

OSA: ИИ- ПОМОЩНИК ДЛЯ РАЗРАБОТКИ НАУЧНОГО ОТКРЫТОГО КОДА

Андрей Гетманов
исследователь и опенсорс-
энтузиаст

ИТМО



ЧТО ДЕЛАЛИ?

Проблема: недостаточное качество и высокая трудоемкость создания "научного опенсорса"

Генерация

Документации и кода для выполнения этих рекомендаций

Рекомендации

По улучшению open source репозитория

OSA

(Open Source Advisor)

открытая библиотека на Python для повышения качества научного open source

На входе

- репозиторий с черновым программным кодом
- научная статья (если есть)
- запрос пользователя (целевой язык, доп. контекст)

СЦЕНАРИЙ ИСПОЛЬЗОВАНИЯ

программный код с комментариями (на английском или русском языке)

README и документация (на английском языке)

автоматически сформированный Pull Request со всеми изменениями

На выходе

ЧТО УЖЕ РЕАЛИЗОВАНО?

по состоянию
на 25.02.25

- Генерация README для различных типов репозиториев (инструмент, статья) на основе анализа кода и доступных шаблонов и примеров
- Генерация docstrings для автосборки документации по ним
- Автоперевод документации на английский или русский язык
- Автоматическая обработка репозитория (fork -> modify -> Pull Request) на GitHub
- Поддержка различных режимов работы с LLM (локальный, облачный)

