Statistica Sinica Preprint No: SS-2026-0110	
Title	Statistical Disclosure Control for External Ai via Transformation Groups
Manuscript ID	SS-2026-0110
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0110
Complete List of Authors	Lingxuan Shao
Corresponding Authors	Lingxuan Shao
E-mails	shao_lingxuan@fudan.edu.cn
Notice: Accepted author version.	

Statistica Sinica: Newly accepted Paper

Statistical Disclosure Control for External AI via Transformation Groups

Lingxuan Shao

Department of Statistics and Data Science, School of Management, Fudan University

ORCID: 0000-0001-8061-7913

Abstract: Small enterprises increasingly rely on external AI for data analytics, which exposes sensitive summary statistics to the AI provider. We propose a practical framework for statistical disclosure control. It lets enterprises use advanced AI while withholding the statistics they deem confidential. Each data owner first applies a random, reversible transformation drawn from a pre-verified group, and submits only the transformed data. An inverse mapping then recovers the target statistics: the recovered output has the same distribution as the target. We introduce notions of traceability and transitivity for transformation groups and verify them through a two-step procedure. A multi-sample test based on Haar-measure sampling checks traceability; an orbit-based analysis then confirms transitivity. We implement the framework on a pottery-company dataset with five LLMs. In each task the test is not rejected under a suitable model and sampling measure, so that the external AI cannot identify the protected statistics from the submitted data.

Key words and phrases: Group orbit, Haar measure, Multi-sample distribution test.

1. Introduction

1.1 Balancing AI Utility and Data Security

In an era where the adoption of artificial intelligence has transitioned from a strategic advantage to a competitive necessity (Yesuf & Fields 2025), a growing number of small enterprises confront a critical dilemma. These organizations typically possess a valuable, proprietary dataset X_0 whose analysis could significantly enhance their statistical forecasting and strategic decision-making. To achieve this, they seek to leverage advanced, externally hosted AI models. However, this intention is starkly tempered by a paramount concern: the imperative to prevent the disclosure of sensitive business information—such as specific forecasts, proprietary metrics, or design parameters—to the external AI provider. This fundamental tension between leveraging superior external computational resources and safeguarding confidential commercial information defines the core problem addressed in this work.

For example, an artisan pottery company may hold a decade of detailed sales records, including pricing, marketing expenditure, customer ratings, and proprietary design attributes. Such data could train a model to forecast future demand—essential for production planning—yet the company cannot risk exposing its predictions or upcoming design specifications to outsiders. This concrete scenario underscores the broader tension outlined above and motivates the need for practical methods that balance statistical utility with information security. Interventions available to the enterprise act on one of two layers: a semantic layer,

Statistica Sinica: Newly accepted Paper

in which the instructions given to the AI are modified, and a numerical layer, in which the data themselves are modified.

The first approach is prompt engineering with semantic obfuscation, whereby the enterprise deliberately rephrases, omits, or generalizes key descriptive elements within the instructions provided to the AI. While this method aims to shield sensitive context by altering the semantic presentation of the task, it may distort or remove the precise intent and nuanced specifications that the model requires for accurate analysis (Elsharif et al. 2025). Consequently, the AI's comprehension of the problem becomes misaligned with the true objective, leading to outputs that are systematically biased or irrelevant—a compromise that fundamentally undermines the goal of data-driven decision-making.

In the numerical layer, the enterprise instead preserves the semantic integrity of the task and modifies only the data. Several families of methods act at this layer: local differential privacy, which perturbs each entry before the data leave the enterprise (Demelius et al. 2023); synthetic data generation; cryptographic approaches such as fully homomorphic encryption and secure multi-party computation; and reversible transformations of the data. This paper studies the last family. Section 1.2 compares these families in detail. In the reversible transformation method, the enterprise shares the transformed data $t(X_0)$ with the external AI, alongside the original, unambiguous problem description. The AI model therefore operates within the correct contextual and semantic framework, processes $t(X_0)$, and produces an output $A\{t(X_0)\}$. The core of this method lies in the enterprise's subsequent

Statistica Sinica: Newly accepted Paper

application of a bespoke recovery mapping R_t to this output, which is designed to invert the data transformation and yield an output whose distribution matches that of the true target $A\{X_0\}$.

Our work adopts this second path to achieve statistical disclosure control. Algorithm 1 formalizes the proposed transformation approach, which is designed to satisfy two essential and complementary objectives:

- **Traceability**: The enterprise must be able to reconstruct the target statistic $A\{X_0\}$ from the transformed output $A\{t(X_0)\}$, in the sense that the recovered output has the same distribution as the target.
- **Disclosure Control**: given the data $t(X_0)$ and the resulting output, the external AI provider cannot identify the target statistic $A\{X_0\}$.

This framework offers a structured and testable pathway for enterprises—especially those with limited in-house computational resources—to harness state-of-the-art AI capabilities without disclosing the statistics they designate as confidential.

1. Transform the original dataset X_0 into $t(X_0)$;
2. Submit $t(X_0)$ to an external AI provider and obtain the result $A\{t(X_0)\}$;
3. Apply a recovery mapping R_t to $A\{t(X_0)\}$ to obtain $R_t(A\{t(X_0)\})$;

Algorithm 1: General framework for statistical disclosure control.

Statistica Sinica: Newly accepted Paper

1.2 Related Literature

Our work is situated within the specific subfield of statistical disclosure control (SDC) (Hundepool et al. 2012, Green et al. 2024). Unlike general privacy-preserving techniques that aim for universal data protection, this paper focuses on preventing the leakage of specific, sensitive aggregates derived from data. Our framework advances this objective by introducing a transformation-based mechanism tailored for the external AI paradigm, which differs fundamentally from established approaches in its threat model, mechanism, and efficiency profile.

We clarify the distinction between our method and five key classes of techniques:

- Fully Homomorphic Encryption (FHE) enables arbitrary computations on encrypted data (Gong et al. 2023). However, its profound computational and communication costs render it largely impractical for training or querying complex AI models. In contrast, our method relies on lightweight, pre-processing transformations. The transformed data $t(X_0)$ remains structurally compatible with standard AI models, enabling practical delegation without requiring specialized, homomorphic-aware algorithms.

- Differential Privacy (DP) provides a rigorous framework for protecting individual privacy in statistical releases by injecting calibrated noise into queries or the data (Demelius et al. 2023). Its core mechanism is designed to obscure the contribution of any single record. Our protection target is different in kind: we do not aim to

Statistica Sinica: Newly accepted Paper

protect the contribution of individual records but to keep a designated statistic—for instance a strategic forecast—from being identified. This yields statistic-level rather than record-level confidentiality. The two families also differ in protection unit and in the form of the guarantee: a DP guarantee holds uniformly over neighboring datasets and concerns the contribution of a single record, whereas ours concerns one statistic and states which values remain possible, conditional on the sampling distribution being undisclosed. The two families are therefore complementary rather than mutually substitutable.

- Secure Multi-Party Computation (MPC) is architected for scenarios where multiple parties jointly compute a function without revealing their private inputs (Li et al. 2023). Our setting is fundamentally simpler: a single data owner delegates computation to a single external AI provider. The intricate interactive protocols of MPC are unnecessary for this two-party, client-server paradigm and would introduce disproportionate complexity and latency compared to our non-interactive, transformation-based approach.

- Synthetic Data Generation creates artificial datasets to preserve the statistical properties of the original data (Chapelle & Falissard 2023). A critical step in many such methods, however, involves estimating and releasing aggregate parameters (e.g., moments, correlations) to fit a generative model. This process directly discloses the very

Statistica Sinica: Newly accepted Paper

global statistics our framework is designed to protect. Our method circumvents this by ensuring the target statistic remains obscured throughout the pipeline; the external AI provider works on data from which this specific quantity cannot be inferred, even indirectly.

