AI Buddy (Visual Studio) Extension & Hosting DeepSeek LLM Locally

|  |  |
| --- | --- |
| Bert O’Neill | |
| Department: | Software Development |
| Document Title: | Visual Studio AI Prompt Integration |
| Author: | Bert O’Neill |
| Date: | 18-Feb-2025 |
| Version | 1.0 |
| SharePoint |  |
| Document Goals | Provide a user guide, on setting up Ollama, hosting a LLM (locally) and integrating the *AI Buddy* extension within Visual Studio. |

# Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 18/02/2025 | 1.0 | Initial draft | Bert O’Neill |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Introduction

This document provides details on how your company can setup a LLM Hoster locally, and connect the Visual Studio AI Buddy extension to it. For example, be able to prompt a DeepSeek LLM with the security of knowing that you’re prompt hasn’t left your company’s network.

## Purpose

This Document provides a comprehensive overview on how your company can keep its AI prompts secret and not leak any sensitive information, even when using a LLM model from DeepSeek.

## Scope

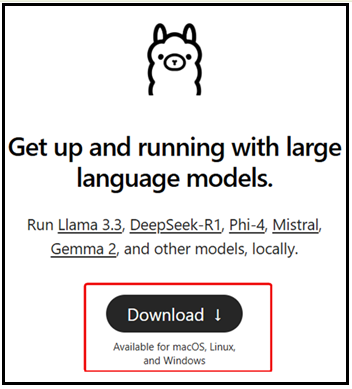
The scope of this document is to convey the steps involved in connecting to and using AI prompts from within Visual Studio to your local Hoster.

# Prerequisites

* Knowledge of a LLM Hoster (like Ollama)
* Know your Hoster API
* Know your Hoster’s API key
* Know the (Ollama) LLM model you wish to use (for e.g. DeepSeek-R1).

# LLM Hoster (Ollama) Setup

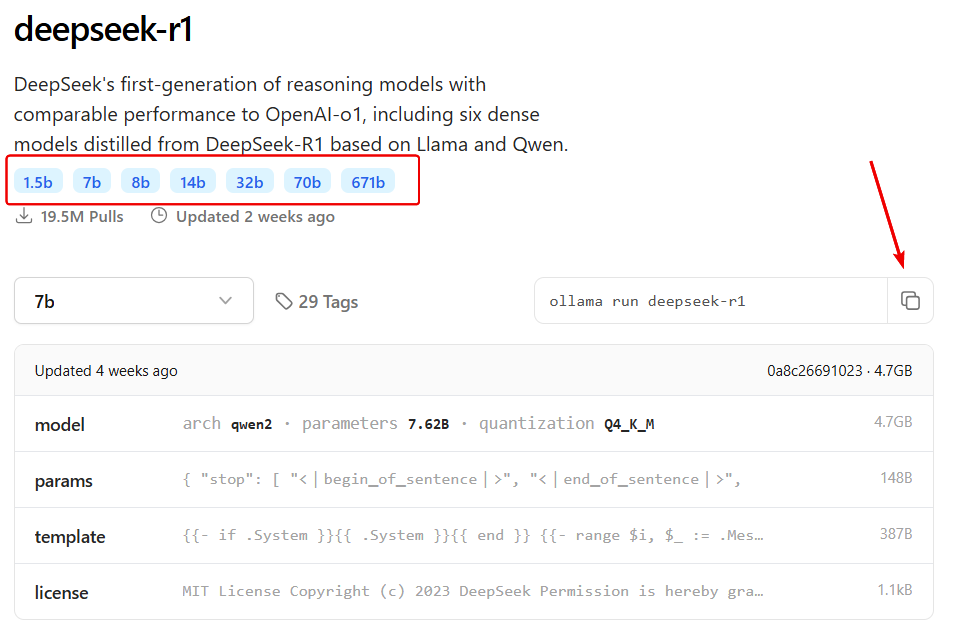
Download and install Ollama using the latest installer for your environment - <https://ollama.com/>

