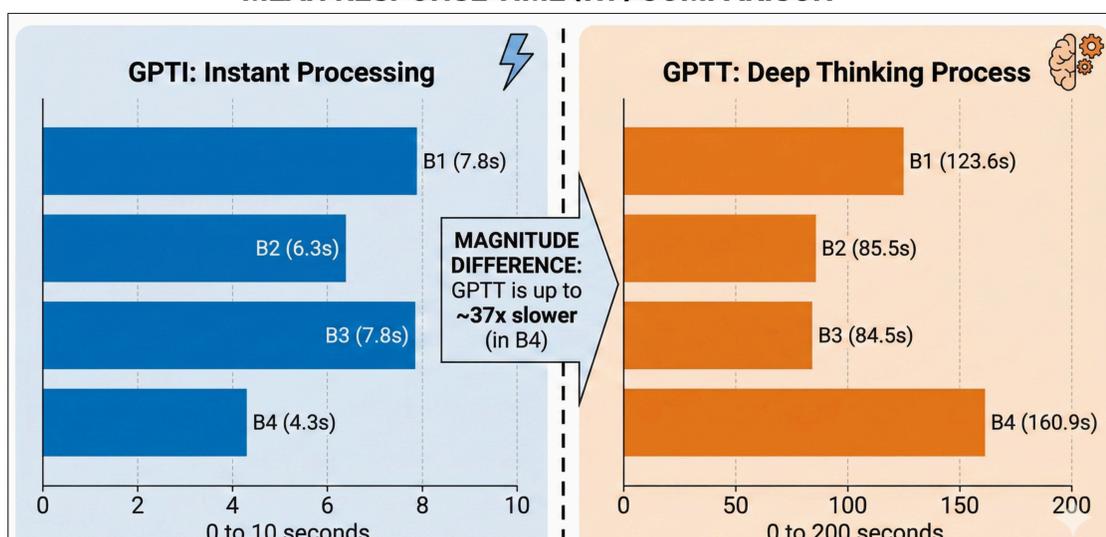
GPTI vs GPTT

Agreement in correct/incorrect responses between models was analysed using McNemar's test per block.

Results

ACCURACY & STATISTICAL SIGNIFICANCE: GPTI vs. GPTT ACROSS CLINICAL TASKS

Blue = GPTI Orange = GPTT

p=0.375 (ns) p=1.0 (ns) **p=0.002** p=not defined (GPTT constant 100%)

- B1: Availability — GPTI 80%, GPTT 95%
- B2: Posology — GPTI 90%, GPTT 95%
- B3: IV Compatibility — GPTI 45%, GPTT 95%
- B4: NGT Admin — GPTI 90%, GPTT 100%

MEAN RESPONSE TIME (RT) COMPARISON

GPTI: Instant Processing
- B1 (7.8s)
- B2 (6.3s)
- B3 (7.8s)
- B4 (4.3s)
0 to 10 seconds

MAGNITUDE DIFFERENCE: GPTT is up to ~37x slower (in B4)

GPTT: Deep Thinking Process
- B1 (123.6s)
- B2 (85.5s)
- B3 (84.5s)
- B4 (160.9s)
0 to 200 seconds

Conclusion and relevance

GPTT demonstrated superior reliability across all evaluated domains. The performance gap was most critical in IV compatibility, where GPTT achieved statistically significant superiority (p = 0.002).

GPTI showed comparable accuracy to GPTT in Posology and NGT administration, coupled with lower RT. GPTI was faster across all areas.

Currently, GPTT's latency is a necessary trade-off for clinical safety.

Scan for Abstract & Code

Contact data: fernando.lago@carm.es
Abstract Number: 5PSQ-023