1. Introduction

The audiovisual turn in social and human sciences has produced a paradox. Research communities that depend on video and kinematic data — gesture researchers, conversation analysts, clinical scientists, educational psychologists — now have access to more computational tools than ever before, yet the relationship between those tools and the theoretical constructs that give their outputs scientific meaning remains poorly specified. A pose estimator returns joint coordinates; whether those coordinates correspond to a *gesture stroke*, a *grooming movement*, or measurement noise depends on an interpretive framework the estimator knows nothing about.

At the same time, the regulatory and ethical landscape for audiovisual data has shifted sharply. GDPR, institutional review requirements, and emerging AI Act obligations mean that video of identifiable people can rarely be shared, re-used, or deposited in public archives without transformation — yet the most common transformations (face blur, body pixelation) have analytic costs that are poorly understood and rarely documented.

The result is a fragmentation that affects infrastructure design at every level: which modalities to capture, how to store and process them, what privacy interventions are applied and when, and how outputs are made available for synthesis and re-use.

We propose the **utility triad** (AU, RU, GU) as a principled framework for diagnosing and designing against this fragmentation. Our contribution is primarily conceptual: we name three dimensions of utility that are often conflated or left implicit, show how they interact, and argue that responsible infrastructure must operationalise all three explicitly.

2. Background

2.1 *Data-driven and theory-driven research*

Maass et al. (2018) identify four challenges for information systems research in the era of big data: (1) reconciling data-driven and theory-driven knowledge creation; (2) selecting data and analytical techniques for theory testing; (3) solving problems that require cross-disciplinary collaboration; (4) sharing data and models across research teams. Their framework was developed primarily for text and tabular data; the multimodal audiovisual case extends each challenge in important ways.

Challenge 1 is especially acute for gesture and nonverbal behaviour research, where automated detection produces features (joint velocities, facial AUs) whose theoretical interpretation requires domain knowledge that the algorithm does not possess. Challenge 4 is compounded by privacy constraints that do not apply to most text or tabular datasets: video cannot simply be deposited — it must first be transformed, and the transformation affects the data's analytic utility.

2.2 *Privacy–utility trade-offs in video research*

The empirical study of masking effects on downstream analytic quality is sparse (Owoyele et al., 2026). MaskBench provides the first standardised benchmark: face-and-body blurring preserves 88–95% of pose estimation quality across estimators, while pixelation and solid fill show estimator-dependent degradation. These results make the privacy–utility trade-off empirically tractable for the first time, but they also show that the trade-off is not binary: different masking strategies have different costs for different analysis pipelines.

2.3 *Generative AI in research data workflows*

Large-scale generative models — diffusion models for video, LLMs for text, pose-conditioned generation for gesture — have introduced a new category of research affordance: the ability to *synthesize* data rather than merely *analyse* it. Synthetic corpus generation, model training on de-identified data, and cross-modal hallucination are all instances of this category. The governance literature for these use cases in SSH contexts is in its infancy.

3. The Utility Triad

We define the utility triad as three distinct dimensions of what researchers need from a multimodal data infrastructure:

AU (*Analytic Utility*). The signal extractable from a corpus using available analytic methods. Every pipeline decision — masking, segmentation, extraction, alignment — either preserves or reduces AU. Maximising AU is the primary research goal of any data collection effort, and infrastructure that does not foreground AU tradeoffs silently degrades the science it is meant to support.

RU (*Reflexive Utility*). The capacity to reason explicitly about what preprocessing and privacy-preserving transformations *cost* in AU. RU is not privacy compliance; it is the meta-analytic awareness that makes compliance decisions legible. A researcher with high RU can state: "face-blur preserves 91% of pose quality for MediaPipe but only 78% for ViTPose on this corpus" — and can make a principled choice between masking strategies on the basis of her research question, not merely her legal obligation. RU is particularly important for SSH early career researchers, who rarely have the computational background to assess masking costs independently.

GU (*Generative Utility*). The capacity to synthesize new data, representations, or research artifacts from existing corpora using generative AI. GU includes synthetic corpus generation for training and augmentation, style-conditioned synthesis for data enrichment, and the generation of pose- or speech-conditioned multimodal representations. GU is the most powerful tier of the triad and the most governance-demanding: without an operationalised RU layer, generative processes may bypass the privacy protections that make the original data ethically usable. At the risk frontier, GU encompasses deepfake production and re-identification attacks — failure modes that responsible infrastructure must anticipate and prevent.