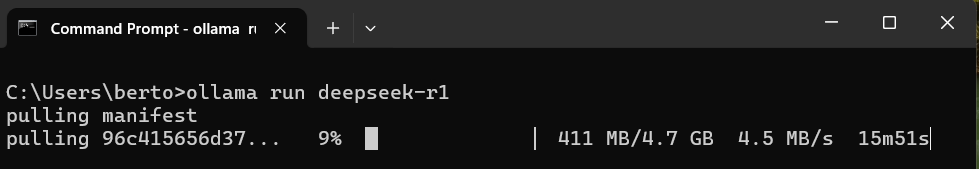


# Download LLM and Start Hoster

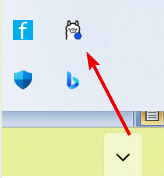
Select the appropriate LLM - <https://ollama.com/search>

Or go directly to the DeepSeek model - <https://ollama.com/library/deepseek-r1> and select the appropriate parameter related model. Then click on the copy button (red arrow). Paste this command into a Terminal\Command dialog.

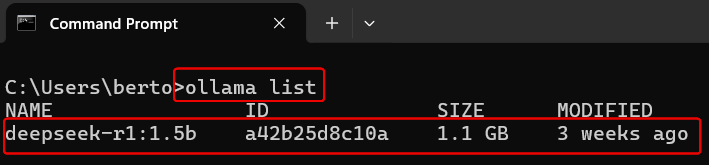




When the model has been downloaded and running, you can verify that Ollama is running from the System Tray:

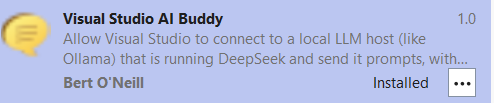


Or from a command prompt run the command ollama list – this will list the available models:



