

Improving Travo test coverage with AI agents on Large Language Models

<https://codimd.math.cnrs.fr/s/9Os27bNKw#>

Travo^[1] is a lightweight open source Python toolkit turning GitLab forges into flexible management solutions for computer assignments.

It is currently actively developed at LISN and Paris Saclay University.

Travo is used in production in a dozen large classes at Université Paris-Saclay and UQAM, and many other smaller classes.

Travo test coverage is still around 50% and needs to be improved: due to its complex interactions with the forge, the task can be overwhelmingly time consuming when manually implemented.

For this internship we propose to use AI agents connected to Large Language Models to increase the test coverage of Travo.

Alshahwam et al (2024) ^[2] describes the deployment of TestGen-LLM at Meta, an AI agent dedicated to unit test improvement.

We plan to follow and test the workflows implemented there using some available open-source tools.

- cover-agent by qodo ^[3]: an open-source implementation of Meta's TestGen-LLM
- LiteLLM by BerriAI ^[4]: an open-source API to interface users and agents with LLMs
- StarCoder2 by BigCode Project ^[5]: a family of code generation models trained on a dataset derived from the Software Heritage archive, accessible through the HuggingFace Hub.

The intern will practice StarCoder2 LLMs prompting for test specification.



Depending on the results of our first experiments, fine tuning on Python source code might be applied.

Thanks to the interoperability functionalities offered by LiteLLM, some benchmarking on different LLMs might also be performed.

Applicants should send a CV and a motivation letter to

✉ chiara.marmo@universite-paris-saclay.fr

-
1. <https://travo-cr.gitlab.io/travo/> 
 2. <https://doi.org/10.48550/arXiv.2402.09171> 

3. <https://github.com/Codium-ai/cover-agent> 
4. <https://litellm.vercel.app/> 
5. <https://github.com/bigcode-project/starcoder2> 