

Reproduction Guide for Mockingbird

2024-11-25

Introduction

Mockingbird ("Robotless.Mocking") is a module of our robot software platform "Robotless". Robotless is still under development, and we plan to publish it soon. Mockingbird depends on an early version of Robotless (included in this supplemental material), please note that this version of Robotless does not imply its final performance.

Requirements

Operating Systems: Robotless platform and Mockingbird can work on systems that supported by [.NET9.0](#) (including Windows, Linux, Android, etc.).

Hardware: The platform itself has no specific requirements for hardware capabilities.

Software:

- An instance of [MongoDB](#) server. Version 7.0 is used in this research.
- API Token for [OpenAI Platform](#).

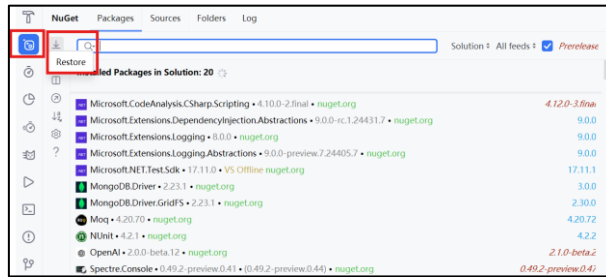
IDE: This platform is developed in C# with [JetBrains Rider](#). Other C# IDEs are also acceptable.

Procedures

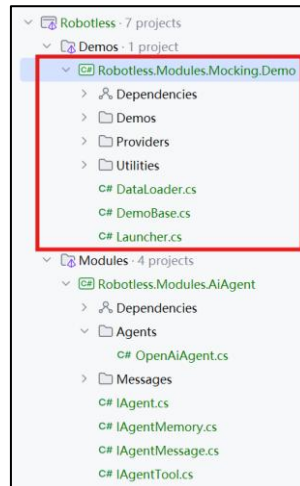
In this section, we will show you how to set up the environment and reproduce the experiment results. Additionally, we will also show you how to enable the source code live preview window for Substitution Script.

Reproduce Experiment Results

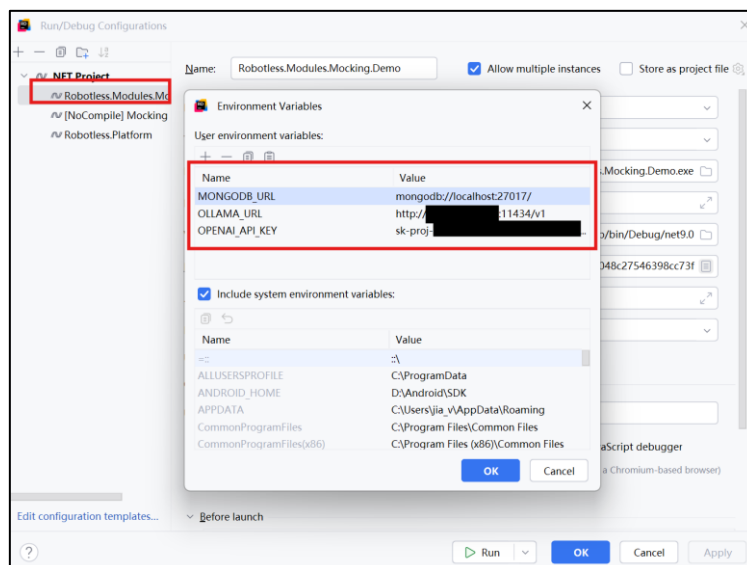
1. Install [.NET 9](#);
2. Setup a MongoDB server instance: either deploy a [community server](#) instance locally or set up a [MongoDB Atlas](#) instance.
3. Open solution "Robotless.sln" in an IDE ([JetBrains Rider](#) is used in this document).
4. Download dependencies by clicking "NuGet Restore" button:



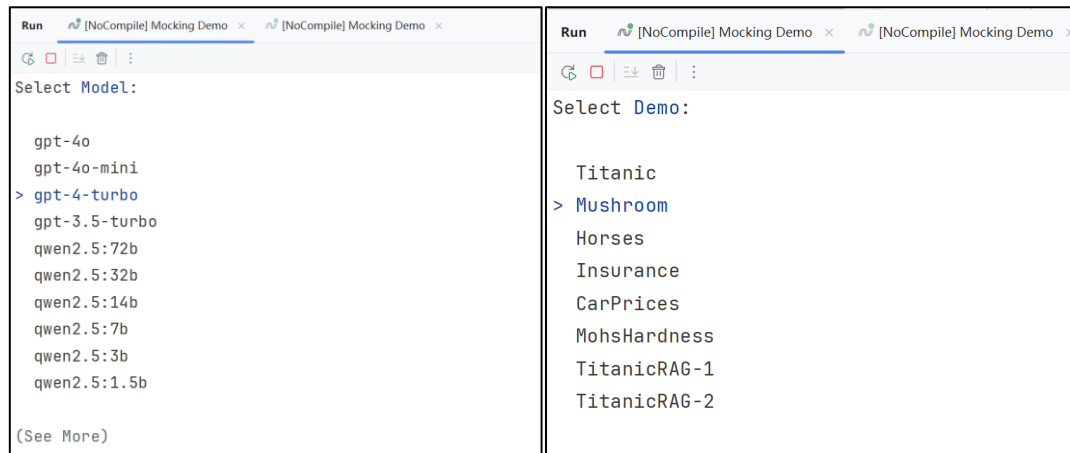
- Find project "Robotless.Platform.Mocking.Demo" under solution folder "Demos":



