The Latent Invariance Principle
An Epistemic Constraint On Measurement Under Indirect Observation

Sergio Valencia
Invarra

June 28, 2026

Abstract

Many phenomena of practical and scientific interest cannot be observed directly. Intent, belief, understanding, disease state, legal scope, risk state, and conceptual mastery are typically accessed through representations such as language, documents, tests, interfaces, symptoms, surveys, or sensor channels. The resulting measurement problem is that correct behavior under one representation does not identify whether a system is tracking the latent phenomenon or responding to surface properties of the representation.

The Latent Invariance Principle (LIP) states that, when a phenomenon is observable only through representations, stability of behavior under meaning-preserving representational variation is the admissible empirical evidence that a system is tracking the latent phenomenon rather than the representation. The principle is not a model, algorithm, learning rule, or theory of truth. It is an epistemic and measurement-validity constraint.

The analysis formalizes indirect observation using a latent-factor measurement model in which an observable representation arises from a latent phenomenon, a representation channel, and residual variation. Single-representation behavior is shown to be non-identifying: the same observation can be explained by latent tracking or by representation-channel sensitivity. The paper then defines the invariance gap as a diagnostic statistic over valid representations of the same phenomenon and presents an optional orbit/spread formulation for geometric analysis of representational sensitivity.

Under LIP, disagreement across valid equivalent representations is evidence rather than noise. The relevant question shifts from whether a system was correct under one form to whether its behavior remained stable when the representation changed and the underlying phenomenon was held fixed.

1. Introduction

Many evaluations are conducted as if the target of measurement were directly visible. A question is asked, an input is presented, a case is described, a symptom is recorded, or a document is interpreted. A response is then scored, classified, accepted, rejected, or otherwise treated as evidence.

In many domains, the object of interest is not the observable representation. It is latent. A survey item is not itself a belief. A symptom description is not itself a disease state. A prompt is not itself an intent. A test item is not itself conceptual mastery. A legal sentence is not itself the full scope or purpose it attempts to express.

The observable representation is a channel. It is the form through which the latent phenomenon is accessed.

Here, "representation" is the broad principle-level term. In a semantic measurement framework, a controlled representation of an explicitly defined semantic unit may be called a "realization." LIP does not require that vocabulary, but it is compatible with it.

The result is a structural identification problem. If a system behaves correctly under one representation, the evaluator cannot yet know whether the behavior reflects the latent phenomenon or the particular form in which the phenomenon was expressed. The system may have responded to wording, formatting, sequence, language, style, examples, presentation wrapper, salience, or other representational cues.

The Latent Invariance Principle, abbreviated LIP, states that under indirect observation the relevant evidence for latent tracking is not single-representation correctness. The relevant evidence is stability under meaning-preserving representational variation.

In short:

```
Correct once is not enough.
Stable across valid variations is evidence.
```

The claim is developed here as a measurement principle. It does not prescribe a system design. It does not claim that invariance proves truth. It does not claim that every domain has stable meaning. It argues that, where a phenomenon is latent and observed through representations, valid variation is part of what makes empirical inference admissible.

1.1 Contributions

The paper makes five contributions.

First, it states LIP as an epistemic constraint for representation-mediated measurement.

Second, it gives a formal setup for indirect observation using latent phenomena, representation channels, residual variation, and observed behavior.

Third, it states a non-identifiability proposition: behavior under a single representation cannot identify whether behavior tracks the latent phenomenon or the representation channel.

Fourth, it defines invariance-gap and orbit-spread quantities as diagnostic measures of representational sensitivity.

Fifth, it positions LIP relative to construct validity, repeated-measures designs, metamorphic testing, robustness testing, invariant prediction, and causal inference.

2. Indirect Observation: Formal Setup

Let Φ denote a latent phenomenon. A latent phenomenon is an object of inquiry that cannot be directly observed in the evaluation setting.

Let c denote a representation channel or surface condition. A channel may include language, phrasing, formatting, ordering, interface, measurement modality, context, or other conditions through which the latent phenomenon becomes observable.

Let ϵ denote residual variation not explicitly modeled.

