version 2025-11-3

Benchmarking Automated Algorithm Discoveries at the TBU in Zlín

Martin Zhelyazkov

Project planning| 2026

Tomas Bata University in Zlín

TBU Library

Table of Contents

1. Introduction

The internship focuses on benchmarking Large Language Model (LLM) driven frameworks for **Automated Algorithm Discovery (AAD)**. Recent research has demonstrated that LLMs can not only generate code but also discover novel algorithms for solving computational problems.

The objective of this project is to systematically compare several algorithm discovery frameworks and evaluate their computational requirements, performance characteristics, and feasibility for real-world applications.

This document describes the background of the project, the objectives, scope, planning, stakeholders, risks, and reporting structure of the internship project.

2. Research Lab Background

2.1 AI Research Lab

The internship is conducted within an artificial intelligence research laboratory focused on developing advanced machine learning and algorithm discovery systems.

One of the main research initiatives of the lab is **EASE**, an internal framework designed to explore automated algorithm discovery using Large Language Models.

The lab investigates how AI systems can automatically design algorithms, evaluate them, and iteratively improve them.

2.2 Research Area

The field of **Automated Algorithm Discovery (AAD)** has gained significant attention due to recent research breakthroughs such as **FunSearch** and **AlphaEvolve**, which demonstrate the ability of LLMs to generate novel algorithms.

Several frameworks have emerged in this research area, including:

- LLAMEA

- LLM4AD
- EASE (the UTB lab's internal framework)

However, there is currently limited benchmarking that compares these frameworks in terms of:

- computational requirements
- scalability
- algorithm quality
- hardware usage

This project aims to perform a structured benchmarking study of these frameworks.

3. Current Situation

Although algorithm discovery frameworks are rapidly evolving, their real-world applicability is still uncertain.

Many frameworks have been proposed, but there is limited knowledge regarding:

- how well they perform across different problem domains
- how much computational power they require
- how scalable they are when applied to complex problems

Furthermore, organizations interested in applying automated algorithm discovery often face challenges related to hardware requirements, execution time, and reliability of generated algorithms.

As a result, there is a need for systematic benchmarking that evaluates the performance and practicality of these frameworks.

4. Project Objective

The main objective of this project is to benchmark multiple LLM-driven frameworks for automated algorithm discovery and evaluate their performance.

The project will focus on:

- implementing operational versions of LLAMEA, LLM4AD, and EASE
- defining hardware requirements for the frameworks
- designing an automated benchmarking pipeline
- defining evaluation metrics for algorithm quality
- creating a baseline dataset of benchmark problems
- executing experiments to compare framework performance
- evaluating the feasibility of applying these frameworks in real-world scenarios

By the end of the internship, the project should provide a clear comparison between the frameworks and insights into their strengths and limitations.

5. Technical Environment

The project will be conducted using several software tools and technologies.

Development Environment

- Python for implementation and experimentation
- Jupyter notebooks for exploration and analysis

Frameworks

- LLAMEA
- LLM4AD
- EASE

Machine Learning Infrastructure

- Large Language Models for algorithm generation
- local computing resources or cloud infrastructure for experiments

Data and Benchmarking

- curated benchmark datasets
- evaluation metrics for algorithm quality and efficiency

These technologies form the experimental environment used during the project.

6. Project Scope

6.1 In Scope

The following tasks are included in the project:

- setting up and running LLAMEA, LLM4AD, and EASE
- defining benchmarking criteria and evaluation metrics
- designing and implementing a benchmarking pipeline
- running benchmark experiments
- analysing algorithm quality and computational cost
- testing frameworks on selected real-world examples
- documenting results and findings
- DEPLOY ON DOCKER

6.2 Out of Scope

The following tasks are not part of the project:

- redesigning the frameworks themselves
- developing entirely new algorithm discovery frameworks
- deploying frameworks into production systems
- maintaining infrastructure after the internship

The project focuses on evaluation and experimentation.

7. Risk Analysis

- High Computational Requirements
- Benchmark Design Difficulty
- Scalability Limitations

8. Project Planning

8.1 - Phase 1 - week 1/2
- Onboarding and Research

8.2 - Phase 2 - week 1/2
- Framework Implementation

8.3 - Phase 3 - week 3/4
- Benchmark Design

8.4 - Phase 4 - week 5/6
- Benchmark Execution

8.5 - Phase 5 - week 7/8
- Stability Optimization & Extended Benchmarking

8.6 - Phase 6 - week 9/12
- Buffer, Final Analysis and Reporting

9. Definition of Done

The project will be considered complete when the following conditions are met:

- LLAMEA, LLM4AD, and EASE are successfully implemented and executed
- a working benchmarking pipeline is developed
- evaluation metrics for algorithm quality are defined
- benchmark experiments are completed
- results are analysed and documented
- a final internship report is submitted
- the project results are presented to the research team.