MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

AUSTRALIA: WHAT HAPPENS IF NOTHING CHANGES? QUESTIONS FOR THE FUTURE OF HEALTH TECHNOLOGY ASSESSMENT

A discussion prompted by international AI large language model interrogations of HTA knowledge bases

Paul C Langley Ph.D Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 665 MAY 2026

www.maimonresearch.com

Tucson AZ

INTRODUCTION

"The issue may no longer be whether measurement inversion exists, but whether institutions should prepare for its consequences."

Over the past six months, Maimon Research has undertaken more than 230 AI large language model interrogations of HTA knowledge bases associated with agencies, academic centers, journals, pharmacy schools, and formulary guideline structures across 28 countries. The findings have been remarkably consistent. Responses to a canonical diagnostic framework repeatedly suggest weak recognition of representational measurement principles together with strong endorsement of assumptions supporting utilities, QALYs, simulation frameworks, and cost-effectiveness claims.

These findings have been communicated through country assessments, institutional interrogations, correspondence with agencies and academic centers, and supporting transition proposals. While discussion has begun, responses have generally been cautious and measured. This is perhaps unsurprising. Questions concerning the scientific foundations of an established discipline rarely generate immediate institutional reaction, particularly where current frameworks remain deeply embedded within teaching programs, publication practices, reimbursement systems, and research infrastructures.

Yet muted response should not be confused with absence of significance. Scientific questions often require time before their broader implications become apparent. If the interrogation findings continue to attract attention internationally, then the issue may no longer concern particular agencies or institutions. It may increasingly concern the future intellectual foundations of HTA itself.

The purpose of this paper is therefore not to revisit technical arguments regarding representational measurement. Rather, it is to explore several broader questions that may arise if the implications of these interrogations continue expanding internationally. What happens if nothing changes? Could current concerns prove temporary? What are the implications for graduate education, research centers, and institutional preparedness? And perhaps most importantly: if methodological transition eventually becomes necessary, should institutions begin preparing now?

QUESTION 1: "IF CURRENT HTA METHODS HAVE BEEN ACCEPTED FOR FORTY YEARS, WHY SHOULD ANYTHING CHANGE NOW?"

This is perhaps the most common reaction to criticisms of current HTA practice. The underlying assumption is straightforward: if a methodology has been used internationally for four decades, supported by governments, universities, journals, reimbursement agencies, and thousands of publications, then surely it must possess scientific legitimacy. Institutional persistence becomes confused with scientific validity.

The difficulty is that scientific history provides many examples where widespread acceptance and methodological validity diverged. Entire fields can stabilize around conventions and assumptions that later prove difficult or impossible to defend. Longevity demonstrates institutional success; it

does not establish scientific truth. Scientific disciplines evolve not because practices are old, but because they continue to satisfy the standards required for credible claims.

The central issue raised by recent HTA interrogations is not whether current approaches have generated useful information. Clearly, they have. Economic evaluations, epidemiological studies, healthcare utilization analyses, and policy assessments have contributed substantially to healthcare decision-making. The concern lies elsewhere: whether many quantitative claims in HTA satisfy the requirements necessary for lawful measurement.

The AI large language model interrogations undertaken across more than 230 institutions and knowledge environments in 28 countries point to a remarkably consistent pattern. The axioms of representational measurement appear weakly represented or absent within contemporary HTA frameworks. Instead, arithmetic operations are often undertaken independently of demonstrated scale properties. Utilities are multiplied by time; preference scores are aggregated; simulation models generate claims extending decades into the future. Yet there is often little discussion of dimensional homogeneity, admissible transformations, or whether latent constructs possess measurement properties sufficient to support arithmetic manipulation.

This does not necessarily imply fraud, incompetence, or bad faith. Disciplines frequently develop around inherited assumptions that become normalized through teaching, publication practices, reimbursement frameworks, and professional incentives. Once embedded, methodological conventions can persist largely unchallenged.

The issue therefore is not: “Why change after forty years?” The question is instead: “What happens if foundational assumptions become increasingly difficult to defend?” Once representational measurement enters the discussion, institutional age no longer protects a methodology. Four decades of acceptance do not exempt HTA from the standards expected of scientific inquiry.