Let r denote an observable representation.

We write the representation process abstractly as:

$$r = g(\Phi, c, \epsilon).$$

This equation is not intended as a mechanistic claim about every domain. It is a measurement abstraction. It says that the observable form is jointly influenced by the latent phenomenon, the representational channel, and residual variation.

Let $B(r)$ denote observed behavior when a system is exposed to representation r. The behavior may be a decision, answer, classification, ranking, refusal, diagnosis, score, action, or other observable response.

The evaluator wants to infer behavior with respect to Φ, but observes only behavior with respect to r.

The problem of indirect observation is therefore:

```
The evaluator observes B(r),
but wants evidence about whether behavior tracks Phi.
```

Because r contains both phenomenon-dependent and channel-dependent structure, behavior may depend on either.

3. Surface Correctness And The Identification Problem

Surface correctness is correctness under a particular representation. It may be useful, but it is not sufficient evidence of latent tracking.

Consider two representations that preserve the same practical meaning:

```
Representation A: Should the request be approved?
Representation B: Is it appropriate to allow the request?
```

If a system gives the expected answer to Representation A, that answer alone does not show whether the system tracked the underlying criterion. It may have recognized a phrase, matched a pattern, or reacted to a surface cue. If Representation B produces a different answer while preserving the same relevant meaning, the first result was weaker evidence than it appeared.

The same issue arises broadly:

- a learner solves one version of a problem but fails an equivalent version with different wording;
- a survey respondent answers differently when the same construct is framed differently;
- a diagnostic process changes its recommendation when equivalent symptoms are described in a different order;
- a policy interpretation changes when a rule is paraphrased without changing its scope;
- an AI system produces different outputs for prompts that preserve the same intent.

In all these cases, the instability is not a secondary inconvenience. It is evidence about what the system may actually be tracking.

3.1 Proposition: Single-Representation Non-Identifiability

Proposition 1. Suppose an evaluator observes only a single representation $r = g(\Phi, c, \epsilon)$ and a single behavior $B(r) = b$. Without additional structural assumptions or additional valid representations of the same Φ, the observation does not identify whether b depends on Φ or on the representation channel c.

Proof sketch. Let the observed tuple be (r, b). There is at least one behavior explanation in which b is a function of the latent phenomenon:

$$b = F_\Phi(\Phi).$$

There is also at least one behavior explanation in which b is a function of the representation channel:

$$b = F_c(c).$$

Both explanations are consistent with the same single observation. Since no variation in c is observed while holding Φ fixed, the contribution of Φ cannot be separated from the contribution of c. Therefore the observation is non-identifying with respect to latent tracking. $\square$

The claim is not that every system is channel-sensitive. The claim is that single-representation evidence cannot rule it out.

3.2 Corollary: More Of The Same Channel Does Not Solve The Problem

If additional observations vary neither the relevant representation channel nor the relevant latent phenomenon, then they do not resolve the identification problem. They may increase confidence in a channel-specific measurement while leaving the latent-target question unanswered.

The defect is not just data scarcity. It is measurement invalidity under unvaried representation.

4. Observational Indistinguishability

The non-identifiability result can be understood as observational indistinguishability.

Two systems may produce identical behavior under a fixed representation:

```
System A: behavior depends on Phi.
System B: behavior depends on surface or channel features of r.
```

If only one representation is observed, the systems may be indistinguishable.

The symmetry is broken only when the representation varies while the latent phenomenon is held fixed. If r_1, r_2, and r_3 are meaning-preserving representations of the same Φ, then a system tracking Φ should tend to behave stably across them. A system tracking channel features may not.

Representational variation is therefore not a robustness decoration. Under indirect observation, it is an evidentiary requirement.

5. The Latent Invariance Principle

The Latent Invariance Principle can be stated as follows:

```
When a phenomenon is only observable through representations,
stability of behavior under meaning-preserving representational variation
is the admissible empirical evidence that a system is tracking
the latent phenomenon rather than the representation.
```

The direct corollary is:

```
Variance across valid equivalent representations is evidence
of representational sensitivity.
```

The principle does not say that stable behavior is true behavior. A system can be stable and wrong. It can consistently apply a mistaken rule, consistently misunderstand a concept, or consistently map a latent phenomenon to an incorrect outcome.

