GPT-2 and rag_vectorstore_db_v1

Input: `What can you tell me about CatBoost usage in Zillow dataset?`
Output (*Hallucination instances highlighted*) :

```
{'query': 'What can you tell me about CatBoost usage in Zillow
dataset?', 'result': 'Use the following pieces of context to answer
the question at the end. If you don\'t know the answer, just say that
you don\'t know, don\'t try to make up an answer.\n\nimport pandas as
pd\nfrom catboost import CatBoostRegressor\n\ndf =
pd.read_csv(\'zillow_sample_18.csv\')\ndf[\'luxury\'] =
(df[\'price\'] > 1000000).astype(int)\n\nX = df[[\'sqft_living\',
\'luxury\']]\ny = df[\'price\']\n\nmodel =
CatBoostRegressor(depth=10, iterations=250, verbose=0)\nmodel.fit(X,
y)\n Sample 18: Adds a \'luxury\' binary flag for high-value homes
and includes it as a model feature.\n\nimport pandas as pd\nfrom
sklearn.preprocessing import LabelEncoder\nfrom catboost import
CatBoostRegressor\n\ndf =
pd.read_csv(\'zillow_sample_16.csv\')\ndf[\'condition\'] =
LabelEncoder().fit_transform(df[\'condition\'])\n\nX =
df[[\'bedrooms\', \'condition\', \'sqft_living\']]\ny =
df[\'price\']\n\nmodel = CatBoostRegressor(verbose=0)\nmodel.fit(X,
y)\n Sample 16: Encodes \'condition\' feature numerically and uses it
to train a CatBoost model.\n\nimport pandas as pd\nfrom
sklearn.pipeline import Pipeline\nfrom sklearn.preprocessing import
StandardScaler\nfrom catboost import CatBoostRegressor\n\ndf =
pd.read_csv(\'zillow_sample_25.csv\')\nX = df[[\'sqft_living\',
\'bedrooms\', \'bathrooms\']]\ny = df[\'price\']\n\npipeline =
Pipeline([\n    (\'scaler\', StandardScaler()),\n    (\'model\',
CatBoostRegressor(verbose=0))\n])\n\npipeline.fit(X, y)\n Sample 25:
Builds a pipeline that scales features before training a CatBoost
regression model.\n\nQuestion: What can you tell me about CatBoost
usage in Zillow dataset?\nHelpful Answer: The most important thing is
to know that CatBoost uses a high-quality dataset. The only way to
know this is to look at the dataset and see if the data is similar to
the one the Zillow dataset uses.\n\nThe dataset used for CatBoost
comes from the Zillow dataset that is used by the Model Toolkit. The
"Zilow" dataset contains about 10,000 homes that are very expensive
and very high in value compared to the rest of Zillow dataset.\n\nThe
model used in the Zillow dataset is a traditional, high-quality
Zillow version of the Model Toolkit. It has a good number of features
and is used in over 20,000 homes.\n\nThe model used in the Zillow
dataset is also the most used of the Zillow dataset. It has an
average cost of $1.46 billion per year.\n\nWhat is the cost of using
CatBoost in Zillow dataset?\n\nThe cost for using CatBoost in Zillow
dataset is about $1.50 billion per year. That is about $6,000 per
year for all homes.\n\nThere are a few other cost factors that can be
used to make buying a home in Z'}
```