Why AI Proof-of-Concepts Do Not Necessarily Prove Anything: Journey from PoC to Product

MSR43

Hanan Irfan, MSc[1], **Tushar Srivastava, MSc**[2]
[1]ConnectHEOR, Delhi, India; [2]ConnectHEOR, London, UK | Email: tushar.srivastava@connectheor.com

KEY MESSAGE

In HEOR, AI proof-of-concepts demonstrate possibility, not readiness. The danger in HEOR AI isn't a failed PoC - it's a successful one that gets deployed before it's ready.

BACKGROUND

- Artificial Intelligence (AI) is increasingly adopted across Health Economics and Outcomes Research (HEOR) to support evidence synthesis, literature review, real-world evidence generation, economic modelling, and technical reporting.
- Proof-of-concepts (PoCs) have become a common way to demonstrate innovation and , attract internal or external buy-in, and showcase the potential utility of AI-enabled tools.
- However, a technically successful PoC does not demonstrate that the solution is **reliable, reproducible, scalable, or suitable** for use in high-stakes HEOR and Health Technology Assessment (HTA) workflows.[1][2]

The gap between "it works in a demo" and "it works in production" is where most AI-enabled HEOR tools fail.

OBJECTIVES

1. Examine limitations of AI PoCs in HEOR and why they do not demonstrate real-world suitability
2. Identify methodological, governance, and operational gaps that must be addressed before such systems can be considered product-ready
3. Synthesize these observations into a practical maturity framework to support the transition from PoC to robust HEOR product development.

METHODS

Study Design

- Qualitative cross-initiative analysis of AI-enabled HEOR use cases
- Included literature review, economic modelling, real-world evidence generation, and technical writing workflows
- Sample comprised both early-stage PoCs and mature, production-oriented systems
- Enabled comparison between feasibility-focused and operational implementations

Assessment Framework

Each PoC was evaluated against the requirements of a production-ready HEOR product across five dimensions(see Results).

Synthesis

- Failure modes grouped into higher-order categories
- Synthesized into a PoC-to-product maturity framework tailored to HEOR/HTA
- Framework structured as a lifecycle model reflecting evidentiary and governance requirements

RESULTS

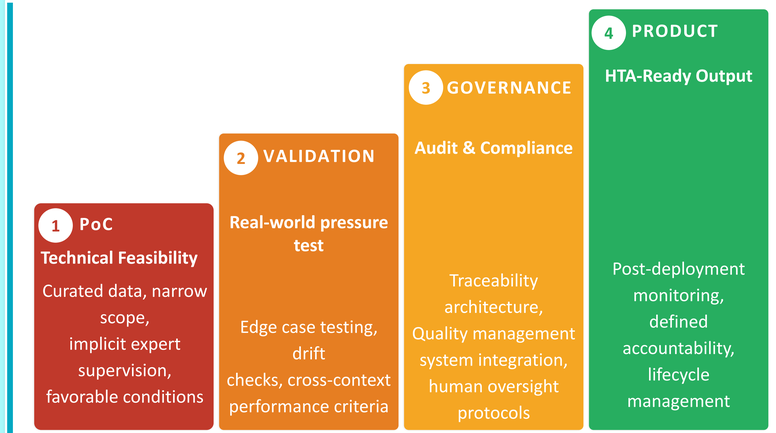
Five Assessment Dimensions: PoC Reality vs Product Requirement

DIMENSION	PoC REALITY	PRODUCT NEED
Test Data Representativeness	Curated datasets, cleaner than real project data	Diverse, messy, variable real-world HEOR inputs
Robustness to Edge Cases & Data Drift	Untested beyond narrow task definitions	Validated across disease areas, designs, endpoints
Transparency & Auditability	Plausible outputs without traceability architecture	Full source-to-output audit trail
Human-in-the-loop (HITL) Integration	No focus on HITL during system design	Embedded in review, validation workflows
Accountability Post-Deployment	Unclear ownership, implicit expert oversight	Defined roles for errors, escalation, monitoring

Five Failure Modes Observed During Scale-Up

Loss of Traceability	CRITICAL	Plausible outputs but lacked documentation to explain data sources, reasoning steps, and reproducibility. In HEOR/HTA, absence of traceability limits auditability and makes outputs insufficient despite apparent validity.
Unclear Accountability	CRITICAL	Poorly defined ownership for validation, escalation, error correction, model updates, and approval authority. This creates risk in HEOR contexts where outputs influence strategic and regulatory decisions.
Inconsistent Cross-Context Performance	HIGH	Systems performed well in initial settings but degraded across different therapeutic areas, evidence types, and use cases. This reflects optimisation for specific PoC conditions rather than generalisability across the HEOR domain.
Sensitivity to Evolving Evidence	HIGH	PoCs lacked mechanisms for feedback, revalidation, and monitoring as evidence and inputs changed. As a result, their validity eroded over time in dynamic HEOR tasks.
Dependence on Human Oversight	HIGH	PoC performance relied on expert users correcting and refining outputs in real time, masking system limitations. This dependence limited scalability and created risks in broader deployment without similar oversight.

PoC-TO-PRODUCT MATURITY FRAMEWORK



Successful transition requires substantial redesign beyond the PoC: explicit validation thresholds, reproducibility controls, governance checkpoints, and post-deployment monitoring aligned with HTA expectations.

The PoC Illusion

"Can the AI produce an impressive output on a narrow input once?"

Curated inputs → Expert oversight → Favourable judgment

The Product Question

"Can it be scaled, validated, monitored, and trusted repeatedly?"

Governance → Auditability → Reproducibility → Impact

CONCLUSIONS

- In HEOR, AI PoCs demonstrate **vague possibility, not readiness.** Treating PoCs as evidence of validity risks premature promise of tools that fail under real-world tasks, HTA scrutiny, cross-context variability, evolving evidence, and formal quality requirements.
- The relevant question is not whether a prototype can produce an impressive output once, but whether the system can be **scaled, validated, monitored, and trusted repeatedly** where analytical credibility matters.

AI development in HEOR must be approached as a disciplined lifecycle , not a one-time proof of technical feasibility.

References :
1.Lekadir, et al; 2025; BMJ Information Fusion
2.Sanchez, et al, 2026;
3. Srivastava T, et al, Integration of generative AI with human expertise : a hybrid intelligence framework. *Adv Ther.*

Financial Disclosure: Authors are employees of ConnectHEOR Limited. No external funding received. No conflict of interest to declare.

Poster presented at ISPOR in Pennsylvania, Philadelphia, USA, May 17-20, 2026