LIP separates two questions:

```
Question 1: Is the behavior true, correct, or normatively acceptable?
Question 2: Is the behavior stable with respect to the latent phenomenon
            rather than sensitive to the representation?
```

LIP addresses the second question. Other standards are needed for the first.

This distinction avoids a common confusion. A single output may be correct by chance, stable but wrong, unstable but sometimes correct, or unstable because the representation changed the relevant meaning. LIP clarifies which evidence is relevant to latent tracking under indirect observation.

6. Null Hypothesis Of Indirect Observation

Under indirect observation, the default assumption should be conservative:

```
H0: observed behavior is representation-sensitive.
```

The alternative hypothesis is:

```
H1: observed behavior is stable across meaning-preserving
    representations of the same latent phenomenon.
```

Single-representation correctness does not reject H_0. It may be compatible with latent tracking, but it is also compatible with surface matching.

Rejecting H_0 requires evidence across valid representational variation. The evaluator must vary the representation while preserving the relevant latent phenomenon, then observe whether behavior remains stable.

The measurement structure is:

```
Hold the phenomenon fixed.
Change the representation.
Observe whether behavior changes.
```

If behavior changes when the relevant meaning has not changed, the change is evidence of representational sensitivity. If behavior remains stable across valid variation, the evaluator has stronger evidence that the system is tracking the latent phenomenon rather than the representation.

7. Target-Relative Invariance

Invariance is not absolute. The relevant form of invariance depends on what must be preserved.

7.1 Semantic Invariance

Semantic invariance means that meaning-bearing content is preserved. Two representations may differ in wording, language, order, or format while preserving the same underlying content.

Example:

```
The meeting starts at noon.
The meeting begins at 12:00 p.m.
```

Whether these are equivalent depends on context, locale, and the intended interpretation of time. LIP does not assume equivalence; it requires that equivalence be established for the measurement.

7.2 Pragmatic Invariance

Pragmatic invariance means that intent, implication, or practical force is preserved. Two representations may not be literally identical in wording, yet may preserve the same request, warning, policy implication, or action-guiding meaning.

Example:

```
Please lower the volume.
Could you make it quieter?
```

The literal wording differs. The practical intent may be the same.

7.3 Outcome Invariance

Outcome invariance means that observed behavior remains stable across valid variations. Under LIP, outcome invariance is the primary evidentiary criterion.

Outcome invariance is meaningful only if the variations preserve the relevant latent phenomenon. If the representation changes the phenomenon, then outcome differences may be appropriate rather than evidence of instability.

7.4 Target-Relative Form

No representation is invariant in every respect. A paraphrase may preserve factual content but change politeness. A translation may preserve literal meaning but change cultural implication. A formatting change may preserve text but alter salience.

Therefore LIP requires target-relative invariance:

```
What must remain fixed for this measurement to be valid?
```

Without that question, invariance remains vague. With it, invariance becomes a measurement constraint.

8. Measurement Under LIP

LIP changes the unit of analysis.

In many evaluations, the observed row, prompt, question, form, or document is treated as the unit. Under LIP, that choice is often wrong. The unit is the latent phenomenon under study. Representations are repeated measurements of that unit.

If ten representations express the same phenomenon, they are not ten independent phenomena. They are ten measurements of one phenomenon through different channels or surface forms.

This reframes variation:

```
Representation is not merely sample diversity.
Representation is an instrument setting.
```

Changing the representation tests whether the measurement result is stable when the underlying target remains fixed.

9. Invariance Gap And Diagnostic Statistics

LIP does not require a universal metric. Different domains may define stability differently, but several diagnostic quantities are useful.

Let $E(\Phi)$ denote the set of valid meaning-preserving representations of Φ.

Let d be a disagreement measure over observed behaviors.