- The Matrix Masking framework (Duncan & Pearson 1991) unifies various disclosure limitation techniques, enabling systematic control over data utility and disclosure risk. The General Additive Data Perturbation (GADP) method (Muralidhar et al. 1999) further refines this approach by generating perturbed values that preserve all pairwise correlations, thus eliminating statistical bias for multivariate normal data. While these methods share our goal of protecting aggregates through transformation, they are designed for static data release or database query answering, and often require strong distributional assumptions or complex conditional generation. In contrast, our method is tailored for the interactive AI paradigm, offering a lightweight, pre-processing transformation that remains compatible with standard AI models without imposing computational overhead or requiring a priori distributional knowledge.

In summary, our framework aims to provide a precise, client-side mechanism for preventing the disclosure of pre-defined summary statistics, rather than offering blanket data-level privacy. This focus allows it to be both computationally lightweight—imposing minimal burden on the enterprise—and highly effective at preserving the utility of the data for its intended analytical purpose. This combination is particularly aligned with the operational

and security constraints of small enterprises seeking to leverage advanced external AI.

2. Candidate Properties, Transformation Groups, and the Threat Model

2.1 Candidate properties and transformation groups

This paper focuses on AI-driven tasks that extend beyond the computation of simple descriptive statistics (e.g., means or variances) to encompass more complex and strategic operations, such as prediction and statistical inference. A critical characteristic of the AI models we consider is their stochasticity. Unlike deterministic models that consistently produce the same output for a given input, stochastic models incorporate randomness in their generation process, leading to inherently variable outputs. We concentrate on this stochastic setting because it presents a richer and more statistically interesting challenge; the deterministic case, while simpler, is not our primary focus. Stochastic models offer several advantages: they can generate a diverse set of plausible solutions, help mitigate systematic biases across repeated queries, allow for the analysis of output distributions, and more faithfully reflect the uncertainty present in real-world data. Consequently, they are particularly valuable for tasks requiring exploration and robust decision-making. Formally, the AI operator $A\{\cdot\}$ in our framework is therefore treated as a random map.

A core challenge arises because the internal algorithm of the external AI model A is typically a black box. The selection of a suitable transformation-recovery pair (t, R_t) must therefore be based on high-level properties that the AI task may or may not possess. Each of

Statistica Sinica: Newly accepted Paper

them is treated as a candidate hypothesis: the enterprise specifies the corresponding family of transformations, and the family is retained only if it survives the test of Section 5 at the true dataset. The examples below therefore define candidates rather than assumptions.

Example 1 (Candidate transformation groups and the corresponding transformations)**.** The following are candidate properties of the black-box AI operator A, each with the transformation and recovery pair (t, R_t) that would protect summary statistics if the property held.

- **Scale Homogeneity:** If the operator A is homogeneous of degree one (i.e., $A\{\lambda X\} = \lambda A\{X\}$), one can apply the transformation $t(X) = \lambda X$ using a random positive scalar λ. The AI processes the transformed data, yielding output $A\{t(X)\} = \lambda \cdot A\{X\}$. The original result is recovered via the inverse scaling $R_t(y) = y/\lambda$.
- **Linearity:** If A behaves approximately as a linear map (i.e., $A\{PX\} \approx PA\{X\}$ for a matrix P), one can apply $t(X) = PX$ with a random non-singular matrix P. The AI then operates on $t(X)$, producing output $A\{t(X)\} \approx P \cdot A\{X\}$. Recovery is achieved by applying the inverse transformation: $R_t(y) = P^{-1}y$.
- **Angle-Preserving:** If the operator A is covariant under isometric transformations (i.e., $A\{PX\} = PA\{X\}$ for any rotation matrix P), one can apply $t(X) = PX$ with a random rotation matrix P with $\det P = 1$. The AI processes the transformed data, yielding output $A\{t(X)\} = P \cdot A\{X\}$. Recovery is achieved by applying the inverse rotation: $R_t(y) = P^{-1}y = P^T y$.

The transformations described in Example 1 illustrate how specific properties of A, if they hold, lead to practical protection schemes. To build a robust and reusable framework, however, we consider not isolated maps but structured families of transformations from

Statistica Sinica: Newly accepted Paper

which a specific t can be sampled randomly for each execution of Algorithm 1. A natural and algebraically tractable structure for such a family is that of a group—a set closed under composition and inversion. Remarkably, these three motivating examples in Example 1 correspond directly to classic matrix groups:

Example 2 (Transformation groups corresponding to Example 1)**.** Let the data space be $\mathcal{X} = \mathbb{R}^p$.

- **Scaling Group:** The set of all positive scale transformations,

$$G_{\text{scale}} = \{t_\lambda : X \mapsto \lambda X \mid \lambda \in \mathbb{R}_{>0}\},$$

  forms a group isomorphic to the multiplicative group of positive real numbers $(\mathbb{R}_{>0}, \times)$. The group operation (composition) corresponds to scalar multiplication: $t_{\lambda_1} \circ t_{\lambda_2} = t_{\lambda_1 \lambda_2}$, and the inverse of t_λ is $t_{1/\lambda}$.

- **General Linear Group:** The set of all invertible linear transformations,

$$G_{\text{GL}} = \{t_P : X \mapsto PX \mid P \in \text{GL}(p, \mathbb{R})\},$$

  is isomorphic to the general linear group $\text{GL}(p, \mathbb{R})$ of $p \times p$ non-singular matrices. Here, composition corresponds to matrix multiplication, and inversion corresponds to matrix inversion.

- **Rotation Group:** The set of all rotation transformations (without reflections),

$$G_{\text{SO}} = \{t_P : X \mapsto PX \mid P \in \text{SO}(p)\},$$

  is isomorphic to the special orthogonal group $\text{SO}(p) = \{P \in \mathbb{R}^{p \times p} \mid P^\top P = I_p, \det(P) = 1\}$. This group preserves Euclidean distances and angles, and also orientation. Composition and inversion are again given by matrix multiplication and transposition, respectively.

Statistica Sinica: Newly accepted Paper

The groups in Example 2 are candidates, not assumptions. Whether a given AI operator satisfies the corresponding properties is an empirical question. The task itself suggests what to expect. A sufficiently capable AI, asked for a forecast that respects the structure of the data, should return an estimator consistent with statistical intuition. Such an estimator inherits the invariances that the task imposes. Homogeneity of degree one implies invariance under the scaling group G_{scale}: rescaling a group of inputs and the response rescales the forecast by the same factor. Equivariance under rotations implies invariance under the rotation group G_{SO}: when only the relative geometry of the design coordinates matters, rotating them rotates the answer.

2.2 The testing framework at a glance

To study and justify whether a group G is applicable in our proposed framework, we introduce the concepts of *globally and locally traceable groups*, defined with respect to the randomness inherent in the AI operator A. The property of traceability guarantees that for every $t \in G$, a legitimate recovery map R_t exists, so that the enterprise can reconstruct $A\{X_0\}$ from the transformed output in the distributional sense—this formalizes the **Traceability** requirement.

Similarly, we introduce the concepts of *globally and locally transitive groups* to satisfy the **Disclosure Control** objective. A group acting transitively on the relevant data space ensures that the external AI provider cannot identify the sensitive statistic from the trans-

Statistica Sinica: Newly accepted Paper

formed data $t(X_0)$: every candidate value remains compatible with what it observes.

To implement Algorithm 1, we develop methods to test whether a given candidate transformation group G is traceable and transitive. To operationalize this testing in a stochastic context, we sample transformations from G according to a probability measure μ built on the *Haar measure*—the unique (up to scaling) left-invariant measure on a locally compact topological group. The measure μ is then constructed from the Haar measure by reweighting it with a positive weight function, and it is not disclosed to the AI provider. The Haar measure thus provides a principled foundation for sampling transformations from G and for analyzing the resulting distribution of the AI model's outputs. We then construct a formal hypothesis test for local traceability by verifying that the distribution of recovered outputs $R_t(A\{t(X_0)\})$ is identical to the distribution of the true target $A\{X_0\}$. Specifically, we combine random sampling from μ with a suitable multi-sample test for distributional equality; the choice of multi-sample test can be tailored based on the data's dimensionality, sample size, and computational constraints. Once a group G is found admissible by the test of Section 5 for local traceability, checking local transitivity simplifies to a straightforward computational procedure involving the analysis of group orbits.