The implication is not sudden collapse. Existing systems will continue for many years. The issue is preparedness. Institutions may increasingly need to ask whether early consideration of transition pathways is preferable to later reactive adjustment. History suggests that scientific disciplines rarely avoid change indefinitely. The important question is whether HTA chooses to lead that process or simply respond when external pressure eventually makes change unavoidable.

QUESTION 2: “WHAT HAPPENS IF RESEARCH CENTERS SIMPLY CONTINUE WITH BUSINESS AS USUAL?”

At first sight, the answer may appear obvious: perhaps very little happens. Research centers continue to publish papers, attract grants, supervise doctoral students, participate in policy discussions, and contribute to HTA debates. Faculty continue teaching established methods, students continue learning familiar analytical frameworks, and agencies continue relying upon cost-effectiveness models, utilities, and QALY-based evaluations. From this perspective, the system appears stable and self-sustaining.

Indeed, in the short term this may be entirely correct. Academic institutions and research Centers are not highly responsive to methodological criticism, particularly when established practices

remain supported by journals, funding agencies, reimbursement systems, and professional networks. Existing frameworks possess substantial institutional inertia. Entire careers, teaching programs, and research infrastructures are often built around accepted conventions.

The issue, however, concerns the longer term rather than immediate disruption.

The recent AI large language model interrogations across more than 230 institutions and knowledge environments in 28 countries suggest that a recurrent pattern now exists internationally. Across agencies, journals, pharmacy schools, research centers, and academic programs, the same issue repeatedly emerges: weak recognition of representational measurement requirements together with strong endorsement of frameworks dependent upon arithmetic manipulation of constructs whose measurement properties remain unresolved.

If these findings continue to circulate and attract attention, the consequences may become cumulative rather than immediate. Questions begin to emerge regarding the scientific foundations of current HTA methods. Why are utilities treated as though they support multiplication? Why are latent constructs approached through ordinal summation rather than invariant measurement? Why do simulation claims extending decades into the future escape meaningful empirical evaluation? Why does measurement appear to follow arithmetic rather than precede it?

Research centers may initially ignore such questions. The difficulty is that unresolved criticism rarely disappears automatically. It accumulates. Graduate students begin asking questions. Researchers outside HTA notice anomalies. Adjacent disciplines become curious. Eventually, what initially appeared peripheral begins influencing perceptions of scientific credibility.

The risk is therefore not sudden institutional collapse. Grants will continue; journals will continue; conferences will continue. The larger concern is gradual erosion of confidence in the quantitative foundations supporting the discipline. If measurement inversion becomes increasingly visible, centers may find themselves defending inherited assumptions rather than shaping future directions.

The future of HTA may therefore not depend simply upon preserving existing structures. It may increasingly depend upon whether institutions are willing to consider alternatives before external pressures force change. Continuing with business as usual may prove entirely successful. It may also mean surrendering intellectual leadership at precisely the moment when new frameworks begin to emerge.

QUESTION 3: “COULD THIS SIMPLY PROVE TO BE ANOTHER PASSING METHODOLOGICAL DEBATE?”

This is a reasonable question. Health technology assessment has experienced no shortage of methodological disputes over the past four decades. Debates concerning discounting rates, willingness-to-pay thresholds, probabilistic sensitivity analysis, Bayesian versus frequentist approaches, indirect treatment comparisons, utility instruments, and model structures have all generated extensive discussion. Most eventually settled into revised conventions without fundamentally changing the structure of HTA itself.

Viewed in this context, criticisms of representational measurement may initially appear to be simply another technical disagreement. Academic disciplines regularly generate controversies that attract temporary attention before fading from view. The natural response is therefore caution: perhaps the present concerns regarding utilities, QALYs, dimensional homogeneity, and latent construct measurement will ultimately prove no different.

There is, however, an important distinction.