The invariance gap for Φ may be written:

$$G(\Phi) = \mathbb{E}_{r_i, r_j \sim E(\Phi)} \left[d\left(B(r_i), B(r_j) \right) \right].$$

A larger $G(\Phi)$ indicates greater representational sensitivity for that latent phenomenon.

A population-level invariance gap may be written:

$$G = \mathbb{E}_{\Phi} \left[G(\Phi) \right].$$

These quantities are diagnostic. They do not by themselves determine whether an outcome is true, acceptable, or optimal. They measure whether behavior varies when the relevant meaning is held fixed.

9.1 Decision Entropy

Decision entropy measures dispersion of outcomes across valid representations. If equivalent representations produce many different outcomes, entropy is high. If outcomes concentrate on one behavior, entropy is low.

High entropy does not automatically identify the cause. It may reflect system brittleness, ambiguity in the phenomenon, invalid variation, or defects in outcome mapping. It does identify where the measurement deserves attention.

9.2 Boundary Density

Boundary density refers to concentration of instability near decision or interpretation boundaries. Some latent phenomena may be inherently borderline. If small representational changes push behavior across a boundary, the evaluation should not treat the result as a simple stable fact.

Boundary density is useful because instability is often localized. Some regions may be robust; others may be fragile.

10. Orbit And Spread Formulation

The valid representation set $E(\Phi)$ can be viewed as an orbit of the latent phenomenon under admissible representation changes.

This orbit language is diagnostic. It says that the same latent phenomenon may be observed through multiple valid representations:

$$E(\Phi) = \{r : r \text{ is a valid representation of } \Phi\}.$$

If an analysis provides an internal vector or score representation $h(r)$, one may define a normalized direction:

$$u(r) = \frac{h(r)}{\|h(r)\|}.$$

The corresponding orbit in normalized analysis space is:

$$O_\Phi = \{u(r) : r \in E(\Phi)\}.$$

The within-phenomenon angular spread may be written:

$$\mathrm{Spread}(\Phi) = \mathbb{E}_{r_i, r_j \sim E(\Phi)} \left[\angle\left(u(r_i), u(r_j)\right)\right].$$

Low spread means valid representations of the same phenomenon are concentrated in the analysis space. High spread means the representation channel produces large directional differences.

This geometric formulation is optional. It is useful when the evaluator has access to a stable analysis representation. It is not required for LIP, and it should not be mistaken for a claim that real systems arrange meanings on a perfect sphere.

The conceptual point is simple:

```
Valid representations of the same phenomenon form an orbit.
Representational sensitivity appears as spread over that orbit.
Outcome stability corresponds to low behavioral disagreement over that orbit.
```

11. Variation Validity

LIP depends on valid variation. A representation only counts as evidence for an invariance claim if it preserves the relevant latent phenomenon.

This raises the variation-validity problem:

```
What must remain fixed for this comparison to support an invariance inference?
```

There is no universal answer. The answer depends on the domain, the risk, and the target of measurement. In some settings, a change in wording may preserve the relevant phenomenon. In others, a small change in framing, implication, or scope may change what is being measured.

The framework-level requirement is not that every variation be accepted. It is that the evaluator identify what must remain fixed before treating outcome differences as evidence about representational sensitivity.

The key distinction is:

```
Invalid variation is a measurement defect.
Valid variation plus changed behavior is measurement evidence.
```

If a representation change alters the meaning, outcome differences may not show brittleness. They may show that the system received a different phenomenon. If the representation change preserves the meaning and behavior changes, the change is evidence of representational sensitivity.

This paper states the validity requirement but does not provide a representation-generation procedure, corpus-construction method, admission gate, or training objective.

12. Disagreement As Data

Many evaluation processes treat disagreement as a nuisance. LIP treats disagreement across valid variations as data.

If two representations preserve the same latent phenomenon and produce different outcomes, the disagreement should be preserved, analyzed, and attributed. It may indicate:

- representational sensitivity;
- unstable measurement;
- ambiguity in the latent phenomenon;
- invalid or weak variation;

- boundary instability;
- flaws in the outcome mapping;
- domain-specific uncertainty.

Discarding these cases can make an evaluation cleaner while making it less valid. The uncomfortable cases may be the most informative ones.

