Can a Model Explainer Agent Accelerate Manufacturer Responses to Health Technology Assessment Clarification Questions?

MSR297

Tushar Srivastava, Hanan Irfan, Thaison Tong, Shilpi Swami

ConnectHEOR, London, UK

Email: tushar.srivastava@connectheor.com

connectHEOR

> **On one rare disease Excel model and 30 simulated NICE and CDA-AMC clarification questions, an explainer agent located the correct worksheet and formula for all 30 on first pass and cut first-draft time from about 2 hours to 8 minutes per question, with no fabricated references; a senior modeller adjudicated every output.**

1 BACKGROUND

After a model submission, clarification questions can ask the manufacturer to explain any formula in a model that may hold well over 100,000 cells and to trace every parameter to its source; at the National Institute for Health and Care Excellence (NICE), the company has 14 calendar days to reply.[1]

- Errors are common: for instance, only two of the 41 appraisals NICE completed in 2017 were free of reported errors.[4]
- Credibility rests on **traceability** [2,3], and Canada's Drug Agency (CDA-AMC) discourages IFERROR statements because they mask errors.[5]

GAP: Generative-AI has been proposed for HTA work, including modelling [6], but evidence on explaining a submitted model back to its reviewer is lacking.

2 OBJECTIVES

We evaluated a **model explainer agent**, a conversational AI agent that explains an economic model in natural language, on drafting manufacturer responses to HTA clarification questions.

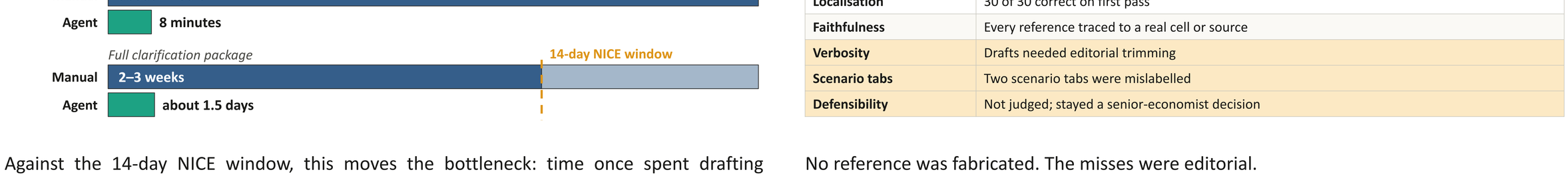
1. Localisation
Measure whether the agent correctly localises the model component each question targets.

2. Speed
Measure time to a first draft and to a complete package against a manual baseline.

3. Faithfulness
Characterise faithfulness and failure modes, and whether speed costs accuracy or ownership.

3 METHODS

The case-study asset was a rare disease Excel cost-effectiveness model: a **Markov cohort model** with 5 health states, quarterly cycles, a lifetime horizon and a payer perspective (36 worksheets, approximately 150,000 cells). The workbook was parsed into a structured, queryable representation, and its documentation was indexed via retrieval-augmented generation (RAG), so exact cell references and prose rationale could both be retrieved.



- **Panel:** 30 technical clarification questions, simulated from NICE and CDA-AMC submissions, spanned structure, parameters, costs, utilities, discounting, extrapolation and scenario requests.
- **Task:** for each question, the agent located the worksheet, cell range and driving formula, explained the rationale and drafted a response.
- **Judging:** a senior modeler compared outputs with a manual baseline for localisation, traceability of every reference, and drafting time.
- **Ownership:** the modeler could query the agent further, reject a draft and make the final call.

4 RESULTS

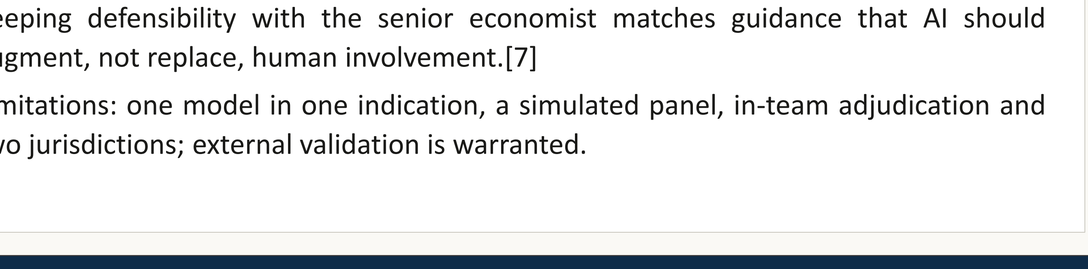
The agent correctly localised the relevant component for **30 of 30 questions (100%)** on first pass. This held even for VBA-automated probabilistic analysis, where the calculation path is partly held in code rather than in visible cell dependencies.

30 / 30 questions localised correctly on first pass

8 min median time to a first draft, down from about 2 hours

~1.5 days for the full package, down from 2–3 weeks

Drafting time: manual baseline vs agent



Adjudicated outcomes and failure modes

Criterion	Finding
Localisation	30 of 30 correct on first pass
Faithfulness	Every reference traced to a real cell or source
Verbosity	Drafts needed editorial trimming
Scenario tabs	Two scenario tabs were mislabelled
Defensibility	Not judged; stayed a senior-economist decision

Against the 14-day NICE window, this moves the bottleneck: time once spent drafting becomes available for senior review and methodological argument.

No reference was fabricated. The misses were editorial.

5 DISCUSSION

Faithfulness follows from the architecture, not the language model. An agent that retrieves from a parsed workbook can only cite cells and passages that exist, whereas unaided language models are known to fabricate references.[8]

- A clarification response has a mechanical part, explaining what the model does, and a judgement part, defending whether it should.
- The agent compresses the mechanical part, which dominates the calendar, while judgement still decides the outcome.
- Keeping defensibility with the senior economist matches guidance that AI should augment, not replace, human involvement.[7]
- Limitations: one model in one indication, a simulated panel, in-team adjudication and two jurisdictions; external validation is warranted.

6 CONCLUSION

A governed, **human-in-the-loop** agent accelerated clarification drafting for 30 simulated questions across **two HTA jurisdictions** while preserving traceability and expert ownership.

- **Speed:** first-draft time fell from about 2 hours to 8 minutes per question, and the full package took about 1.5 days against an estimated [baseline to be defined: TS]; NICE allows 14 calendar days.
- **Accuracy:** 30 of 30 questions localised; every reference traced to a real source.
- **Ownership:** methodological judgement stayed with the senior economist.
- **Next step:** external validation on more models, indications and adjudicators.

Implication: the agent compresses mechanical work, not judgement; under senior accountability, it is a defensible addition to the submission workflow.

REFERENCES

1. NICE. Technology appraisal manual (PMG36), section 5.
2. Eddy DM, et al. Value Health. 2012;15.
3. Büyükkaramikli NC, et al. PharmacoEconomics. 2019;37.
4. Radeva D, et al. Int J Technol Assess Health Care. 2020;36.
5. CDA-AMC. Pharmaceutical Reviews Update, Issue 52. 2024.
6. Fleurence RL, et al. Value Health. 2025;28.
7. NICE. Use of AI in evidence generation (ECD11). 2024.
8. Chelli M, et al. J Med Internet Res. 2024;26.

Financial disclosure: The authors are employees of ConnectHEOR Limited. No external funding was received. The authors have no conflict of interest to declare.

Poster presented at ISPOR Europe 2026 in Vienna, Austria, 8-11 November 2026