This completes the practical workflow: an enterprise can select a candidate group, test its local traceability and local transitivity, and then implement Algorithm 1 to realize statistical disclosure control.

Statistica Sinica: Newly accepted Paper

2.3 The threat model and strengths of disclosure

Before proceeding we state the threat model on which the rest of the paper rests. (i) *Adversary*: the external AI provider, which may attempt to infer confidential quantities from what it receives. (ii) *Adversary Observation*: the submitted dataset $t(X_0)$ together with the full prompt, and the fact that repeated requests originate from the same client. (iii) *Background knowledge*: the AI provider may know the protocol and the transformation group G, and it can evaluate its own model A on any input it chooses; it does not know X_0, it does not know the realized transformation t, and it is not told the sampling distribution μ. (iv) *Adversary Capability*: it may combine its observations with auxiliary information and run inference attacks.

Within this model we distinguish three strengths of disclosure, ordered from the strongest to the weakest: reconstruction of the dataset X_0; exact identification of the target statistic $A\{X_0\}$; and statistical inference that shifts the AI provider's beliefs about the statistic. The first two are linked. If the AI provider could reconstruct X_0, it could evaluate $A\{X_0\}$ on that dataset and thereby identify the statistic; hence a guarantee that excludes identification also excludes reconstruction. The framework developed in Sections 3–6 delivers both: the external AI provider cannot identify $A\{X_0\}$ from the data the enterprise submits, and therefore cannot reconstruct X_0 either. The third strength is left open. Excluding inference is difficult, because what an AI provider can infer depends on the prior information it holds and not only on the data it receives. Because μ is not disclosed, the AI provider is given no basis on

which to rank those values, which limits but does not remove the leverage of an inference attack.

Remark 1 (Fixed data versus stochastic operator)**.** A crucial distinction between our framework and classical invariance concepts in statistics lies in the source of randomness. In our setting, the original dataset X_0 is *fixed* and non-stochastic; all randomness emanates from the external AI operator A, which we model as a stochastic map. Consequently, the transformed data $t(X_0)$ is itself a fixed quantity for a given realization of t and does not constitute a random variable or a statistic in the traditional sense. Classical invariance theory typically concerns statistics—functions of random samples whose distributions remain unchanged under a group of transformations acting on the sample space. Here, by contrast, we exploit potential invariance properties of the *operator* A itself (e.g., homogeneity, linearity, or equivariance under rotations). The transformation t is chosen not to stabilize a distribution, but to ensure that the action of A on $t(X_0)$ can be related back to its action on the original fixed X_0 via a known recovery map R_t. This operator-centric view of invariance fundamentally differentiates our approach from classical statistical invariance and is key to achieving traceability without requiring distributional assumptions about the data.

3. Traceability and Transitivity

Let the true dataset be denoted by X_0, residing in a *data space* $\mathcal{X}$. The stochastic AI operator A maps an input dataset $X \in \mathcal{X}$ to an output in the *decision space* Θ, which we

write as $A\{X\}$. Formally, A is a mapping

$$A : \Omega \times \mathcal{X} \to \Theta,$$

where Ω is the underlying sample space. We assume Θ is a Euclidean space endowed with its Borel σ-algebra, and for each fixed $X \in \mathcal{X}$, the map $\omega \mapsto A(\omega, X)$ is measurable.

Operationally, A is the end-to-end map that the AI provider implements: given a submitted dataset x and a fixed prompt template π, it returns a random answer $A_\pi\{x\}$, and the framework works with the distribution of that answer. The data space $\mathcal{X}$ may contain multi-source or structurally complex data and thus is assumed to be a topological space. In Algorithm 1, we require choosing a self-homeomorphism $t : \mathcal{X} \to \mathcal{X}$ from the group $\mathrm{Homeo}(\mathcal{X})$. For the subsequent analysis, we assume that $\mathrm{Homeo}(\mathcal{X})$, when equipped with the compact-open topology, forms a locally compact Hausdorff topological group. This assumption is standard in topology and transformation group theory, and it is satisfied under a mild topological condition on $\mathcal{X}$: namely, when $\mathcal{X}$ is a locally compact Hausdorff space that is also locally connected (Arens 1946, McCleary 2006). In particular, the transformation groups considered in Example 2 all satisfy these conditions, ensuring that the assumption holds throughout our discussion. We refer the reader to Section S2 of the Supplement for the definitions of the compact-open topology and of the local compactness and Hausdorff properties, together with standard examples.

With the topology provided by $\mathrm{Homeo}(\mathcal{X})$, we now introduce the central concept of a *globally traceable transformation*, which formalizes the existence of a deterministic recovery

procedure for the target statistic.

Definition 1 (Globally traceable transformation and group)**.** Let $t \in \text{Homeo}(\mathcal{X})$ be a self-homeomorphism of the data space. We say t is *globally traceable* if there exists a homeomorphism $R_t \in \text{Homeo}(\Theta)$ of the decision space such that

$$R_t\left(A\{t(X)\}\right) \overset{d}{=} A\{X\} \qquad \text{for all } X \in \mathcal{X},$$

where $\overset{d}{=}$ denotes equality in probability distributions over Θ. A subgroup $G \subset \text{Homeo}(\mathcal{X})$ consisting entirely of globally traceable transformations is called a *globally traceable group*.

An important algebraic property of globally traceable transformations is that they naturally form a subgroup of the homeomorphism group, as established by the following lemma.

Lemma 1 (Closure of globally traceable transformations under composition and inversion)**.** *Let $\mathcal{G} := \{t \in \text{Homeo}(\mathcal{X}) \mid t \text{ is globally traceable}\}$. Then $\mathcal{G}$ is a subgroup of $\text{Homeo}(\mathcal{X})$ under composition.*

Remark 2 (Recovery maps form an anti-isomorphic group)**.** For a given globally traceable transformation t, the recovery map R_t is not necessarily unique. However, the proof of Lemma 1 shows that one can consistently choose a family $\{R_t\}_{t\in G}$ such that the mapping $t \mapsto R_t$ is a group anti-homomorphism (i.e., $R_{t_2 \circ t_1} = R_{t_1} \circ R_{t_2}$). Consequently, the selected family forms a group that is anti-isomorphic to the original traceable group G.

The existence of a globally traceable group $G < \mathcal{G}$ (Lemma 1) provides a principled basis for our framework, motivating the use of a group of transformations rather than a single fixed

map. Another core requirement of disclosure control is that a chosen transformation t should distort the AI's output so that $A\{t(X)\}$ reveals no direct information about the true target $A\{X\}$. To enforce this in practice, the specific transformation t must be kept secret from the AI provider.

For a robust security analysis, however, we must consider a stronger adversarial setting: the AI provider may know the group G from which t is drawn, even though it does not know the particular realization. In this scenario, the AI provider could, in principle, compute the set of all possible values $t^{-1}(X)$ from the input dataset X by enumerating over G. This leads to the notion of the *disclosure space* associated with G and the input dataset X:

$$\Theta_{GX} := \{\, \mathbb{E}[\, A\{t^{-1}(X)\}\,] \mid t \in G \,\},$$

provided the expectations exist. Here, the expectation is taken over the internal randomness of the stochastic AI operator A. Intuitively, Θ_{GX} represents the set of all candidate values for the true statistic that are consistent with the knowledge of G. To ensure that every candidate value of the true statistic remains possible, this set must be as large as possible—ideally, it should coincide with the entire decision space Θ. This requirement is formalized by the concept of a *globally transitive group*.

Definition 2 (Globally transitive group)**.** Let $G \subset \mathrm{Homeo}(\mathcal{X})$ be a group. We say G is *globally transitive* if for every input dataset $X \in \mathcal{X}$, the corresponding disclosure space equals the full decision space; that is, $\Theta_{GX} = \Theta$.