# Install *AI Buddy* Extension (Visual Studio)

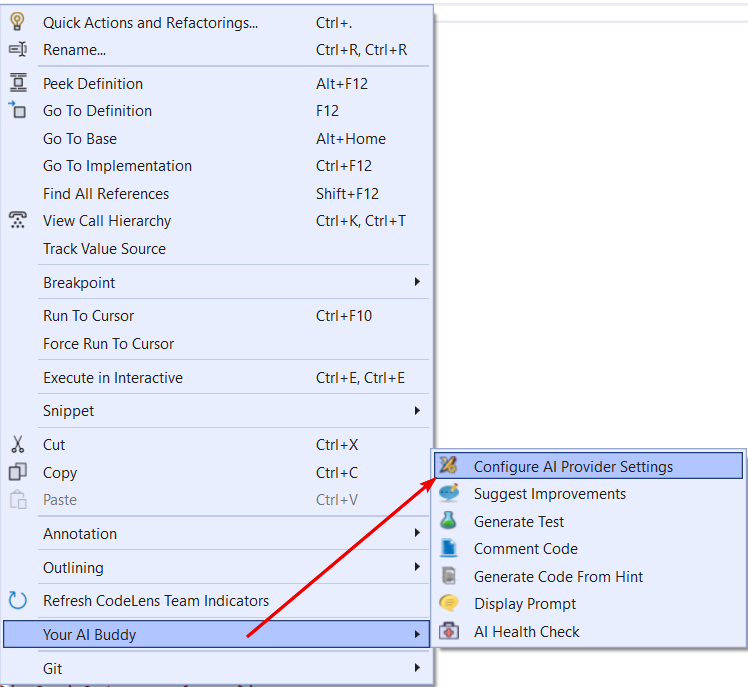
You can either search



# Using AI Buddy

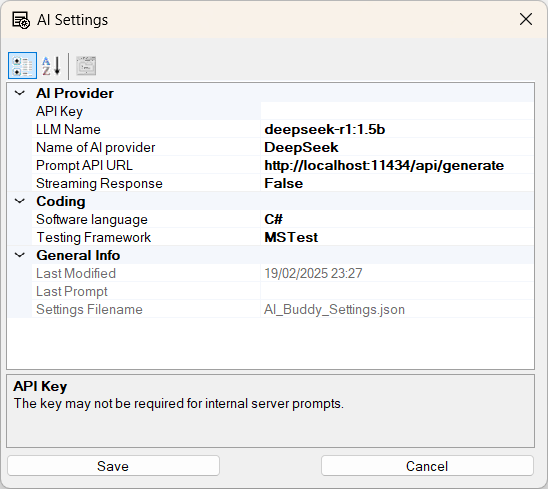
Once you have the *AI Buddy* extension installed, you can initially configure the settings against your local LLM Hoster. To do this, open the editor (as all the menu options are only available from the editor context menu) and right click in the editor pane, and select the menu option *Your AI Buddy* and then select the sub-menu option *Configure AI Provider Settings*.

## Configure AI Provider Settings



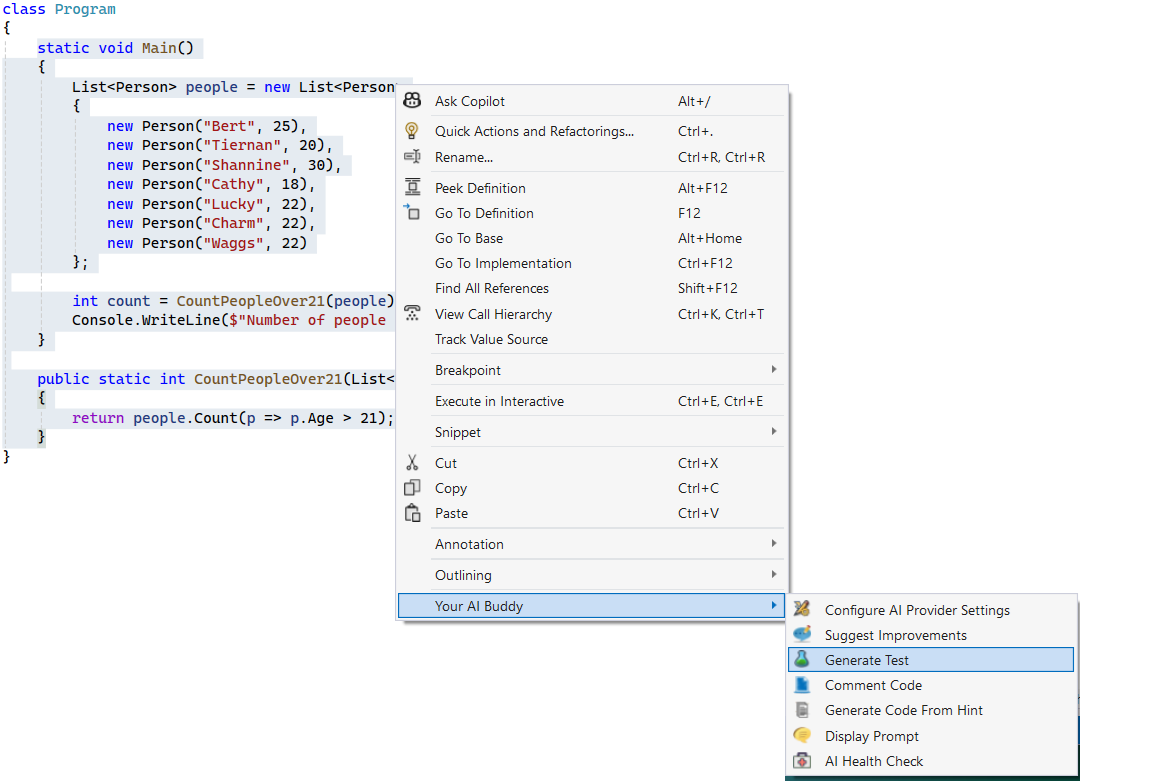
This will display the *AI Settings* dialog, from here you must enter the:

1. Prompt API URL
2. LLM Name
3. Software Language (used in some of the prompts to respond in a particular language)
4. Testing Framework (used in some of the prompts to respond in a particular language)



## Generate Unit Tests

Highlight the code that you wish to generate unit tests for, within the editor. Right click the code and select the submenu item *Generate Unit Tests*.