СПОСОБЫ ДОСТУПА

WEB-APP на базе
Streamlit

DOCKER локальное
развертывание,
взаимодействие
через консоль

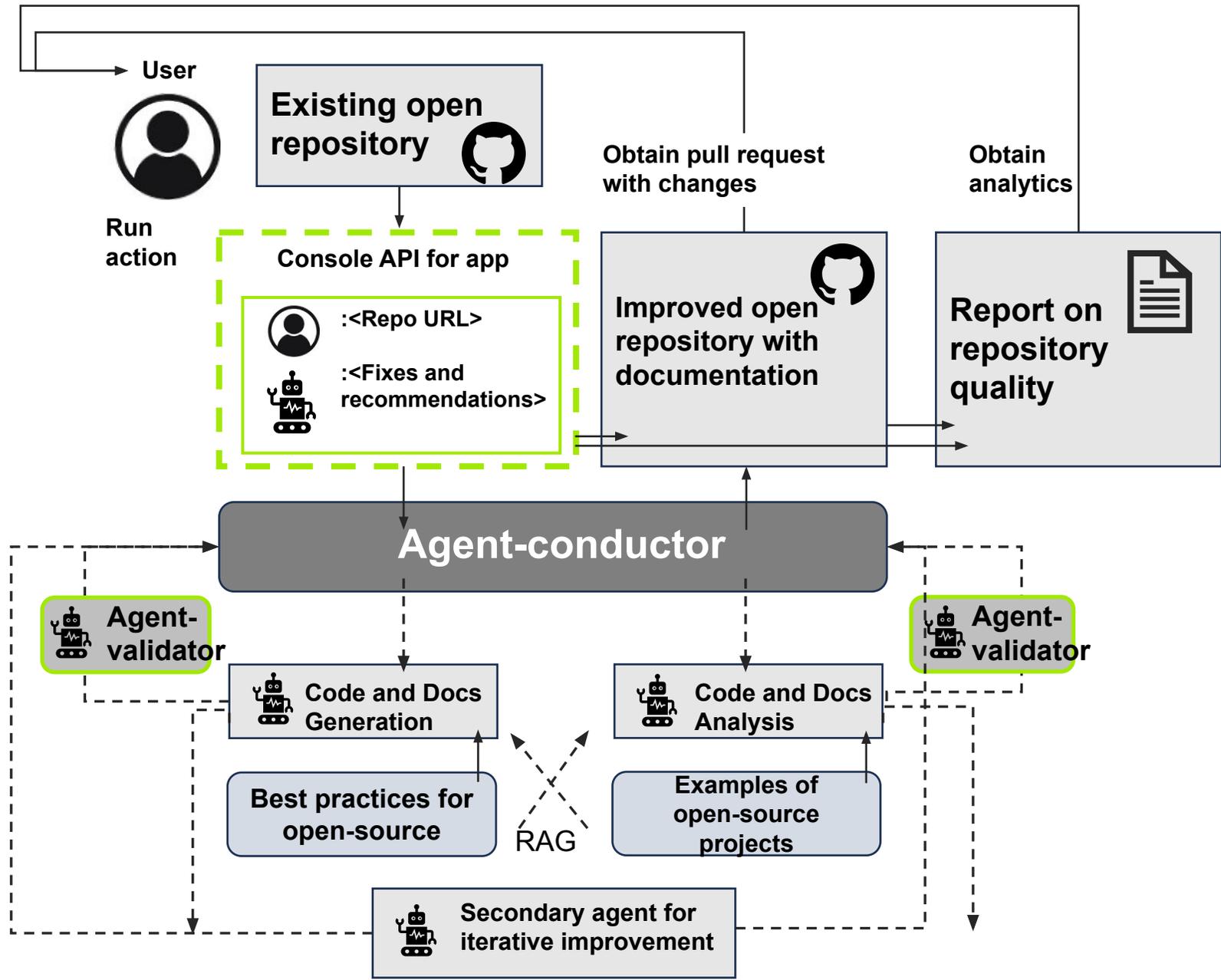
ИИ-ЧАСТЬ

декомпозиция задачи и
имеющегося кода и документации,
итеративное применение
специализированных промптов

БАЗОВАЯ МОДЕЛЬ

локальная Llama-3.1-70B +
работа с OpenAI API

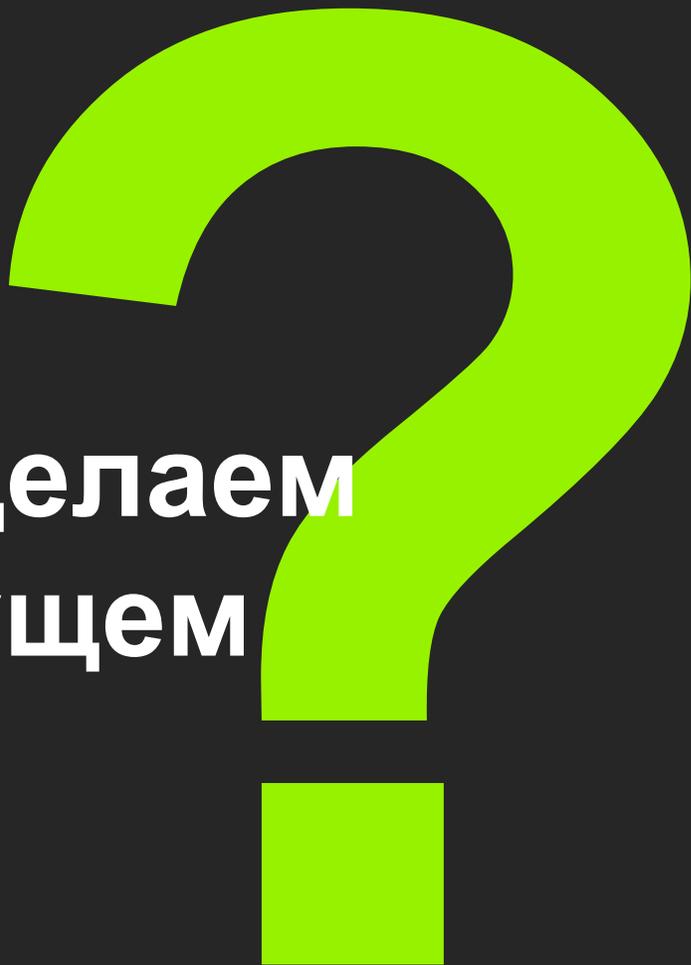
Образ результата



Легенда

- Primary agent
- Secondary agent
- System agent
- Tool usage
- User interface
- Agent external communication

Что сделаем в будущем



- Агентную логику работы
- Анализ репозитория с подготовкой рекомендаций
- Авто-генерация кода для тестов и CI/CD
- Повышение качества за счет RAG
- Интеграция со средой запуска кода
- Алгоритмы оценки качества результата, итеративное улучшение
- Больше моделей

