

Kallia Kleisarchaki
Independent Researcher

When Ground Truth Never Arrives: Data Quality and Information-Set Integrity in Prospective Model Evaluation

Data quality is usually discussed in terms of accuracy, completeness, consistency and measurement error. In many real-world inference problems, however, a more fundamental difficulty arises: the variable of interest is latent, and ground truth never arrives.

This talk examines that problem through prospective modelling of Formula One competition. Rival strategic states and intentions are not directly observable, nor are they subsequently released as labelled data. A model must infer them from the information available to a public observer at the time. Later race events may provide evidence consistent or inconsistent with an earlier inference, but they do not retrospectively reveal the true value of the hidden state. Consequently, conventional evaluation against eventual ground-truth labels is impossible.

This creates two related data-quality problems. First, evaluation must distinguish between what was observable when an inference was made and what became observable only afterwards. Second, retrospective evidence must not be mistaken for ground truth.

Knowing the eventual race outcome can make one interpretation appear obvious after the fact even though several latent-state explanations were compatible with the original information set.

I explore an approach based on pre-specified and timestamped forecasts, preserved failures and abstentions, explicit regime boundaries, and evaluation against the information genuinely available at prediction time. Subsequent observations are treated as evidence about earlier latent states rather than as labels that reveal those states.

The broader argument is that provenance should capture not only where a datum came from, but when it became knowable, under which regime it was valid, and whether it represents an observation, an inference, or retrospective evidence. These distinctions become particularly important when evaluating statistical and AI systems operating under partial observability, where the quantities that matter most may never become directly observable.

The case study illustrates a broader challenge for real-world data science: how should we evaluate inference quality when the object being inferred remains hidden even after the event?