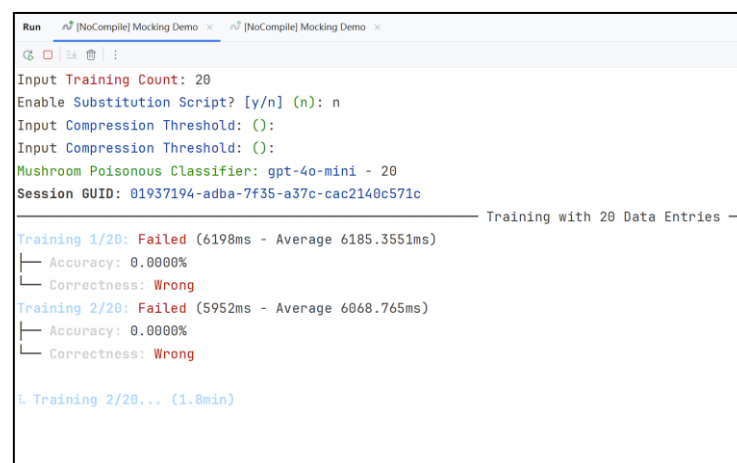
- Configure project environment variables in the red box in the following screenshot.
 "MONGODB_URL" is the connection URL for your MongoDB server instance.
 "OPENAI_API_KEY" is the API key acquired from [OpenAI platform](#), and it is necessary if you choose GPT series as the backend. "OLLAMA_URL" is the API URL for [Ollama](#), a convenient open-source project to host LLMs on your local machines.



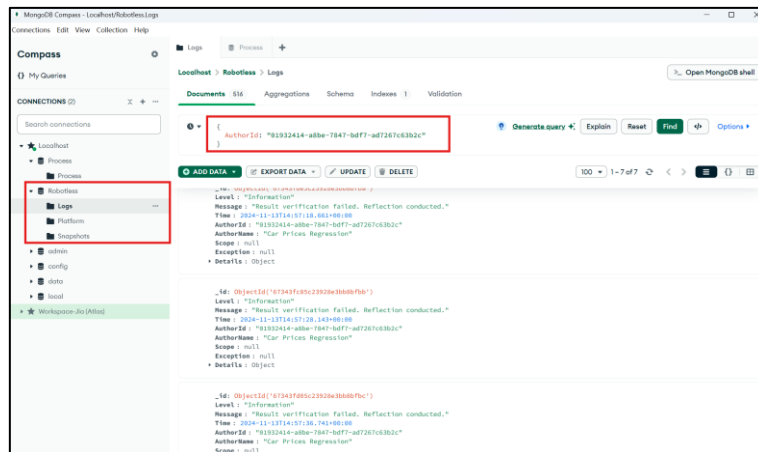
7. Run project "Robotless.Platform.Mocking.Demo" and then choose models and demos to run.



8. Fill in the configuration parameters such as the training count (context length), and then press enter to start this evaluation.



9. Retrieve operation logs from MongoDB and perform analysis. Open MongoDBCompass, and then use the session GUID showed in the console to filter operation logs for this session. Then you can view these logs in Compass or export them in a JSON file by clicking "Export Data" button.



10. Each log contains all necessary information for analysis, such as the input (“Arguments” field), output (“ActualResults” field) and answer (“ExpectedResults” field), reasoning (“Remarks” field) and evaluation metrics (“Metrics” field). When the training stage and the evaluation stage are finished, there will be summary logs for the final metrics.

```

_id: ObjectId('67343fc85c23928e3bb8bfbfb')
Level: "Information"
Message: "Result verification failed. Reflection conducted."
Time: 2024-11-13T14:57:28.143+00:00
AuthorId: "01932414-a8be-7847-bdf7-ad7267c63b2c"
AuthorName: "Car Prices Regression"
Scope: null
Exception: null
Details: Object
  Stage: Object
    !Type: "System.String"
    !Value: "Training"
  Index: Object
    !Type: "System.Int32"
    !Value: 4
  Arguments: Object
    !Type: "MongoDB.Bson.BsonDocument"
    !Value: Object
      brand: "Volvo"
      model: "XC90 Hybrid T8 R-Design"
      modelYear: 2018
      milage: "40,000 mi."
      fuelType: "Plug-In Hybrid"
      engine: "400.0HP 2.0L 4 Cylinder Engine Plug-In Electric/Gas"
      transmission: "Transmission w/Dual Shift Mode"
      exteriorColor: "Black"
      interiorColor: "Black"
      accident: "None reported"
      cleanTitle: true
  Remarks: Object
    !Type: "MongoDB.Bson.BsonString"
    !Value: "I previously given the wrong result 42000, but the correct answer is 4..."
  ActualResults: Object
    !Type: "MongoDB.Bson.BsonInt32"
    !Value: 42000
  ExpectedResults: Object
    !Type: "MongoDB.Bson.BsonInt32"
    !Value: 42000
  Reflection: Object
    !Type: "System.String"
    !Value: "It appears there was a misunderstanding or miscommunication as the est..."
  Metrics: Object
    !Type: "System.Collections.Generic.Dictionary`2[[System.String, System.Private..."
    !Value: Object
      ReflectionThreshold: "0.0000"
      StandardDeviation: "0.0000"
      RootMeanSquareError: "5931.9053"
      ErrorRatio: "0.00%"

```

Operation logs for the experiments in this research are included in this supplemental material.