+ **Апробация и тестирование**

Примеры

Как мы применяли
OSA для проектов
команд ИТМО
и не только

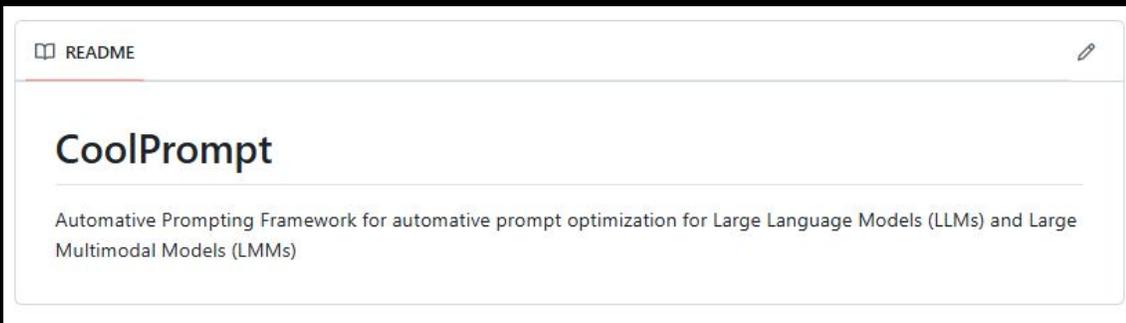


Репозиторий лаборатории КТ

Авто-модификации OSA:
расширение существующего
README

Стало

Было



Overview

CoolPrompt is a software project focused on implementing core functionality through organized code files. It aims to provide a modular, scalable, and maintainable application structure. Targeting developers, it facilitates easier development, debugging, and future enhancements by creating a cohesive and efficient codebase.

Table of contents

- [Core features](#)
- [Installation](#)
- [Examples](#)
- [Documentation](#)
- [Getting started](#)
- [Contributing](#)
- [License](#)
- [Acknowledgments](#)
- [Contacts](#)
- [Citation](#)

Core features

- Modular Codebase:** Files handle key operations, data processing, and flow for scalability and maintainability.
- Structured Architecture:** Organized files contribute to cohesive, efficient application development and performance.
- Enhanced Development:** Facilitates easier debugging, future enhancements, and overall project maintenance.
- Scalability and Maintainability:** Architecture supports modularity for easy scaling and maintenance of the project.
- Efficient Functionality:** Seamless file integration creates a structured and organized codebase for optimal performance.

Installation

Install CoolPrompt using one of the following methods:

Build from source:

- Clone the CoolPrompt repository:

```
> git clone https://github.com/f11pcoin/CoolPrompt
```

- Navigate to the project directory:

```
> cd CoolPrompt
```

- Install the project dependencies:

Репозиторий ХБК ИТМО

проект SequenceCraft, Мария Еремеева

Авто-модификации OSA: генерация README по коду и статье

File	Commit Message	Time
data	added new model and data	10 months ago
utils	added new model and data	10 months ago
weights	added new model and data	10 months ago
.gitignore	code upload	11 months ago
README.md	Initial commit	2 years ago
requirements.txt	added new model and data	10 months ago
run_model.py	added new model and data	10 months ago

Overview

dnaZyme

Было

Стало

File	Commit Message	Time
data	added new model and data	10 months ago
utils	added new model and data	10 months ago
weights	added new model and data	10 months ago
.gitignore	code upload	11 months ago
README.md	OSA recommendations	now
requirements.txt	added new model and data	10 months ago
run_model.py	added new model and data	10 months ago

Overview

iTMO
DNAZYME

license not specified | last commit april 2024 | purebasic 84.7% | languages 2

Built with the tools and technologies:

Python

dnaZyme is a machine learning platform that helps researchers and scientists extract meaningful patterns and insights from complex nucleic acid data. By providing a structured approach to data analysis, dnaZyme enables users to identify hidden relationships and trends, ultimately accelerating discovery and innovation in the field of genomics and biotechnology.



Репозиторий зарубежной команды

Авто-модификации OSA:
генерация README
по коду и статье, перевод
структуры каталогов с
португальского на английский

Было

OSRM-Sapucaia Public

main 1 Branch 0 Tags

Go to file Add file Code

vtor-fmnds Adicionando Mapas de Localização Deb3b2f · 5 months ago 5 Commits

- Artigos Adicionando Artigos 5 months ago
- Códigos Adicionando Códigos 5 months ago
- Diagrama de Fluxo Adicionando Diagrama de Fluxo 5 months ago
- Mapa de Sapucaia Adicionando Mapas de Localização 5 months ago
- LICENSE Initial commit 5 months ago
- README.md Initial commit 5 months ago

README MIT license

OSRM-Sapucaia

Стало

Core features

- Route Optimization:** Analyze and optimize waste management routes for efficiency.
- Simulation Strategies:** Use Monte Carlo simulations to evaluate waste management scenarios.
- Visual Representation:** Complement code with a visual representation of waste management process.
- Comprehensive System:** Enable analysis, optimization, and visualization of waste management routes.
- Sustainability Focus:** Aim for more effective and sustainable waste management practices.