Most previous HTA debates occurred within an accepted methodological framework. Arguments focused on how best to perform cost-effectiveness analysis, how to improve model assumptions, or how to refine statistical methods. The current discussion differs because it raises questions regarding the foundational conditions required before quantitative claims can be made at all. The issue is not whether a simulation model should use one parameter estimate rather than another. The issue is whether the constructs entering the model possess admissible measurement properties in the first place.

The recent AI large language model interrogations provide additional reason for caution. Across more than 230 interrogations in 28 countries, remarkably similar findings have emerged. HTA agencies, academic centers, journals, pharmacy schools, and reimbursement systems repeatedly demonstrate weak recognition of representational measurement principles together with strong endorsement of assumptions supporting utility-based cost-effectiveness analysis. The consistency itself becomes noteworthy.

A recurrent pattern across diverse institutions suggests something broader than isolated methodological disagreement. Questions concerning dimensional homogeneity, arithmetic admissibility, latent trait measurement, and falsifiability are appearing repeatedly rather than sporadically. That repetition may indicate that the issue concerns the structure of HTA itself rather than one particular analytical technique.

This does not guarantee major change. Disciplines can continue operating despite unresolved theoretical difficulties for long periods. Yet persistent foundational criticism differs from ordinary methodological adjustment. Once questions begin involving the scientific status of measures themselves, the discussion becomes more difficult to dismiss as merely technical.

The issue therefore may not be whether this debate disappears next year or the year after. The more important question may be whether institutions should assume it will disappear. Scientific history offers many examples where apparently minor foundational criticisms eventually proved far more significant than initially imagined. Prudent institutions may therefore wish to consider not whether the issue survives, but what preparations may be sensible if it does.

QUESTION 4: “IF CONCERNS REGARDING UTILITIES AND QALYS EXPAND INTERNATIONALLY, WHAT ARE THE IMPLICATIONS FOR GRADUATE EDUCATION?”

This question may ultimately prove more important than arguments over reimbursement systems, simulation models, or cost-effectiveness thresholds. Research methodologies can change gradually

and agencies can revise guidance documents, but graduate education determines how future generations understand the structure of evidence itself. Educational systems reproduce assumptions, analytical habits, and methodological frameworks. If foundational assumptions are not questioned during training, they frequently become accepted as natural and self-evident.

Current HTA and HEOR education programs are largely organized around a familiar sequence. Students are introduced to economic evaluation, utility assessment, QALYs, cost-effectiveness analysis, decision modeling, probabilistic sensitivity analysis, and simulation techniques. These methods are often presented as established analytical tools that support evidence-based healthcare decision-making. Considerable attention is given to statistical sophistication and model construction.

The issue raised by recent interrogations is not that these methods are taught. Rather, it concerns what may be absent from the curriculum.

Across more than 230 interrogations involving agencies, journals, pharmacy schools, and research centers in 28 countries, one pattern repeatedly appears: weak recognition of representational measurement requirements. Concepts such as dimensional homogeneity, admissible transformations, latent versus manifest attributes, falsifiability, and Rasch measurement often occupy little or no visible role within HTA-related knowledge bases. Measurement frequently appears as an assumed background condition rather than a prerequisite for quantitative claims.

If concerns regarding utilities, QALYs, and simulation frameworks continue to expand internationally, future students may ask increasingly uncomfortable questions. Why was multiplication undertaken before establishing ratio properties? Why were preference scores treated as though they possessed interval or ratio measurement characteristics? Why were latent constructs represented through ordinal summation rather than invariant measurement structures? Why was representational measurement largely absent from graduate instruction?

These are not trivial questions because they concern the scientific legitimacy of the methods students are being trained to use.

The issue is not that existing curricula suddenly become invalid. Graduate programs evolve continuously. New techniques, analytical frameworks, and research methods appear regularly. The concern is whether HTA education has inadvertently normalized measurement assumptions that were never explicitly examined.

If this proves correct, curriculum revision becomes more than an administrative adjustment. It becomes a question of scientific responsibility. Students entering health economics, pharmacy, outcomes research, and healthcare policy programs may increasingly require explicit understanding of measurement principles before proceeding to arithmetic manipulation and modeling exercises.

