

AI-powered technical lineage for Python integrations

AI feature description	Uses generative AI to translate Python code into column-level technical lineage for calculated columns and measures, with support for two data sources: Cloud Storage: Python integration via your cloud provider of choice. Git Repository: Python integration via your preferred Git host.
Feature optional?	Yes. These are optional integrations as opposed to an optional feature.
Model	Model Type: Google Vertex AI Gen AI?: Yes Owner: Google LLC
Intended use	These integrations are intended to take Python code from customer's configured data sources and translate it to readable technical lineage in the Collibra platform.
Model training/retraining data	Collibra relies on the training data leveraged by Google Vertex AI as shipped. For more information, please see the Google Vertex AI Product Documentation .
Model input	A piece of code (in case of plain Python scripts, the whole file) with summarized context from other files/pieces of code.
Model output	A JSON file describing technical lineage.
Is input data used to train the model for general product improvement purposes?	No
Data storage and retention	All Collibra storage locations are referenced here . Retention is concurrent with customer environment retention.
Human in the loop?	Yes. While this is partially automated, the user can delete any output if deemed incorrect or unwanted.
Is personal data processed?	Personal data is not intended to be included in the model input data, but if it is, its processing is subject to the customer's data privacy terms and conditions with Collibra under its Master Cloud Agreement. You can find our standard, customer facing privacy terms here .
Other considerations	Please note these are optional integrations that must be appropriately configured by the user. Data processed will be the raw data from the configured data sources, whether that is the cloud storage source or the git repository.