Installation

Install OSRM-Sapucaia using one of the following methods:

Build from source:

- Clone the OSRM-Sapucaia repository:

```
> git clone https://github.com/f11pcoin/OSRM-Sapucaia
```
- Navigate to the project directory:

```
> cd OSRM-Sapucaia
```

OSRM-Sapucaia Public

OSA 2 Branches 0 Tags

Go to file Add file Code

This branch is 3 commits ahead of vitor-fmnds:OSRM-Sapucaia:main

- ftpcoin OSA recommendations 2f2bc0 · 8 hours ago 8 Commits
- articles OSA recommendations 8 hours ago
- codes OSA recommendations 8 hours ago
- flow diagram OSA recommendations 8 hours ago
- map of sapucaia OSA recommendations 8 hours ago
- LICENSE Initial commit 5 months ago
- README.md OSA recommendations 8 hours ago

OSRM-SAPUCAIA

license MIT link commit September 2024 Jupiter robotbot 100.0% languages 1

Overview

OSRM-Sapucaia focuses on optimizing waste management routes through route analysis, Monte Carlo simulations, and visual representations. This project aims to enhance waste collection efficiency, evaluate different strategies, and visualize the waste management process. It targets organizations seeking to improve their waste management practices for sustainability.



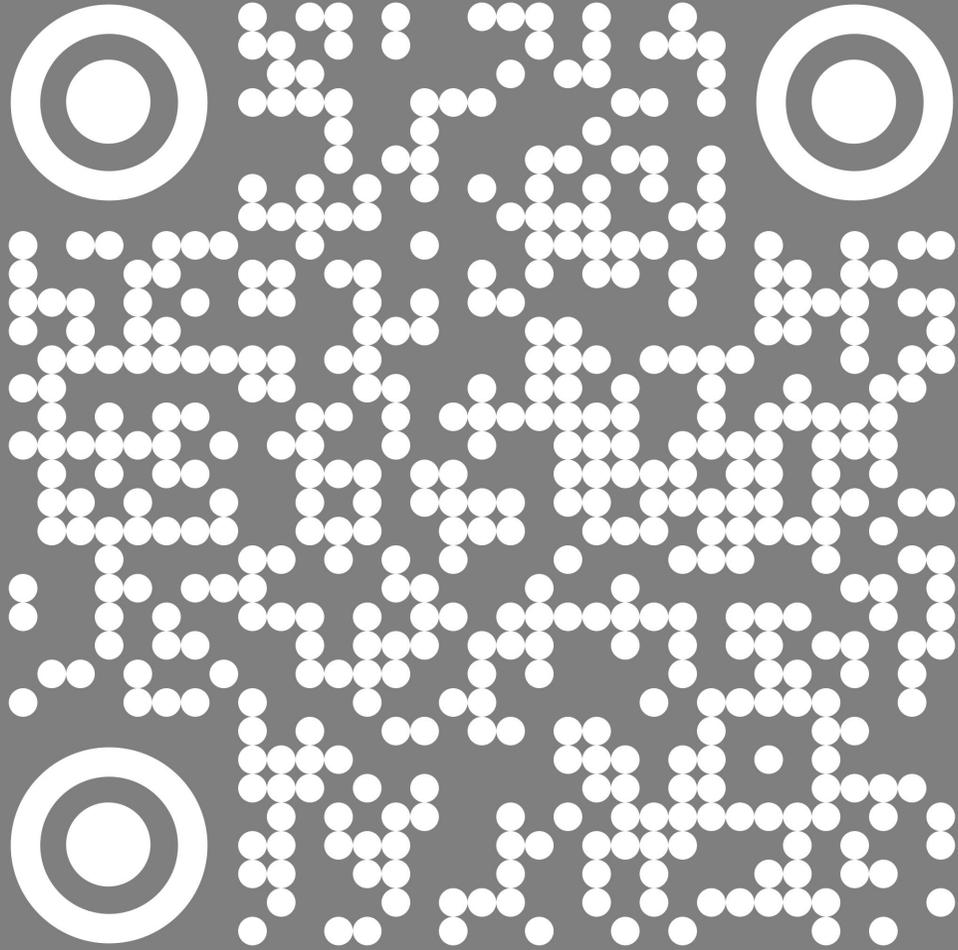
This is mega! Let me
share with the team



The repository was well organized and the
information was clear. It was impressive
what the tool did automatically

use osa

- **Репозиторий** – создавайте issue с предложениями
- **Telegram-чат** – задавайте вопросы по использованию
- Запускайте и улучшайте свои репозитории
- **Звезды** – ставьте и поддерживайте проект



- **Репозиторий** – создавайте issue с предложениями
- **Telegram-чат** – задавайте вопросы по использованию
- **Запускайте и улучшайте свои репозитории**
- **Звезды** – ставьте и поддерживайте проект