

Drift type

Semantic drift

What changes

The meaning of outputs or terms diverges from intent

Model drift

Prediction accuracy decays over time

Concept drift

The relationship between inputs and correct outputs changes

Data drift

Input data distribution shifts

Hallucination

The model fabricates content not supported by evidence

Typical detection signal

Embedding distance from reference, declining entailment scores

Accuracy, precision, or recall declining against labeled data

Performance drop on new data that was valid under old assumptions

Statistical tests (KS test, PSI) on feature distributions

Faithfulness scoring against source material

Example

A chatbot's definition of "premium customer" gradually shifts from the governed definition

A fraud model's false positive rate increases quarter over quarter

A pandemic shifts buying patterns, making pre-2020 training data misleading

A demographic shift in the customer base changes the age distribution of model inputs

A model generates a citation that does not exist