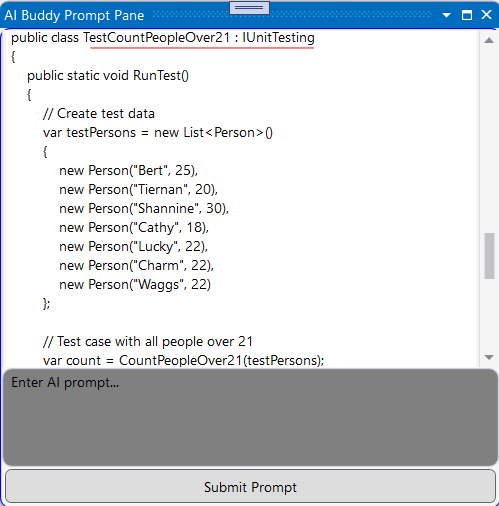


Your selected code will be pasted into a generic prompt for DeepSeek to analysis. An hourglass will appear as DeepSeek is generating the response.

A screenshot of a computer program

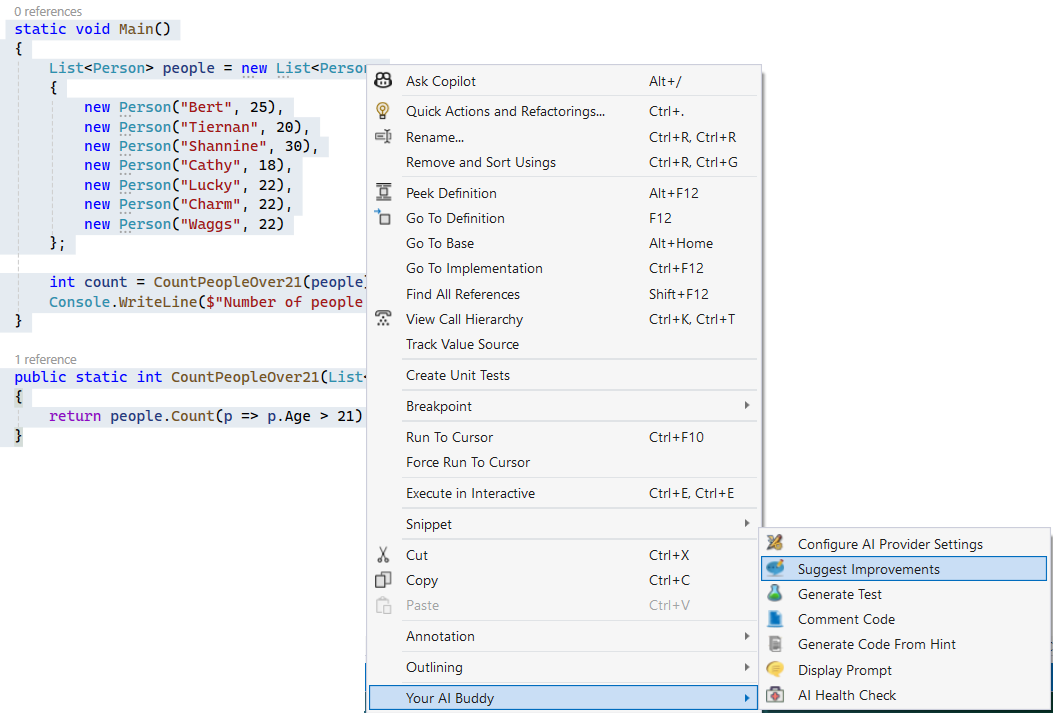
AI-generated content may be incorrect.

This will then generate several Unit Tests based on your *Testing Framework* and *Programming Language* settings. From here, you can copy and paste the code into your test class.