The implication is not immediate abandonment of existing courses. Rather, institutions may need to ask whether future graduates should merely learn current conventions or whether they should also understand the foundational assumptions upon which those conventions depend. Ultimately,

the future direction of HTA may be determined less by agencies and journals than by what students are taught in classrooms today.

QUESTION 5: "COULD RESEARCH CENTERS SIMPLY WAIT UNTIL EXTERNAL AGENCIES DECIDE WHETHER CHANGE IS NECESSARY?"

This is an understandable position. Research centers operate within wider institutional environments shaped by reimbursement agencies, government policy frameworks, journals, grant systems, and professional organizations. If bodies such as PBAC, PHARMAC, NICE, CADTH, ICER, or international organizations continue endorsing existing approaches, then there may appear little reason for individual centers to move independently. Why take methodological risks if larger institutions have not yet signaled the need for change?

From an administrative perspective, waiting may seem entirely rational. Established HTA frameworks remain deeply embedded within healthcare decision-making systems. Funding mechanisms continue operating; journals continue publishing cost-effectiveness studies; agencies continue requesting utility-based analyses; and graduate students continue learning accepted techniques. Existing systems possess considerable institutional stability. In the short term, maintaining alignment with prevailing practice may appear both safe and practical.

The difficulty is that waiting also transfers intellectual leadership elsewhere.

The recent AI large language model interrogations suggest that a recurrent pattern of measurement inversion now extends across agencies, academic centers, journals, pharmacy schools, and research environments internationally. If representational measurement concerns continue attracting attention, then future discussions may increasingly focus not simply upon analytical refinements but upon the foundational legitimacy of quantitative claims themselves.

Historically, scientific change rarely begins through centralized administrative directives. New ideas often emerge at the margins of established systems before gradually influencing broader practice. Universities and research centers have traditionally played a central role in that process. Their purpose is not merely to reproduce existing assumptions but to examine whether those assumptions remain defensible.

Waiting for agencies to decide may therefore create an unintended consequence. Institutions that delay engagement risk becoming passive recipients of change rather than participants in shaping it. If methodological transition eventually becomes necessary, those who entered the discussion early may influence curriculum development, research priorities, and future standards. Those who remained observers may later find themselves adapting to frameworks developed elsewhere.

The issue is not whether research centers should immediately abandon existing methods. Existing HTA systems will continue operating for many years. The question is whether institutions should begin exploring alternatives before external developments force reactive adjustment.

In practical terms, there is little disadvantage in early examination of representational measurement, Rasch latent trait measurement, evaluable claims, and dimensional homogeneity. Exploration does not imply commitment. It simply preserves intellectual flexibility.

The future of HTA may ultimately depend not on who waited longest, but on who recognized early that foundational questions were beginning to change. Research Centers therefore face a choice: wait for others to define the next stage of HTA, or participate in shaping it themselves.

QUESTION 6: "WHAT IS THE ACTUAL RISK OF DOING NOTHING?"

The immediate response may be: perhaps there is very little risk at all. Research centers continue publishing papers, agencies continue making reimbursement decisions, journals continue accepting submissions, and graduate programs continue teaching established methodologies. Existing HTA structures appear stable and internationally entrenched. From this perspective, concerns regarding representational measurement may seem largely theoretical and disconnected from day-to-day research activity.

Indeed, the risk is not sudden institutional collapse. No research center will close tomorrow because of questions surrounding dimensional homogeneity or Rasch measurement. Grants will continue to be awarded; cost-effectiveness analyses will continue to be published; agencies will continue requesting economic models. Existing systems possess substantial inertia and can persist for long periods despite unresolved methodological questions.

The more important issue concerns gradual rather than immediate consequences.

Over the past six months, more than 230 AI large language model interrogations across 28 countries have produced a remarkably consistent pattern. Across agencies, research centers, journals, pharmacy schools, and HTA environments, there appears weak recognition of representational measurement together with strong endorsement of assumptions supporting utilities, QALYs, simulation frameworks, and cost-effectiveness claims. Repetition across multiple countries and institutions suggests that the issue is not isolated. The pattern itself increasingly becomes evidence.

