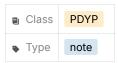
Open Source Al Coding assistants comparison



name	workspace integration	community & support	type of model	open source	possibility to self- host
CodeT5	None	-	family of LLMs	Yes	Yes
PolyCoder	-	-	LLM	Yes	Yes
FauxPilot	Vscode Visual Studio Vim Jetbrains (?work in progress)	GitHub based wiki and discussion forum	LLM	Yes	Yes
Tabby	Vscode IntelliJ VIM/NeoVIM	GitHun repository Slack Twitter/X LinkedIn Newsletter	LLM	Yes	Yes
CodeGeeX	Vscode Jetbrain AndroidStudio HBuilderX		large pre-trained model based on transformers	Yes	Yes
Jedi	Vim Vscode Emacs Kate Atom GNOME builder Gedit	Github repository			Yes

Languages support

name	Python	C++	JavaScript	TypeScript	Golang	Rust	Ruby
CodeT5	У		у		у		у
PolyCoder	У	у	у	у	у	У	у
FauxPilot							
Tabby	У	у	У	у	у	У	у
CodeGeeX	У	у	у	у	у		
Jedi	У						
name	PHP	Java	Kotlin	C#	Kotlin	Solidity	Shell
CodeT5	у	у					
PolyCoder	У	у		у			
FauxPilot							
Tabby	у	у	у	у	у	У	
CodeGeeX	у	у		у			у
I a alt							

name	R	Cuda	css	TeX	С	Scala
CodeT5						
PolyCoder					у	у
FauxPilot						
Tabby						
CodeGeeX	у	у	у	у	у	
Jedi						

CodeGeeX

https://github.com/THUDM/CodeGeeX

- pre-trained on 20+ programming languages
- docker for hosting
- training material:
 - o open-source code datasets: the Pile and CodeParrot
 - o Public GitHub repositories, filtered with the specified criteria
- HumanEval-X new benchmark proposed for evaluation of generated code

PolyCoder

https://github.com/VHellendoorn/Code-LMs

- · LLM model itself
- training 12 programming languages:
 - supervised learning
 - GPT-NeoX toolkit
 - o Public Githun repositories, filtered with the specified criteria
- Docker for hosting (base image = 5.4 GB)

CodeT5

https://github.com/salesforce/CodeT5

- FAMILY of encoder-decoder models
- training:
 - CodeSearchNet data
 - CodeXGLUE benchmark
 - NL pre-trained models based on Transformer architectures
 - PL pre-trained models:
 - CuBERT
 - CodeBERT
 - GPT
- besides autocompletion:

- o text-to-code
- o code summarization

FauxPilot

https://github.com/fauxpilot/fauxpilot

- · Docker for hosting
- SalesForce CodeGen models inside of NVIDIA's Triton Inference Server with the FasterTransformer backend

Tabby

https://github.com/TabbyML/tabby?tab=readme-ov-file

- self-contained
- supports consumer-grade GPUs
- todo publikacje

Jedi

https://github.com/davidhalter/jedi?tab=readme-ov-file

Other possible functionalities

- explaining code
- · recommending alternatives
- bug fixing
- · vulnerability detection
- refactoring

Comparison criteria

- latency
- accuracy standardized test (simple algorithm, like sort) in a fresh IDE setup and check for empiric results
- · security/privacy and data handling
- training material for the model