A globally transitive group guarantees that, from the perspective of an AI provider who knows G but not the selected t, every value of the target statistic remains possible: the disclosure space coincides with the entire space Θ. Consequently, the two algebraic properties introduced—*global traceability* and *global transitivity*—underpin the two requirements of Algorithm 1: traceability ensures the enterprise can recover the target statistic, while transitivity ensures that every candidate value of the target statistic remains possible for the AI provider. Once a group G satisfying both properties is identified, the enterprise can select a transformation t from G to execute Algorithm 1.

Having introduced the notions of global traceability and transitivity, the natural next step is to select and verify a group G that satisfies both properties. In theory, such a group would provide the strongest possible guarantees for Algorithm 1. In practice, however, these global conditions—requiring the defining equalities to hold for every dataset $X \in \mathcal{X}$—are often excessively stringent. Verifying them would necessitate an infeasible number of AI queries across the entire data space, incurring prohibitive computational and financial costs.

Crucially, the security objective in our setting is to protect the specific, confidential summary statistic $A\{X_0\}$ derived from the true dataset X_0. It is therefore sufficient—and far more practical—to require that the traceability and transitivity guarantees hold specifically at X_0. This motivates the following weaker, localized versions of our core definitions.

Definition 3 (Locally traceable group)**.** Let $G \subset \text{Homeo}(\mathcal{X})$ be a group. We say G is *locally*

traceable at X_0 if for every $t \in G$, there exists a recovery map $R_t \in \text{Homeo}(\Theta)$ such that

$$R_t\left(A\{t(X_0)\}\right) \stackrel{d}{=} A\{X_0\}.$$

An individual transformation $t \in G$ satisfying this condition is called *locally traceable at* X_0.

The equality in this definition is an idealization, used to state the property without reference to a sample size. The operational version that is actually implemented—which allows for the error inherent in testing with finitely many transformations and queries—is given in Definition 5.

Definition 4 (Locally transitive group)**.** Let $G \subset \text{Homeo}(\mathcal{X})$ be a group, and let $GX_0 := \{t(X_0) \mid t \in G\}$ denote the orbit of X_0 under G. We say G is *locally transitive at* X_0 if for every dataset X in this orbit, the corresponding disclosure space equals the full decision space; that is, $\Theta_{GX} = \Theta$ for all $X \in GX_0$.

The local transitivity property ensures that, for any transformed data input $X \in GX_0$ derived from X_0, every candidate value of the true statistic $A\{X_0\}$ remains possible for the AI provider. A key simplifying observation is that these local properties are invariant across the entire orbit of X_0. If the group is locally traceable and transitive at X_0, it automatically inherits these properties at any point in the orbit GX_0.

Lemma 2 (Invariance on the orbit)**.** *Let* $G \subset \text{Homeo}(\mathcal{X})$ *be a group and let* X_1, X_2 *be any two points in the orbit* GX_0*. Then:*

- *G is locally traceable at X_1 if and only if it is locally traceable at X_2.*
- *G is locally transitive at X_1 if and only if it is locally transitive at X_2.*
- *The disclosure spaces coincide:* $\Theta_{GX_1} = \Theta_{GX_2} = \Theta_{GX_0}$.

Lemma 2 significantly reduces the verification burden. To execute Algorithm 1 securely, it suffices to confirm that the chosen group G is both locally traceable and locally transitive specifically at the true dataset X_0. The enterprise can then simply select a transformation t from G, provide $t(X_0)$ to the external AI model, and later recover the target $A\{X_0\}$ using the corresponding map R_t. This local formulation preserves the core guarantees of the framework while remaining practically feasible.

4. Sampling on the Group

To securely execute Algorithm 1, we must verify that a chosen transformation group G is both locally traceable and locally transitive at the true dataset X_0. This verification also requires collecting statistical evidence from the external AI model without ever disclosing X_0 itself. In our setting, the dataset X_0 is fixed, while the randomness stems from the stochastic AI operator $A\{\cdot\}$. The central challenge is to design AI queries whose inputs depend on G and X_0 in a way that preserves the confidentiality of X_0.

Recall that local traceability (Definition 3) requires $R_t\left(A\{t(X_0)\}\right) \overset{d}{=} A\{X_0\}$ for all $t \in G$. Direct evaluation of this condition necessitates knowledge of $A\{X_0\}$, which would

transmit X_0 to the external AI model and thus leak the information of X_0. Fortunately, local traceability is equivalent to the following pairwise condition:

$$R_{t_1}\left(A\{t_1(X_0)\}\right) \stackrel{d}{=} R_{t_2}\left(A\{t_2(X_0)\}\right) \qquad \text{for all } t_1, t_2 \in G. \tag{4.1}$$

This reformulation eliminates direct reference to $A\{X_0\}$ (if the tested $t_1, t_2 \neq \mathrm{id}$) and allows us to test purely from transformed queries, thereby preserving disclosure control.

Since G may contain infinitely many elements, checking every pair (t_1, t_2) is infeasible. Instead, we adopt a statistical approach: we sample transformations from G according to a suitable probability measure and test the condition (4.1) on this sample. This necessitates endowing the group G with a sampling distribution. The natural candidate is the Haar measure, the unique (up to scaling) left-invariant measure on a locally compact topological group. We first ensure that our groups of interest admit such a measure. Without loss of generality, we may assume G is closed (otherwise, we take its closure, which remains a group).

Lemma 3 (Existence of Haar measure on transformation groups)**.** *Let* $\mathrm{Homeo}(\mathcal{X})$ *be a locally compact Hausdorff topological group, and let* G *be a closed subgroup of* $\mathrm{Homeo}(\mathcal{X})$*, equipped with the subspace topology inherited from the compact-open topology. Then* G *is a locally compact Hausdorff topological group and therefore admits a left-invariant Haar measure* $\tilde{\mu}$*.*

The left-invariance of $\tilde{\mu}$, meaning $\tilde{\mu}(tH) = \tilde{\mu}(H)$ for any measurable $H \subset G$ and any $t \in G$, ensures that the measure is "uniform" with respect to the group structure. We use

it only as a base. The measure from which we actually sample is a probability measure μ obtained by reweighting $\tilde{\mu}$ with a positive weight function $\omega : G \to \mathbb{R}_{>0}$ chosen by the enterprise,

$$d\mu(g) = \omega(g)\, d\tilde{\mu}(g), \qquad \int_G \omega(g)\, d\tilde{\mu}(g) = 1.$$

Since μ is not disclosed to the AI provider, the choice of ω does not enter its knowledge.

Example 3 (Haar measures for the prototypical groups)**.** We specify the topological structure and Haar measures for the three groups introduced in Example 2. In each case, $\mathcal{X} = \mathbb{R}^p$.

- **Scaling Group $G_{\mathbf{scale}} \cong (\mathbb{R}_{>0}, \times)$:** As a one-dimensional Lie group, its Haar measure is $d\tilde{\mu}(\lambda) = d\lambda/\lambda$. To obtain a probability measure that decays exponentially both near zero and infinity, we choose the weight function $\omega(\lambda) = C^{-1}\exp(-\lambda - 1/\lambda)$, where $C = \int_0^\infty \exp\left(-\lambda - \frac{1}{\lambda}\right)\frac{d\lambda}{\lambda} \approx 0.2278$ is the normalizing constant, so that $\int_G \omega(\lambda)\, d\tilde{\mu}(\lambda) = 1$. The probability measure is then given by

$$d\mu(\lambda) = \omega(\lambda)\, d\tilde{\mu}(\lambda) = \frac{1}{C}\exp\left(-\lambda - \frac{1}{\lambda}\right)\frac{d\lambda}{\lambda}, \quad \lambda > 0.$$

- **General Linear Group $G_{\mathbf{GL}} \cong \mathrm{GL}(p, \mathbb{R})$:** This is a non-compact Lie group. Its Haar measure can be expressed as $d\tilde{\mu}(P) = |\det(P)|^{-n}\, dP$, where dP denotes the Lebesgue measure on the p^2 matrix entries. Since $\tilde{\mu}(G_{\mathrm{GL}}) = \infty$, a probability measure must be constructed via a weight function, e.g., using a matrix-variate distribution that decays sufficiently fast (such as a Gaussian-like weight).