If these findings continue attracting attention, then questions may slowly shift from methodological details toward scientific foundations. Why are arithmetic operations undertaken without established measurement properties? Why are latent constructs represented through preference scores and ordinal aggregation? Why do simulation claims often remain beyond practical empirical evaluation? Why does measurement appear secondary to modeling?

Initially these questions may seem peripheral. Yet scientific credibility rarely changes abruptly. More commonly, confidence erodes gradually as unresolved anomalies accumulate. Criticisms that once appeared technical begin influencing broader perceptions regarding legitimacy, educational priorities, and research direction.

The risk therefore is not operational disruption. The risk is that institutions increasingly find themselves defending inherited assumptions whose foundations become progressively more

difficult to justify. Research centers may move from shaping methodological development to protecting established conventions. Intellectual leadership becomes defensive rather than exploratory.

Perhaps the greatest risk is therefore one of preparedness. Existing HTA systems may continue functioning successfully for years. But if methodological transition eventually becomes necessary, institutions that delayed engagement may find themselves responding reactively rather than leading constructively.

The issue is not whether change is inevitable. The issue is whether prudent institutions should prepare for the possibility. As the scouting principle suggests: *Be prepared.* Scientific disciplines rarely remain unchanged indefinitely; the question is whether HTA chooses to anticipate change or waits until circumstances leave fewer choices available.

CONCLUSION: PREPAREDNESS, LEADERSHIP AND THE NEXT STAGE OF HTA

The purpose of these questions has not been to argue that HTA institutions should immediately dismantle existing teaching programs, reimbursement structures, or research agendas. Existing frameworks are deeply embedded within healthcare systems and will continue to influence decision-making for many years. Nor does the presence of measurement inversion imply that all current HTA activity lacks value. Much descriptive research, epidemiology, health services analysis, and policy evaluation remains informative and important.

The issue raised by the recent AI large language model interrogations is narrower but potentially more significant. Across more than 230 interrogations in 28 countries, a remarkably consistent pattern has emerged: weak recognition of representational measurement principles together with strong endorsement of assumptions supporting utilities, QALYs, simulation models, and cost-effectiveness claims. If these findings continue attracting international attention, then questions concerning the scientific foundations of HTA may become increasingly difficult to ignore.

The Australian findings are particularly noteworthy. Interrogations of Australian research centers and HTA agencies produced profiles closely aligned with those observed internationally. At the same time, these assessments have attracted increasing attention. The PBAC interrogation alone received more than 550 visits within a relatively short period following release. Whether this level of interest reflects agreement, curiosity, skepticism, or concern is unclear. Nevertheless, it suggests that questions surrounding representational measurement and HTA methodology are now entering wider discussion rather than remaining isolated technical observations.

The practical question therefore becomes one of preparedness. What should research centers and academic programs do if foundational assumptions underlying current HTA frameworks become increasingly contested? Waiting for consensus may appear sensible, but scientific transitions rarely begin through universal agreement. Paradigm shifts more commonly emerge through gradual exploration, local experimentation, and early institutional leadership.

Universities and research centers occupy a distinctive position in this process. Their purpose extends beyond reproducing established conventions. They also provide environments in which assumptions may be examined, challenged, and revised. This raises the possibility that some centers may begin considering curriculum changes before broader institutional agreement emerges.

Such changes need not be revolutionary. Initial steps might involve introducing explicit discussion of representational measurement, dimensional homogeneity, latent versus manifest attributes, Rasch measurement, and evaluable claims alongside existing HTA frameworks. Students could be exposed both to current conventions and to the foundational assumptions upon which those conventions depend. The objective would not be immediate replacement of existing approaches but greater methodological transparency and scientific preparedness.

Historically, institutions that engage early in periods of scientific transition often help define the future direction of disciplines. Others adapt later. Whether HTA is approaching such a transition remains uncertain. But if representational measurement increasingly becomes part of the international discussion, then institutions may eventually face a choice: lead, follow, or defend inherited assumptions.

The issue may therefore no longer be whether measurement inversion exists. The more important question may be whether institutions are prepared for the possibility that the next stage of HTA development has already begun.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.