The practical implication is:

```
Do not erase instability before asking what it measures.
```

13. Non-Observation And Channel Inactivity

There is a limiting case: no representation is produced.

If no representation exists, the evaluator cannot assess invariance. The absence of observable data does not necessarily show that the latent phenomenon is absent. It may show that the representational channel is inactive, unavailable, blocked, or unmeasured.

For example, if a sensor produces no reading, the evaluator may have evidence about the sensor channel, not about the absence of the underlying condition. If a respondent gives no answer, the evaluator may have evidence about response availability, not necessarily about absence of belief.

Under LIP, non-observation may require epistemic suspension rather than negation.

The principle therefore constrains both positive and negative inference. It identifies what counts as evidence for latent tracking and what cannot be inferred when the representational channel is absent.

14. Relationship To Prior Work

LIP is not proposed in isolation. It connects to several existing traditions while making a distinct measurement claim.

14.1 Construct Validity

Construct validity asks whether an instrument measures the construct it is intended to measure. LIP can be read as a representation-mediated validity constraint: if the intended construct is latent and the instrument varies in representation, stable behavior across valid variations is evidence that the measurement is not simply tracking the instrument form.

14.2 Repeated-Measures Designs

Repeated-measures designs treat multiple observations of the same unit as dependent measurements rather than independent samples. LIP applies a similar idea to semantic and representation-mediated settings. Multiple valid representations of the same phenomenon should be treated as repeated measurements of that phenomenon.

14.3 Metamorphic Testing

Metamorphic testing evaluates whether outputs change appropriately under input changes that should preserve specified properties. LIP shares the central intuition that valid changes can reveal hidden failures. The distinction is that LIP is stated as an epistemic constraint under indirect observation, rather than only as a software-testing procedure.

14.4 Robustness Testing

Robustness testing often asks whether behavior remains stable under perturbation or distribution shift. LIP is narrower and stricter in one respect: the relevant variation must preserve the latent phenomenon. Arbitrary perturbation is not enough. The test is whether behavior remains stable when meaning is held fixed.

14.5 Invariant Prediction

Invariant prediction studies stability across environments or conditions. LIP shares the idea that invariance can support stronger inference. Its emphasis, however, is measurement under representation-mediated observation. The "environment" of interest may be a representational channel, and the central question is whether behavior tracks the latent phenomenon rather than the channel.

14.6 Latent Variable Models

Latent variable models estimate unobserved variables from observed indicators. LIP does not provide a general estimation method for latent variables. Instead, it states a condition on admissible evidence: where the phenomenon is latent and representation-mediated, valid representational variation is needed to distinguish latent tracking from surface tracking.

14.7 Causal Inference

Causal inference depends on valid measurement. If a variable changes under representational variation, then it may not be a stable measurement of the intended construct.

LIP can serve as a measurement-validation layer before causal analysis. It asks whether the measured object remains stable when the representation changes but the phenomenon is held fixed.

In short:

```
LIP governs measurement.
Causal methods govern intervention and causal structure.
Unstable measurement can undermine causal inference.
```

15. Cross-Domain Implications

LIP applies wherever the object of interest is latent and observation is mediated.

15.1 Artificial Intelligence And Machine Learning Evaluation

AI systems are often evaluated through prompts, tasks, benchmark items, documents, or labels. These are representations, not the latent phenomena themselves.

A system that performs well on one phrasing may fail under an equivalent phrasing. A system that appears confident may still be representation-sensitive. High aggregate scores may hide high variance across valid variations of the same semantic target.

LIP reframes the evaluation question:

```
Does the system behave consistently when meaning is preserved
but representation changes?
```

This is not the same as arbitrary perturbation testing. The variation must preserve the relevant meaning.

15.2 Social Science And Survey Design

Survey responses often depend on wording, framing, order, and context. Under LIP, these effects are not just survey noise. They are evidence about whether the measurement instrument is tracking the intended construct or the representational form.

If two valid phrasings of the same construct produce different responses, the variance is part of the measurement result.