- **Rotation Group $G_{\mathbf{SO}} \cong \mathrm{SO}(p)$:** This is a compact Lie group. Its Haar measure $\tilde{\mu}$ is finite and can be normalized to yield the uniform probability distribution over rotations.

  - For $n = 2$, the group $\mathrm{SO}(2)$ can be parameterized by an angle $\theta \in [0, 2\pi)$. The Haar measure is $d\tilde{\mu}(\theta) = d\theta$, and the corresponding uniform probability measure is $d\mu(\theta) = \frac{1}{2\pi}d\theta$.

– For $n \geq 3$, the Haar measure is the unique rotation-invariant distribution on the special orthogonal group.

Equipped with a probability measure μ on G, we can now design the sampling procedure. We draw n i.i.d. transformations from μ, denoted by $\mathcal{T} := \{T_i\}_{i=1}^n$. For each sampled transformation, we query the AI model m times independently, meaning each input $T_i(X_0)$ is submitted m times. The AI provider returns the corresponding stochastic outputs

$$\mathcal{A} := \{A_{ij}\{T_i(X_0)\}\}_{1\leq i\leq n,\, 1\leq j\leq m}.$$

Conditioned on the input $T_i(X_0)$, the outputs $\{A_{ij}\{T_i(X_0)\}\}_{j=1}^m$ are assumed i.i.d. realizations of $A\{T_i(X_0)\}$.

Finally, we apply the corresponding recovery maps to obtain the processed data

$$\mathcal{Y} := \{Y_{ij}\}_{1\leq i\leq n,\, 1\leq j\leq m} := \{R_{T_i}\left(A_{ij}\{T_i(X_0)\}\right)\}_{1\leq i\leq n,\, 1\leq j\leq m}. \tag{4.2}$$

This dataset $\mathcal{T}, \mathcal{A}, \mathcal{Y}$ will be used to construct hypothesis tests for local traceability and local transitivity, as detailed in the following sections.

5. Hypothesis Testing for Local Traceability

We now develop a hypothesis testing procedure for local traceability based on the collected data $\mathcal{Y}$ in (4.2). Conditional on the sampled transformations $\mathcal{T}$, the processed data $\mathcal{Y}$ consists of n independent groups: $\mathcal{Y} := \bigcup_{i=1}^n \{Y_{ij}\}_{j=1}^m$, where each group $\{Y_{ij}\}_{j=1}^m$ contains m i.i.d. realizations from the distribution of $R_{T_i}\big(A\{T_i(X_0)\}\big)$.

Local traceability (4.1) requires that all these distributions are identical, i.e.,

$$R_{T_{i1}}\big(A\{T_{i1}(X_0)\}\big) \stackrel{d}{=} R_{T_{i2}}\big(A\{T_{i2}(X_0)\}\big) \qquad \text{for all } T_{i1}, T_{i2} \in G.$$

Verifying this condition reduces to a multi-sample distribution equality test. We may employ any suitable non-parametric test for this purpose, such as those based on maximum mean discrepancy (Rizzo & Székely 2010, Zhang & Smaga 2022), graph-based methods (Chen & Friedman 2017), energy distances (Székely & Rizzo 2013), permutation tests (Arboretti et al. 2025), or extensions of Kolmogorov–Smirnov and Anderson–Darling tests to multiple samples (Finner & Gontscharuk 2018, Scholz & Stephens 1987). The specific test can be chosen according to the data's dimensionality, sample size, and computational resources.

Formally, let $S(\mathcal{Y}) \in \{0, 1\}$ denote the test statistic of the chosen multi-sample test, where $S(\mathcal{Y}) = 1$ indicates rejection of the null hypothesis that all groups share the same distribution. We assume this test satisfies standard consistency properties.

Assumption 1 (Properties of the multi-sample test)**.** Consider the multi-sample testing problem:

$$H_0^{\text{multi-sample}} : \text{All groups } \{Y_{ij}\}_{j=1}^{m} \text{ have the same distribution,}$$

$$H_1^{\text{multi-sample}} : \text{Not all groups share the same distribution.}$$

Let S be a test statistic such that:

1. Size control: Under $H_0^{\text{multi-sample}}$, $\mathbb{P}\big(S(\mathcal{Y}) = 1 \mid \mathcal{T}\big) \leq \alpha$.

2. Power: Under $H_1^{\text{multi-sample}}$, for any fixed configuration of distinct distributions among the groups, $\mathbb{P}\big(S(\mathcal{Y}) = 1 \mid \mathcal{T}\big) \to 1 \quad \text{as } m \to \infty$.

Using such a test, we can construct a hypothesis test for local traceability itself. Let G be the candidate transformation group. The null hypothesis is that G is locally traceable at X_0. The alternative hypothesis must accommodate the possibility that the map $t \mapsto R_t\big(A\{t(X_0)\}\big)$ may not be continuous in t. We therefore formulate the alternative as the existence of two open sets in G over which the distributions differ.

H_0: G is locally traceable at X_0.

H_1: There exist non-empty open sets $U_1, U_2 \subset G$ such that,

for any $t_1 \in U_1$ and $t_2 \in U_2$,

$$R_{t_1}\left(A\{t_1(X_0)\}\right) \overset{d}{\neq} R_{t_2}\left(A\{t_2(X_0)\}\right),$$

where $\overset{d}{\neq}$ denotes that the two random variables do not follow the same distribution.

If the mapping $t \mapsto R_t\big(A\{t(X_0)\}\big)$ is continuous in distribution with respect to t, then H_1 simplifies to the existence of some pair $t_1, t_2 \in G$ with different distributions.

The following theorem establishes that the multi-sample test S, applied to the processed data $\mathcal{Y}$, yields a valid test for local traceability.

Theorem 1 (Testing local traceability)**.** *Suppose Assumption 1 holds. Then the test based on $S(\mathcal{Y})$ satisfies:*

1. *Size control under H_0: $\mathbb{P}_{H_0}\big(S(\mathcal{Y}) = 1\big) \leq \alpha$.*

2. *Power under* H_1*:* $\mathbb{P}_{H_1}\big(S(\mathcal{Y}) = 1\big) \to 1$ *as* $m, n \to \infty$.

Theorem 1 shows that, by leveraging the Haar measure for random sampling, a standard multi-sample distribution test can be directly employed to verify local traceability within our framework. This approach allows the enterprise to test the key recoverability property without ever exposing the original dataset X_0, thereby maintaining strict disclosure control throughout the verification process.

It is important to clarify the practical interpretation of the hypothesis test for local traceability. We do not prove that the null hypothesis holds; rather, we adopt a group G in Algorithm 1 if there is insufficient evidence to reject local traceability. This approach is motivated by the nature of modern AI models—highly nonlinear, stochastic, and rarely satisfying exact algebraic invariants. Our test thus acts as a statistical screening mechanism: if the recovered outputs $\{R_{T_i}(A\{T_i(X_0)\})\}$ show no detectable distributional differences across a random sample from G, we regard G as operationally admissible for disclosure control. The following definition makes this notion precise.

Definition 5 (Admissibility)**.** A group G is called *admissible at level* α for a given sampling measure μ and pair (n, m) if the test does not reject local traceability at X_0 at level α.

Non-rejection does not assert that the distributions are equal: it asserts that no deviation larger than the smallest effect detectable at (n, m) was found. This reflects the standard statistical logic that a hypothesis test can rule out violations, but cannot certify absolute truth.

Remark 3 (Sampling error and the choice of (n, m))**.** The test can fail to detect a violation for two reasons. The first is sampling: if the set of transformations at which local traceability fails has mass q under μ, then all n draws miss it with probability $(1-q)^n$, so n is set from the smallest mass q that is deemed material and the tolerated probability δ of missing it. The second is power: even when such a transformation is drawn, the discrepancy must be separated from noise with the m repeated queries, so m is set from the smallest discrepancy that is to be detected.