3.1 *Triad geometry and interactions*

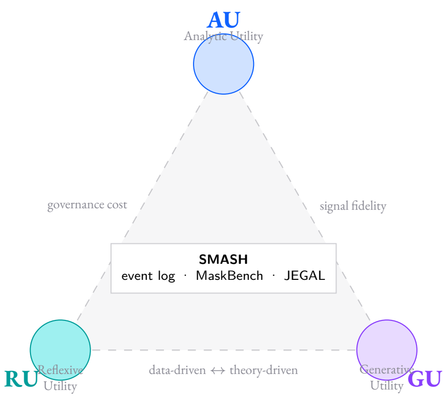

Figure 1: The utility triad. AU at the apex, RU and GU at the base vertices. SMASH instantiates all three.

The geometric arrangement reflects theoretical structure (Figure 1). AU is the apex: the primary research goal to which the other two vertices are instrumental. RU and GU form the base, because they are both means to AU and in irreducible tension with each other: increasing GU (synthesising more, more freely) exerts pressure on RU (making it harder to audit and constrain what was generated from what source data). The base edge is thus the governance frontier. The edge labels reflect three key tradeoffs: the left edge (AU–RU) represents the governance cost of achieving high AU while maintaining reflexive accountability; the right edge (AU–GU) represents the signal fidelity cost of introducing synthetic data into analysis pipelines; the base (RU–GU) represents the data-driven/theory-driven divide that Maass et al. (2018) identified as Challenge 1.

4. Instantiation: SMASH

The SMASH system (Pouw et al., 2026) instantiates the utility triad through three architectural elements:

AU: the canonical event log. Every module input and output — automated or human-annotated — is recorded as a typed, timestamped, provenanced event. This gives researchers complete visibility into what analytic operations have been applied to the corpus, and with what confidence.

RU: MaskBench. MaskBench provides standardised benchmarks for the privacy–utility tradeoff across masking strategies (face blur, body blur, pixelation, solid fill), pose estimators (MediaPipe, OpenPose, ViTPose), and corpus types. A researcher loading a MaskBench result into SMASH sees immediately what analytic cost her institution's preferred masking strategy imposes.

GU: JEGAL embedding substrate. JEGAL (Joint Embedding for Gesture and Language Alignment) provides a cross-modal representational space that supports both analysis (coupling scores, synchrony measures) and synthesis (generation conditioned on gesture or speech inputs). The provenance schema extends to synthetic events: every generated item is labelled `provenance: synthetic`, preserving the audit chain that RU depends on.

5. Research Agenda

The utility triad identifies specific open problems at each vertex:

AU agenda. Long-range cross-corpus alignment; estimator-independent feature representations; theory-aware codebook matching (McNeill gesture types onto kinematic features).

RU agenda. Formal re-identification risk metrics for multimodal data; accessible tooling for SSH ECRs without computational backgrounds; IRB-compatible audit log formats; FAIR metadata standards for masking operations.

GU agenda. Governance frameworks for synthetic corpus generation; deepfake detection benchmarks for SSH audiovisual data; differential privacy for generative fine-tuning; epistemic laundering audits for LLM-assisted annotation.

6. Discussion

The utility triad is a conceptual framework, not an evaluation rubric. Its primary contribution is to *name* three things that infrastructure designers and researchers treat as one — "data quality" — and to make their interdependence explicit.

A system that maximises AU without operationalising RU produces high-quality analyses on data that cannot be shared or audited. A system that enables GU without an RU layer produces synthetic data that cannot be traced to ethically handled sources. Neither is responsible infrastructure.

The triad also has normative implications for research community practice. Peer review of computational SSH work rarely asks: *what did de-identification cost this study's analytic quality? Were synthetic data items labelled as such in the analysis?* The utility triad provides vocabulary for making these questions standard.

7. Conclusion

We have proposed the utility triad — AU, RU, GU — as a framework for what researchers need from multimodal data infrastructure, and shown that SMASH instantiates all three tiers. The central claim is that AU, RU, and GU are not competing goals