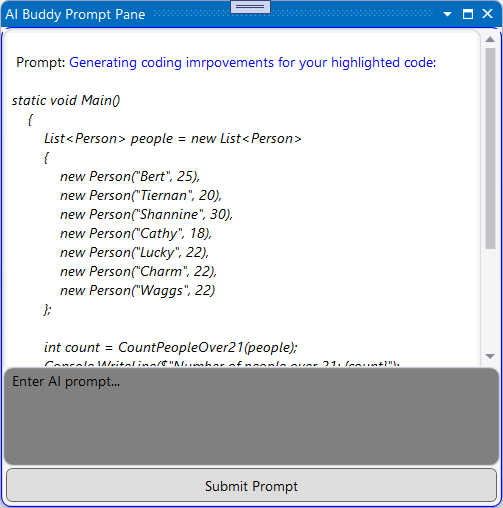


## Suggest (Coding) Improvements

Highlight the code you wish to get coding improvements for, then right click the code and select the submenu item *Suggest Improvements*.



Your highlighted code will be pasted into the generic prompt dialog and DeepSeek will analysis and then return with a response.



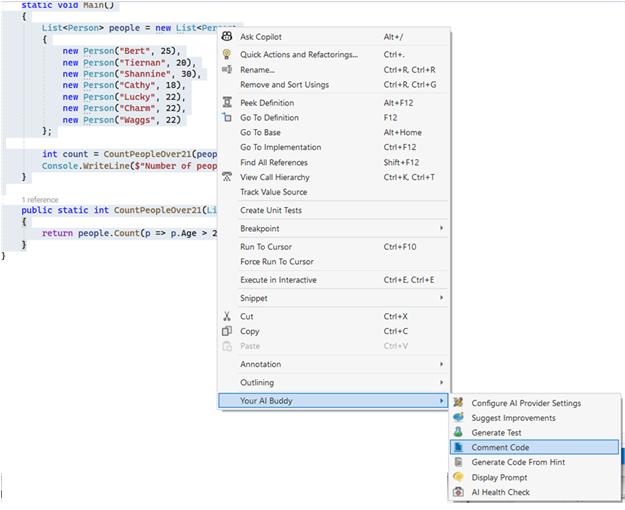
Below, you can see the response from DeepSeek for your prompt.

A screenshot of a computer error

AI-generated content may be incorrect.

## Comment Code

Highlight the code you wish to generate comments against, right click the code and select the submenu item *Comment Code.*



Your code will be pasted into the generic prompt dialog. Which DeepSeek will take as your prompt.

A screenshot of a computer program

AI-generated content may be incorrect.

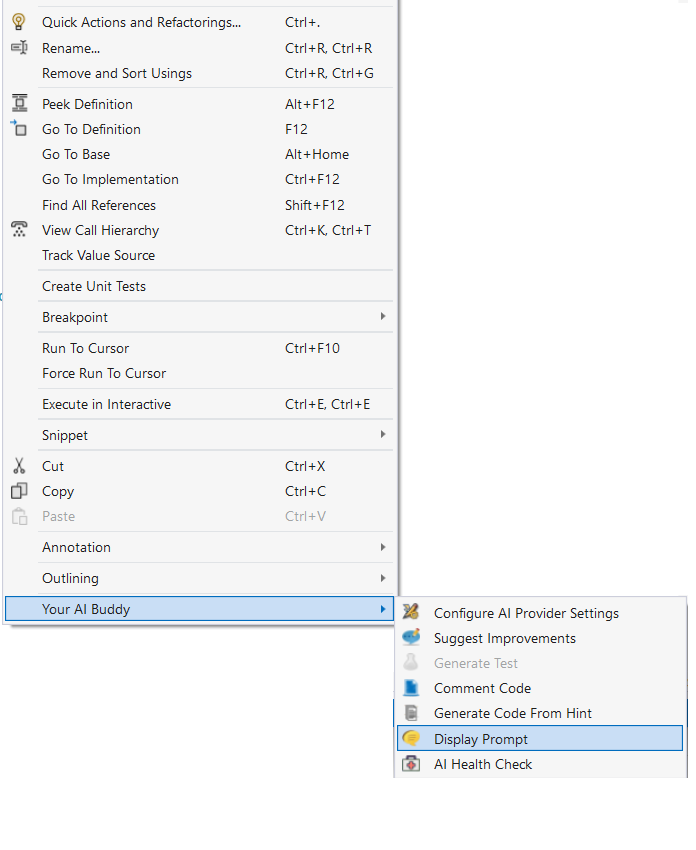
DeepSeek will respond with comments on your code, which you can use to comment your code or paste into a document (a User Guide for e.g.).