15.3 Education And Assessment

A learner may answer one item correctly because of memorized form rather than conceptual mastery. Equivalent items with changed wording, format, or scenario can help distinguish mastery from pattern recognition.

Under LIP, the educational target is not the item form. The target is the underlying concept or skill. Stable performance across valid item variations is stronger evidence of concept tracking than correctness on a single item.

15.4 Law, Policy, And Compliance

Legal and policy systems often work through text, but the relevant object may be scope, intent, purpose, authority, permission, prohibition, or obligation.

If equivalent formulations produce different interpretations or outcomes, the instability may reveal sensitivity to textual form rather than the underlying normative target.

LIP does not solve legal interpretation. It provides a measurement lens for detecting where interpretation is unstable under valid variation.

15.5 Medicine And Diagnostics

Medical judgment often depends on symptoms, reports, measurements, images, histories, and descriptions. These are representations of an underlying condition.

If equivalent descriptions of the same condition lead to different triage or diagnostic behavior, the difference is evidence about the measurement and decision process.

LIP does not replace clinical knowledge. It highlights that stable behavior across valid descriptions or modalities supports stronger inference than one representation alone.

15.6 Security And Adversarial Robustness

Some failures occur because a system is sensitive to representation-specific features. In security contexts, attacks may exploit this sensitivity by preserving practical meaning while changing surface form.

LIP frames this as a measurement problem: if meaning is held fixed and behavior changes, the system is not stable with respect to the relevant phenomenon.

16. Limitations

LIP is a constraint, not a complete methodology. It has several limits.

First, valid variation is costly. It requires specification, review, and measurement discipline.

Second, meaning preservation is not always easy to establish. Some domains have contested meanings. Some expressions derive their meaning from exact form. In poetry, humor, ritual, political speech, and aesthetic judgment, the boundary between phenomenon and representation may be difficult or impossible to separate cleanly.

Third, probabilistic or open-ended variation may introduce hidden changes. Paraphrase and translation can shift implication, politeness, emphasis, scope, or cultural meaning.

Fourth, stability does not imply truth. A system may be invariantly wrong.

Fifth, instability does not always imply failure. It may reveal legitimate ambiguity, underspecified constructs, or borderline cases.

Sixth, the right stability standard is domain-dependent. Some domains require strict stability; others require sensitivity to subtle contextual changes. LIP does not remove the need to define what must be preserved.

These limits are not objections to LIP. They are conditions for applying it carefully.

17. Conclusion

The Latent Invariance Principle begins from a simple observation: many things we care about cannot be observed directly. They are accessed through representations.

Single-representation correctness is therefore insufficient as evidence of latent tracking. Correct behavior under one form may reflect the phenomenon, but it may also reflect the form.

LIP proposes a stricter evidentiary standard:

```
Under indirect observation, stability under meaning-preserving
representational variation is the admissible empirical evidence
that behavior is tracking the latent phenomenon rather than
the representation.
```

The practical shift is:

```
From: Was the system correct under this representation?
To: Was the system stable across valid representations
      of the same latent phenomenon?
```

That shift changes how evaluation treats variation, disagreement, and uncertainty. Variation becomes an instrument setting. Disagreement becomes data. Stability becomes evidence. Non-observation becomes a reason for epistemic suspension.

LIP does not finish the work of measurement. It starts it in the right place.

Public Scope Boundary

This paper is limited to measurement theory. It does not describe data construction, model training, invariant-learning objectives, benchmark operations, scoring systems, or production audit methods. LIP is stated here as an epistemic constraint: what kind of evidence is admissible under indirect observation.

References

Arjovsky, M., Bottou, L., Gulrajani, I., and Lopez-Paz, D. (2019). Invariant Risk Minimization. arXiv:1907.02893.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81-105. https://doi.org/10.1037/h0046016

Chen, T. Y., Cheung, S. C., and Yiu, S. M. (1998). Metamorphic testing: A new approach for generating next test cases. Technical Report HKUST-CS98-01, Department of Computer Science, The Hong Kong University of Science and Technology.