6. Check on Local transitivity

Having established a statistical test for local traceability, we now address the complementary problem of testing local transitivity—that is, whether a locally traceable group G is also locally transitive at the true dataset X_0. The test will be based on the same random samples $\mathcal{T}$ and $\mathcal{A}$ introduced in Section 4.

The starting point is the algebraic characterization given by Lemma 2, which states that verifying local transitivity at X_0 is equivalent to checking whether the disclosure space $\Theta_{GX_0} = \{\mathbb{E}[A\{t^{-1}(X_0)\}]\}_{t\in G}$ coincides with the whole decision space Θ.

A naive way to estimate Θ_{GX_0} would be to estimate the expected values $\mathbb{E}[A\{t(X_0)\}]$ at the finitely many sampled transformations $t \in \mathcal{T} \subset G$ and then extrapolate to the entire group under some continuity assumption. Such an extrapolation, however, is inherently sensitive to the chosen regularity conditions and does not provide a reliable basis for deciding

whether Θ_{GX_0} actually covers Θ.

We therefore propose a different strategy that exploits the group structure together with the local traceability of G. Because G is locally traceable, for any X in the orbit GX_0 and any $t \in G$ we have the distributional equality $A\{t^{-1}(X)\} \overset{d}{=} R_t(A\{X\})$. Consequently, for each fixed $X \in GX_0$ the disclosure space can be rewritten as

$$\Theta_{GX} = \big\{\mathbb{E}\big[A\{t^{-1}(X)\}\big]\big\}_{t\in G} = \big\{\mathbb{E}\big[R_t(A\{X\})\big]\big\}_{t\in G}.$$

This reformulation suggests a natural estimator. For each sampled transformation T_i we set $X_i = T_i(X_0) \in GX_0$. Using the corresponding AI outputs $\{A_{ij}\{X_i\}\}_{j=1}^m$, we define

$$\widehat{\Theta}_{GX_i} := \Big\{\frac{1}{m}\sum_{j=1}^{m} R_t\big(A_{ij}\{X_i\}\big) \;\Big|\; t \in G\Big\}.$$

Notice that the index t runs over the whole group G, not merely the finite sample $\mathcal{T}$. By Lemma 2 the disclosure spaces Θ_{GX_i} are identical for all $X_i \in GX_0$; hence each $\widehat{\Theta}_{GX_i}$ is an estimator of the same target set Θ_{GX_0}.

To improve the accuracy of the disclosure-space approximation, we pool the data from all sampled points $X_i = T_i(X_0)$ and define the pooled estimator

$$\widehat{\Theta}_{GX_0} := \bigcup_{i=1}^{n} \widehat{\Theta}_{GX_i} = \bigcup_{i=1}^{n} \Big\{\frac{1}{m}\sum_{j=1}^{m} R_t\big(A_{ij}\{T_i(X_0)\}\big) \;\Big|\; t \in G\Big\}.$$

This estimator leverages the full set of nm AI queries, yielding a more comprehensive picture of the disclosure space. When the group G is sufficiently rich, the union over $t \in G$ in each $\widehat{\Theta}_{GX_i}$—and consequently the pooled union $\widehat{\Theta}_{GX_0}$—can readily cover the entire decision space

Θ. Hence, a practical criterion for local transitivity at X_0 is simply that $\widehat{\Theta}_{GX_0} = \Theta$. Example 4 illustrates how this equality is easily satisfied for the three prototypical transformation groups, which is attributed to the previously chosen perspective of transformation groups, rather than any single transformation. Because the condition $\widehat{\Theta}_{GX_0} = \Theta$ can be checked directly from the constructed set, there is no need to devise a standard Neyman–Pearson hypothesis test; verifying the equality itself suffices to conclude that the locally traceable group G is also locally transitive.

Example 4 (Empirical disclosure spaces for the prototypical groups)**.** In practice, the disclosure space is assessed through its empirical estimator $\widehat{\Theta}_{GX_0}$. For each of the three groups introduced in Example 2, we specify the decision space Θ, give the explicit form of $\widehat{\Theta}_{GX_0}$, and explain why this estimator can attain the whole set Θ.

- **Scaling Group $G_{\mathbf{scale}} \cong (\mathbb{R}_{>0}, \times)$:** Here the enterprise is assumed to need a positive-valued forecast (e.g., sales volume, production quantity). A natural choice is therefore $\Theta = \mathbb{R}_{>0}$, representing the set of all possible positive predictions. For sampled scalars $\lambda_i > 0$ $(i = 1, \ldots, n)$, the estimator becomes

$$\widehat{\Theta}_{GX_0} = \bigcup_{i=1}^{n} \Big\{ \frac{1}{\lambda} \cdot \frac{1}{m} \sum_{j=1}^{m} A_{ij}\{\lambda_i X_0\} \;\Big|\; \lambda > 0 \Big\}.$$

  Because the map $\lambda \mapsto \frac{1}{\lambda}$ is surjective onto $\mathbb{R}_{>0}$, and the sample averages $\frac{1}{m}\sum_{j=1}^{m} A_{ij}\{\lambda_i X_0\}$ are positive, the union over i and over all $\lambda > 0$ readily covers the entire positive real line. Hence $\widehat{\Theta}_{GX_0} = \mathbb{R}_{>0} = \Theta$.

- **General Linear Group $G_{\mathbf{GL}} \cong \mathrm{GL}(p, \mathbb{R})$:** In this setting the enterprise is assumed to require a vector-valued prediction that is inherently non-zero (e.g., a non-zero displacement, velocity, or gradient). A natural choice is therefore $\Theta = \mathbb{R}^p \backslash \{0\}$, the set of all possible non-zero vector predictions.

Let $P_i \in \mathrm{GL}(p,\mathbb{R})$ $(i = 1,\ldots,n)$ be the sampled matrices. Then

$$\widehat{\Theta}_{GX_0} = \bigcup_{i=1}^{n} \Big\{ P^{-1} \cdot \frac{1}{m} \sum_{j=1}^{m} A_{ij}\{P_i X_0\} \;\Big|\; P \in \mathrm{GL}(p,\mathbb{R}) \Big\}.$$

Since the map $P \mapsto P^{-1}$ is surjective onto $\mathrm{GL}(p,\mathbb{R})$, and $\mathrm{GL}(p,\mathbb{R})$ acts transitively on $\mathbb{R}^p \setminus \{0\}$, the set $\{P^{-1}v_i : P \in \mathrm{GL}(p,\mathbb{R})\}$ for any fixed non-zero $v_i \in \mathbb{R}^p$ already covers $\mathbb{R}^p \setminus \{0\}$. Taking the union over several samples $v_i = \frac{1}{m}\sum_j A_{ij}\{P_i X_0\}$ makes $\widehat{\Theta}_{GX_0}$ essentially fill the whole space $\Theta = \mathbb{R}^p \setminus \{0\}$ if at least one non-zero v_i exists.

- **Rotation Group $G_{\mathbf{SO}} \cong \mathrm{SO}(p)$:** In this case the enterprise is assumed to be concerned only with directional (angular) information contained in the dataset. A natural choice is to take $\Theta = GX_0$, the orbit of X_0 under rotations, which represents exactly the set of all rotated versions of the original data. With sampled rotations $P_i \in \mathrm{SO}(p)$, the estimator is

$$\widehat{\Theta}_{GX_0} = \bigcup_{i=1}^{n} \Big\{ P^{\top} \cdot \frac{1}{m} \sum_{j=1}^{m} A_{ij}\{P_i X_0\} \;\Big|\; P \in \mathrm{SO}(p) \Big\}.$$

Because the map $P \mapsto P^{\top}$ is a bijection on $\mathrm{SO}(p)$, one set $\{P^{\top}v_i : P \in \mathrm{SO}(p)\}$ coincides with GX_0 if that $v_i = \frac{1}{m}\sum_j A_{ij}\{P_i X_0\}$ still lies in GX_0. The union over multiple samples therefore again equals GX_0, i.e. $\widehat{\Theta}_{GX_0} = \Theta$. This reflects the security intention: by choosing $G = \mathrm{SO}(p)$, the enterprise signals that it only wishes to protect the angular information (the "orientation" of the data), and the estimator $\widehat{\Theta}_{GX_0}$ indeed spans exactly that protected orbit.

