

Научно-технический проект на тему:



# Разработка нейронной сети для повышения читабельности декомпилированного кода на языке Си

Проектная  
деятельность  
СНО НИЯУ МИФИ

# Команда и роли



Кислов Константин  
Александрович,  
С23-712

Организация, презентация



Лялин Максим  
Андреевич,  
Б22-554

Создание и обучение модели

< Бехтин Артем Владимирович, Б22-515 >

Научный руководитель



Божко Артем  
Александрович,  
С23-712

Декомпиляция



Ременяко Владислав  
Денисович,  
С23-712

Тестирование

# Актуальность

Проблема **восстановления исходного кода** из машинного возникает сравнительно часто: анализ кода, **близкого к исходному**, позволяет понять, как работает программа, какие данные использует, куда и что отправляет, а также какие в ней есть слабые места и как она реагирует на аварийные ситуации. Для выполнения данной задачи используются **программы-декомпиляторы**. Однако они генерируют **трудночитаемый** код, что затрудняет процесс его **анализа**. В связи с этим возникла идея использовать для решения данной проблемы бурно развивающиеся в настоящее время технологии ИИ, а именно – модели-**трансформеры**.

# Исследовательская составляющая

**Гипотеза:** на основе современных технологий ИИ возможно создать и обучить нейронную сеть, способную облегчить процесс анализа преобразованного из машинного на язык Си программного кода

**Объект исследования** – декомпиляция программного кода

**Предмет исследования** – применение языковых моделей для анализа программного кода

**Методы исследования:** поисковый, анализ, сравнение, измерение, тестирование, моделирование, программирование

# Цель и задачи

**Цель проекта:** разработка и обучение нейронной сети-трансформера для преобразования декомпилированного кода на языке Си в более читабельный

## Задачи:

### 1. Поиск

< и анализ информации по данной теме >

### 2. Тестирование

< декомпиляторов и анализ их работоспособности >

### 3. Создание

< выборки для обучения модели >

### 4. Разработка

< модели нейросети-трансформера и ее обучение >

### 5. Тестирование

< «чернового» варианта модели и анализ результатов ее работы >

### 6. Модернизация

< нейронной сети и ее дальнейшее обучение >

### 7. Подведение итогов и определение перспектив проекта

# Промежуточные результаты

- (1) Найдены существующие эффективные декомпиляторы (RetDec, Ghidra, IDA);
- (2) В качестве примера для обучения, анализа работы и модернизации архитектуры был найден исходный код модели трансформера, способного переводить текст с немецкого на английский;
- (3) Собран структурированный датасет для тестирования декомпиляторов и анализа их работы (все тестовые данные сгруппированы: C-файл + exe-файл);
- (4) Найден объемный датасет для обучения модели, состоящий из исходного кода около 106 тысяч программ;
- (5) Проведены первичные тесты выбранных декомпиляторов, для генерации декомпилированного кода с целью обучения нашей модели был выбран RetDec;
- (6) В качестве основы для модели нейронной сети было решено рассматривать открытый исходный код OpenAIGPT для модернизации и обучения;
- (7) Настроена и автоматизирована работа RetDec с целью формирования части датасета, состоящей из декомпилированного кода, для обучения нейросети;
- (8) Выборка из исходного кода программ была скомпилирована, далее декомпилирована при помощи настроенного RetDec;
- (9) Были проведены первые этапы обучения нейросети, проанализированы первичные результаты ее работы, определены перспективы дальнейшего развития

GitHub:





**Спасибо за  
внимание!**