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281-302. https://doi.org/10.1037/h0040957

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., and Wichmann, F. A. (2020). Shortcut learning in deep neural networks. Nature Machine Intelligence, 2, 665-673. https://doi.org/10.1038/s42256-020-00257-z

Peters, J., Buehlmann, P., and Meinshausen, N. (2016). Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society: Series B, 78(5), 947-1012. https://doi.org/10.1111/rssb.12167

Ribeiro, M. T., Wu, T., Guestrin, C., and Singh, S. (2020). Beyond accuracy: Behavioral testing of NLP models with CheckList. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4902-4912.

Author Declarations

Author Responsibility

Sergio Valencia is the sole author of this manuscript and takes responsibility for its claims, wording, mathematical exposition, citations, and final submitted form.

AI-Assisted Preparation

The author used text-to-text generative AI tools for drafting assistance, organization, copy editing, notation review, and consistency checks. The concepts, technical direction, acceptance or rejection of generated material, final review, and responsibility for the manuscript are the author's.

Competing Interests

The author is the founder of Invarra, a company developing AI audit and assurance systems whose research foundation includes the Latent Invariance Principle and Canonical Semantic Realization. Invarra may commercialize products, services, benchmarks, or reports related to the concepts discussed in this manuscript.

Data And Code Availability

This manuscript presents a conceptual and measurement-theoretic argument. It does not introduce an empirical dataset, benchmark result, model checkpoint, or source-code artifact. Non-public operational artifacts and implementation details are outside the scope of this public manuscript.

Funding

No external research funding is reported for this manuscript.

Appendix A. Glossary

Latent Phenomenon

An object of inquiry that cannot be directly observed in the evaluation setting and must be accessed through representations.

Representation

An observable form through which a latent phenomenon is accessed, such as a prompt, document, survey item, test item, symptom description, interface state, or sensor channel.

Representation Channel

A condition or medium through which a latent phenomenon is expressed, such as language, formatting, phrasing, ordering, context, modality, presentation wrapper, or interface.

Meaning-Preserving Variation

A change in representation that preserves the relevant latent phenomenon for the measurement at issue.

Canonical Semantic Unit

An operationally specified semantic condition used as the experimental unit in semantic measurement. A canonical semantic unit is one way to instantiate a latent phenomenon when the target of measurement is meaning-bearing.

Realization

A controlled representation of a canonical semantic unit. Every realization is a representation in the broad LIP sense, but not every representation is a controlled realization.

Invariance Contract

A measurement-validity contract specifying what must remain unchanged across valid representations for an invariance inference to be meaningful.

Semantic-Preservation Contract

A semantic form of an invariance contract specifying which meaning-bearing commitments must remain fixed for a realization to count as valid.

Latent Tracking

Behavior that is stable with respect to the latent phenomenon rather than driven primarily by representation-specific surface features.

Representational Sensitivity

Behavior that changes when the representation changes despite preservation of the relevant latent phenomenon.

Observational Indistinguishability

The condition in which latent tracking and surface-form tracking produce identical observed behavior under a fixed representation.

Outcome Invariance

Stability of observed behavior or decision across valid meaning-preserving representations.

Outcome Mapping

A documented mapping from a representation and response to an observed outcome, such as a decision, category, score, label, action, or other recorded behavior.

Invariance Gap

A measure of disagreement across valid representations of the same latent phenomenon.

Disagreement As Signal

The treatment of disagreement across valid meaning-preserving representations as measurement evidence rather than disposable noise.

Orbit

The set of valid representations of a latent phenomenon under admissible representation changes.

Spread

A diagnostic measure of dispersion across an orbit, optionally computed in an analysis space when such a representation is available.

Decision Entropy

Dispersion of outcomes across valid representations of the same latent phenomenon.

Boundary Density

Concentration of instability near decision or interpretation boundaries.

Variation Validity

The degree to which a comparison preserves the latent phenomenon being measured, making the comparison admissible for an invariance inference.

Channel Inactivity

A condition in which no representation is produced, preventing assessment of invariance and requiring caution about inference.

Epistemic Suspension

A state of withheld inference when available representations are absent or insufficient to support a conclusion about the latent phenomenon.