In all three cases, the empirical disclosure space $\widehat{\Theta}_{GX_0}$ constructed via the sampled transformations and the corresponding recovery maps naturally covers the whole decision space Θ that is relevant for the chosen group. This provides a practical demonstration that, for these common transformation groups, the local transitivity requirement is often easily satisfied in practice.

7. Example: A Pottery Company

This section offers a comprehensive description of the proposed method, encompassing the hypothesis testing of local traceability and check on local transitivity, along with the implementation details of Algorithm 1.

7.1 Problem and confidentiality targets

A handicraft company specializing in traditional ceramic art has marketed a pottery collection since 2016, targeting premium artisan segments. The firm's data team has compiled annual sales data alongside relevant product attributes spanning 2016–2025, X_0, as detailed in the Supplement. The company seeks to forecast the global sales volume $\widehat{Y}$ for 2026. Management has two forecasting needs: (1) the point forecast $\widehat{Y}$ of the 2026 sales volume, and (2) a floor that the volume is expected to exceed with probability 0.8, that is, the 20% quantile $Q_{0.2}$ of the predictive distribution. The second target is typical of production planning, where capacity is committed against a pessimistic scenario rather than a central one.

Planned managerial inputs for 2026 include maintaining a unit price of \$269, allocating a marketing budget of \$85,000, anticipating approximately 37 competing artisan brands, targeting a customer quality rating of 4.3, introducing a new pottery diameter of 6.7 inches, and positioning a decorative element at the design coordinates (0.2, 1.2). Correspondingly, the feature vector for prediction is $X_{11} = (2026,\ 269,\ 85,\ 37,\ 4.3,\ 6.7,\ 0.2,\ 1.2)$.

Were there no statistical disclosure requirement, the company would transmit the dataset

Statistica Sinica: Newly accepted Paper

X_0 and an associated query to an external AI, expecting to obtain the forecast of the task. In practice, management intends to safeguard specific confidential information concerning the 2026 collection. First, the reported forecast for 2026, that is, $\widehat{Y}$ or $Q_{0.2}$, must remain confidential due to its strategic importance. Second, the design coordinates $(x, y) = (0.2,\ 1.2)$ for the 2026 collection require protection, as they embody proprietary design patterns vulnerable to competitor imitation.

Remark 4. Since the objective of the proposed method is to ensure confidentiality of all AI-generated outputs, the protected information may include not only the primary target but also other summary statistics requiring disclosure control. In this scenario, the 2026 sales forecast and the 2026 design coordinates constitute sensitive business intelligence.

7.2 Choice of the transformation group

To apply the proposed statistical disclosure control method, we need to select a suitable locally traceable group G of transformations on the dataset space $\mathcal{X} = \mathbb{R}^{10\times 9+8}$. Based on the structure of the sensitive data (the 2026 sales forecast and design coordinates), we consider transformations that act on the entire dataset but preserve the relationships needed for accurate prediction:

- **Uniform scaling on price, marketing expenditure, and sales**: For columns 2 (historical and 2026), columns 3 (historical and 2026), and Y values (historical), we apply a positive scaling factor λ. That is, we multiply all values in these columns by

Statistica Sinica: Newly accepted Paper

the same factor λ.

- **Two-dimensional rotation on all design coordinates**: For all design coordinates in columns 7 and 8 (both historical and 2026), we apply rotations in $\mathbb{R}^2$. For any rotation angle $\theta \in [0, 2\pi)$, we define a transformation that rotates all (x, y) coordinates by θ while leaving other columns unchanged.

The combined transformation group is thus $G = G_{\text{scale}} \times G_{\text{SO}} \cong \mathbb{R}_{>0} \times \text{SO}(2)$. This group structure offers several advantages. Relationships are preserved because rotations maintain Euclidean distances and angles between design points—a geometric property important for many AI models—while uniform scaling preserves the ratios among financial and sales variables. Simple recovery maps are another benefit: the inverse transformation is straightforward, involving only a rotation by $(-\theta)$ applied to the design coordinates and a division of the returned target by the scaling factor λ. Consequently, the decision space is the same for either task. Written in terms of its elements, it is

$$\begin{aligned}\Theta &= \Big\{(z, \hat{x}, \hat{y}) : \ z \in \mathbb{R}_{>0}, \ (\hat{x}, \hat{y}) = R_\theta(0.2,\ 1.2) \text{ for some } \theta \in [0, 2\pi)\Big\} \\ &= \mathbb{R}_{>0} \times rS^1,\end{aligned}$$

where z denotes the target of the task at hand, that is, $z = \widehat{Y}$ for the point forecast and $z = Q_{0.2}$ for the quantile; R_θ is the rotation matrix of angle θ, $r = \sqrt{0.2^2 + 1.2^2} = \sqrt{37}/5 \approx 1.217$, and $rS^1 = G_{\text{SO}}\big((0.2, 1.2)\big) = \{R_\theta(0.2, 1.2)\}$ is the circle of radius r traced by the design coordinates under rotations. Every element of Θ is thus a triple $(z, \hat{x}, \hat{y})$ of the type

Statistica Sinica: Newly accepted Paper

returned by the AI.

7.3 Testing local traceability

Now we proceed to test whether the candidate group G is locally traceable according to the methodology outlined in Section 5. We employ five models of the current generation: Gemini 3.8 Flash, GPT-6.1 Terra, DeepSeek V4.1 Flash, Qwen 3.7 Flash and GLM-5.3 FlashX, with the temperature fixed at its default value 1. Access to these models is available either directly through their official APIs or via third-party aggregation platforms if required. Herein, we do not prescribe a specific algorithm but allow the AI to select the prediction methodology based on its assessment of the commercial context.

The query submitted to each AI model is structured according to the prompts given in Section S3 of the Supplement, where the placeholder "transformed data" is replaced with the corresponding entry from the transformed dataset $T_i(X_0)$; the only difference between the two tasks is the description of the target quantity. Crucially, neither the specific transformation T_i, nor the original dataset X_0, nor the sampling measure μ is disclosed to the AI model.

The evaluation is a 2×2 design that crosses the two forecasting tasks with two sampling measures for the scaling factor. The tasks are the point forecast $\widehat{Y}$ and the 20% quantile $Q_{0.2}$; the measures are the concentrated weight function of Example 3 and a diffuse alternative, the log-uniform measure on $[0.05, 20]$, which assigns comparable mass to small and to large scalings. In each of the four combinations we draw $n = 10$ transformations from μ

Statistica Sinica: Newly accepted Paper

and issue $m = 20$ independent queries per transformation. Local traceability is tested with the distance-based permutation test using the pseudo-F statistic (PERMANOVA-type) with $B = 999$ permutations. The statistic and the exact form of the p-value are given in Section S1.1 of the Supplement.

Table 1: Results in the pottery example: p-values and recovered values of the protected statistic (mean $\pm$ s.d.). The recommended combination for each task is in bold. Here conc. and log-unif. denote the concentrated weight function of Example 3 and the log-uniform measure on $[0.05, 20]$, and F, T and FX abbreviate the model suffixes Flash, Terra and FlashX.