A screenshot of a computer

AI-generated content may be incorrect.

## AI Prompt Dialog

There is a generic prompt dialog, where you can bring up an *AI Prompt*. You don’t need to highlight any code, as this will be a general prompt dialog.



From here, you can enter any prompt message for DeepSeek to analysis and respond to. Enter the prompt into the gray textbox and hit the Submit button.

A screenshot of a computer

AI-generated content may be incorrect.

After hitting the Submit button, the prompt is posted to DeepSeek.

A screenshot of a computer

AI-generated content may be incorrect.

DeepSeek will respond to your prompt as normal (if using the live DeepSeek API, but without your prompt leaving your network).

A screenshot of a computer program

AI-generated content may be incorrect.

### Pasting an Image With Prompt

If you host other LLM’s other than DeepSeek (for e.g. LLaVA 1.6 or Llama 3.2) you can paste (Ctrl V) an image into the Prompt textbox, and submit it along with a prompt. Currently DeepSeek doesn’t accept images as part of the Prompt.

A screenshot of a computer

AI-generated content may be incorrect.

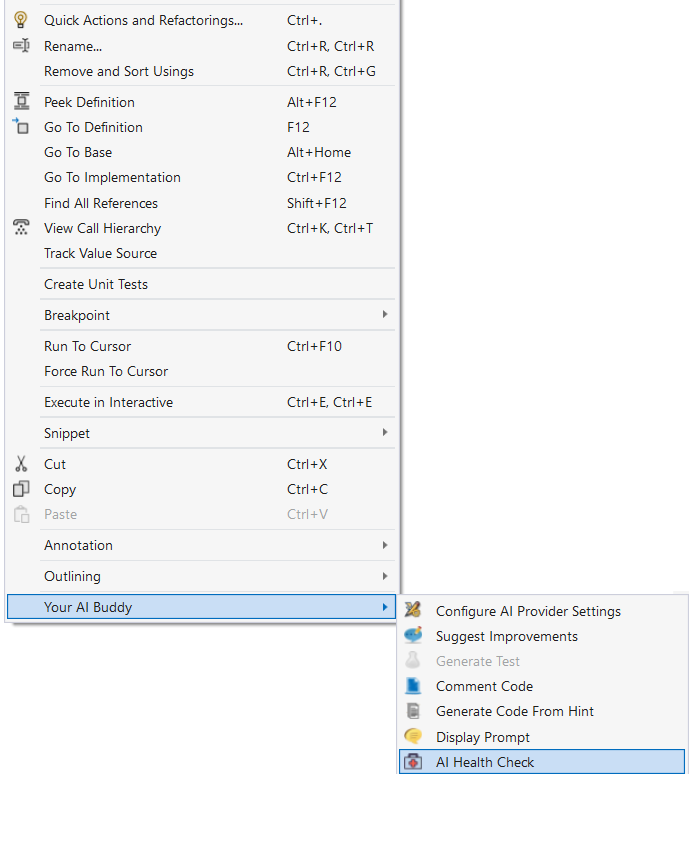
The prompt message will expand the image and post the prompt message to the hosted LLM.

A screenshot of a computer program

AI-generated content may be incorrect.

# Health Check

You may want to check if the Hoster is up and running, you can do this with the submenu item *Health Check*.



A valid response:

A screenshot of a computer

AI-generated content may be incorrect.

An invalid response:

A screenshot of a computer error message

AI-generated content may be incorrect.