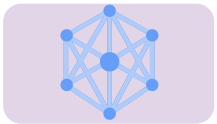
Step 1.
Construct knowledge graphs and Pre-train node embeddings.

Construct knowledge graph and initialise node embeddings



Train node embeddings in edge regression objective

The node embeddings is used to combine with *BERT

Step 2.
Collect consistent GOA and synthesise inconsistent GOA.

GOA with experiment type evidence code will be collected from databases

Apply heuristic rules to modify each consistent instance in simulation of typed inconsistency

OS	OB
IM	IG

Apply concept-recognition tools to extract biological concepts in evidence texts

PubTator

Collect GeneRIF statement indexed by PMID and Gene tuple

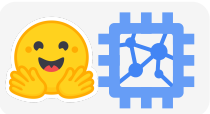
NIH

Human verify true simulation of real-world inconsistency. This data is used to train de-inconsistency model

Step 3.
Train classification model for detection of typed GOA inconsistencies.

A GOA instance is sampled from the synthetic dataset.

The corresponding node embedding is utilised to combine with *BERT.



The classifier calculate logit and classify inconsistency with maximum likelihood.

The suspicious records will be sent back to human for manual quality review.