Point forecast $\widehat{Y}$				Quantile $Q_{0.2}$			
Measure	Model	p-value	Recovered	Measure	Model	p-value	Recovered
conc.	Gemini 3.8F	0.002	3782.6 ± 109.7	conc.	Gemini 3.8F	0.106	3799.7 ± 62.6
	GPT-6.1T	**0.798**	$\mathbf{3893.5 \pm 139.3}$		GPT-6.1T	0.001	3694.8 ± 152.9
	DeepSeek V4.1F	0.037	3761.2 ± 141.6		DeepSeek V4.1F	0.378	3721.2 ± 161.2
	Qwen 3.7F	0.142	3782.3 ± 173.2		Qwen 3.7F	0.112	3617.7 ± 453.7
	GLM-5.3FX	0.428	3946.6 ± 409.0		**GLM-5.3FX**	**0.575**	$\mathbf{3625.0 \pm 294.9}$
log-unif.	Gemini 3.8F	0.099	3755.1 ± 119.0	log-unif.	Gemini 3.8F	0.006	3795.2 ± 61.4
	GPT-6.1T	0.001	3868.9 ± 118.8		GPT-6.1T	0.005	3715.0 ± 165.1
	DeepSeek V4.1F	0.018	3763.5 ± 130.7		DeepSeek V4.1F	0.139	3692.1 ± 162.7
	Qwen 3.7F	0.038	3802.6 ± 166.2		Qwen 3.7F	0.002	3495.5 ± 521.4
	GLM-5.3FX	0.715	3934.7 ± 241.8		GLM-5.3FX	0.002	3667.2 ± 235.5

Table 1 reports the outcome. At level $\alpha = 0.05$, ten of the twenty combinations of model and sampling measure are not rejected. Among the admissible combinations of each

Statistica Sinica: Newly accepted Paper

task, we suggest taking the one with the largest p-value. For the point forecast this selects GPT-6.1 Terra under the concentrated measure (p-value $= 0.798$); for the 20% quantile it selects GLM-5.3 FlashX under the concentrated measure (p-value $= 0.575$).

7.4 Checking local transitivity

To verify the local transitivity property, we analyze the disclosure space. Recall from Section 3 that $\Theta_{GX_0} = \{\mathbb{E}[A\{t^{-1}(X_0)\}] : t \in G\}$. Because the AI returns triples, each element of this set is explicitly

$$\Big(\mathbb{E}\big[z(t)\big],\ \mathbb{E}\big[\hat{x}(t)\big],\ \mathbb{E}\big[\hat{y}(t)\big]\Big), \qquad t \in G,$$

where $(z(t), \hat{x}(t), \hat{y}(t))$, with z as above, denotes the output obtained after applying the recovery map R_t. For the chosen transformation group $G = G_{\text{scale}} \times G_{\text{SO}}$, the estimated disclosure space at X_0 is: $\widehat{\Theta}_{GX_0} = \mathbb{R}_{>0} \times rS^1$. The equality $\Theta_{GX_0} = \Theta$ is immediately evident, confirming that G is also locally transitive at X_0.

For the two recommended combinations the test found no evidence against local traceability at level α for $(n, m) = (10, 20)$, and G is locally transitive at X_0, so that every value of the target statistic remains possible (Section 3). The company can implement Algorithm 1 with the recommended combination for each task, and the external AI cannot identify the reported statistics from the data it submits.

This case study illustrates how an enterprise can operationalize the proposed framework by strategically selecting transformation groups tailored to the specific types of sensitive

Statistica Sinica: Newly accepted Paper

information requiring protection. By applying distinct transformations to different data components—scale transformations to financial and sales variables, and rotations to geometric design coordinates—the company enforces a targeted privacy strategy. This allows it to leverage advanced, externally-hosted AI analytics while maintaining confidentiality over its key strategic statistics.

Supplementary Material

The supplementary material is organized as follows. Section S1 reports the simulations of the hypothesis testing framework, Section S2 collects the notations and mathematical preliminaries, Section S3 gives the dataset and the two prompts used in the example, and Section S4 contains the proofs of the lemmas and theorems.

Acknowledgements

Lingxuan Shao is funded by National Natural Science Foundation of China (12401340 and 12471279) and Shanghai Science and Technology Development Foundation (24ZR1405700). All programs and data used in the numerical studies are available at https://github.com/LingxuanShao/SDC.

Statistica Sinica: Newly accepted Paper

References

Arboretti, R., Barzizza, E., Biasetton, N. & Disegna, M. (2025), 'A review of multivariate permutation tests: Findings and trends', *Journal of Multivariate Analysis* **207**, 105421.

URL: *https://www.sciencedirect.com/science/article/pii/S0047259X25000168*

Arens, R. F. (1946), 'A topology for spaces of transformations', *Annals of Mathematics* **47**(3), 480–495.

URL: *http://www.jstor.org/stable/1969087*

Chapelle, R. & Falissard, B. (2023), 'Statistical properties and privacy guarantees of an original distance-based fully synthetic data generation method', *HAL preprint hal-04233955* .

Chen, H. & Friedman, J. H. (2017), 'A new graph-based two-sample test for multivariate and object data', *Journal of the American Statistical Association* **112**(517), 397–409.

Demelius, L. et al. (2023), 'Recent advances of differential privacy in centralized deep learning: A systematic survey', *arXiv preprint arXiv:2309.16398* .

Duncan, G. T. & Pearson, R. W. (1991), 'Enhancing access to microdata while protecting confidentiality: Prospects for the future', *Statistical Science* **6**(3), 219–239.

Elsharif, W., Alzubaidi, M., She, J. & Agus, M. (2025), 'Visualizing ambiguity: Analyzing linguistic ambiguity resolution in text-to-image models', *Computers* **14**(1), 19.

URL: *https://www.mdpi.com/2073-431X/14/1/19*

Finner, H. & Gontscharuk, V. (2018), 'Two-sample Kolmogorov–Smirnov-type tests revisited: Old and new tests in terms of local levels', *The Annals of Statistics* **46**(6A), 3014–3037.

Statistica Sinica: Newly accepted Paper

URL: *https://doi.org/10.1214/17-AOS1647*

Gong, Y. et al. (2023), 'Practical solutions in fully homomorphic encryption – a survey analyzing existing acceleration methods', *arXiv preprint arXiv:2303.10877* .

Green, E., Ritchie, F. & White, P. (2024), The statbarn: A new model for output statistical disclosure control, *in* 'Privacy in Statistical Databases: International Conference, PSD 2024, Antibes Juan-les-Pins, France, September 25–27, 2024, Proceedings', Vol. 14915 of *Lecture Notes in Computer Science*, Springer, Cham, pp. 284–293.

URL: *https://uwe-repository.worktribe.com/output/12790060*

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K. & de Wolf, P.-P. (2012), *Statistical Disclosure Control*, Wiley Series in Survey Methodology, John Wiley & Sons.

Li, X. et al. (2023), 'A survey of secure computation using trusted execution environments', *arXiv preprint arXiv:2302.12150* .

McCleary, J. (2006), *A First Course in Topology: Continuity and Dimension*, Vol. 31 of *Student Mathematical Library*, American Mathematical Society.

Muralidhar, K., Parsa, R. & Sarathy, R. (1999), 'A general additive data perturbation method for database security', *Management Science* **45**(10), 1399–1415.

Rizzo, M. L. & Székely, G. J. (2010), 'DISCO analysis: A nonparametric extension of analysis of variance', *The Annals of Applied Statistics* **4**(2), 1034–1055.

URL: *https://doi.org/10.1214/09-AOAS245*

Scholz, F. W. & Stephens, M. A. (1987), 'K-Sample Anderson–Darling Tests', *Journal of the American Statistical*

Statistica Sinica: Newly accepted Paper

Association **82**(399), 918–924.

URL: *https://doi.org/10.1080/01621459.1987.10478517*

Székely, G. J. & Rizzo, M. L. (2013), 'Energy statistics: A class of statistics based on distances', *Journal of Statistical Planning and Inference* **143**(8), 1249–1272.

URL: *https://www.sciencedirect.com/science/article/pii/S0378375813000633*

Yesuf, Y. & Fields, Z. (2025), 'Artificial intelligence adoption as a driver of innovation and competitiveness in SMEs: A bibliometric and systematic review', *F1000Research* **14**, 1187. PMID: 41477651, PMCID: PMC12749562.

URL: *https://doi.org/10.12688/f1000research.171494.1*

Zhang, J.-T. & Smaga, L. (2022), 'Two-sample test for equal distributions in separate metric space: New maximum mean discrepancy based approaches', *Electronic Journal of Statistics* **16**(2), 4090–4132.

URL: *https://doi.org/10.1214/22-EJS2033*

Department of Statistics and Data Science, School of Management, Fudan University

E-mail: shao_lingxuan@fudan.edu.cn