Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.8

Дубинина
Евгения Олеговна

Модели и средства автоматизации разработки программного кода на основе искусственных нейронных сетей

АВТОРЕФЕРАТ
на соискание степени магистра наук
по специальности 7-06-0611-03 – искусственный интеллект

Научный руководитель
Захаров В.В.
кандидат технических наук

Минск 2026

Краткое введение

Общая характеристика работы

Цель и задачи проводимых исследований

Целью диссертационной работы является повышение эффективности процесса разработки интеллектуальных систем на платформе OSTIS за счет создания инструментария автоматической генерации семантического кода (SC-код) на основе специализированных локальных нейросетевых моделей.

Указанная цель определяет следующие задачи исследования:

а) Провести анализ существующих методов автоматической генерации программного кода и выявить ограничения глобальных языковых моделей (LLM) применительно к предметно-ориентированным языкам графовой структуры.

б) Разработать методику формирования обучающего корпуса данных для языка SC-код, объединяющую парсинг существующих репозиториев и генерацию синтетических примеров на основе формальной грамматики.

в) Осуществить адаптацию и дообучение (Fine-Tuning) большой языковой модели с использованием методов низкоранговой декомпозиции матриц весов (LoRA) для задачи генерации семантических структур.

г) Исследовать и применить методы оптимизации инференса (квантование, оптимизация KV-кэша) для обеспечения возможности запуска модели на пользовательском оборудовании с ограниченными ресурсами.

д) Разработать архитектуру и программную реализацию комплекса, включающего локальный сервер инференса и плагин для интегрированной среды разработки (IDE), поддерживающий контекстно-зависимое автодополнение.

е) Провести экспериментальные исследования разработанной системы для оценки синтаксической корректности генерации, производительности и потребления системных ресурсов.

Объектом исследования является процесс автоматизированного проектирования баз знаний и компонентов интеллектуальных систем.

Предметом исследования являются методы, алгоритмы и программные средства адаптации больших языковых моделей для генерации кода на языке SC-код в условиях ограниченных вычислительных ресурсов.

Связь работы с приоритетными направлениями исследований и запросами реального сектора экономики

Тема диссертации соответствует приоритетному направлению научной, научно-технической и инновационной деятельности на 2021–2025 годы «Цифровые информационно-коммуникационные и междисциплинарные технологии, основанные на них производства» (подпункт 1.3 «Искусственный интеллект, робототехника и роботизированные комплексы, цифровые двойники, технологии больших данных, системы управления и анализа бизнес-процессов, технологии блокчейн»), утвержденному Указом Президента Республики Беларусь от 07.05.2020 г. № 156.

Диссертационное исследование выполнено на базе учреждения образования «Белорусский государственный университет информатики и радиоэлектроники» (кафедра интеллектуальных информационных технологий). Работа проводилась в русле развития технологии проектирования интеллектуальных систем OSTIS (Open Semantic Technology for Intelligent Systems), разрабатываемой научной школой кафедры, и направлена на автоматизацию процессов семантического моделирования знаний.

Связь с запросами реального сектора экономики подтверждается выполнением исследований в интересах ООО «Интелиджент Семантик Системс» — резидента Парка высоких технологий, специализирующегося на разработке интеллектуальных систем и коммерциализации платформы OSTIS.

Актуальность работы для реального сектора обусловлена необходимостью снижения порога вхождения для разработчиков интеллектуальных систем и ускорения процессов наполнения баз знаний. Разработанный программный инструментарий решает проблему дефицита кадров, владеющих специализированным языком SC-code, и обеспечивает защиту коммерческой тайны за счет использования локальных моделей искусственного интеллекта, не требующих передачи данных в облачные сервисы.

Результаты диссертации внедрены в производственный процесс ООО «Интелиджент Семантик Системс» в виде интеллектуального плагина для среды разработки, что позволило автоматизировать рутинные операции по написанию кода и повысить производительность труда разработчиков баз знаний.

Личный вклад

Диссертационное исследование является квалификационной научной работой, выполненной соискателем самостоятельно. Личный вклад автора в получение научных результатов заключается в следующем:

Разработана методика формирования и подготовки обучающего датасета для низкоресурсного языка SC-code. Автором лично собран и структурирован корпус данных объемом более 250 000 строк кода, объединяющий реальные примеры из открытых репозиториев платформы OSTIS и синтетические семантические структуры, сгенерированные по разработанным автором правилам, что позволило решить проблему нехватки данных для обучения.

Осуществлена адаптация и оптимизация большой языковой модели Llama 2 (Code Llama) для задач генерации семантического кода (SC-код). Автором предложен и реализован подход к дообучению (fine-tuning) модели на специализированном датасете с использованием метода низкоранговой адаптации (LoRA) и последующим применением 4-битного квантования, что обеспечило возможность локального запуска модели на оборудовании с ограниченными ресурсами (CPU).

Спроектирована архитектура и выполнена программная реализация инструментального средства в виде плагина для среды разработки PyCharm. Автором самостоятельно разработаны модули лексического и синтаксического анализа SC-кода (лексер и парсер на базе Grammar Kit), реализован механизм построения абстрактного синтаксического дерева (AST) и создан локальный HTTP-сервис (микросервис инференса) для интеграции нейросетевой модели в интерфейс IDE согласно архитектурному паттерну «Local Sidecar».

Проведено экспериментальное исследование эффективности разработанного инструментария. Автором разработана методика тестирования, подготовлен набор тестовых запросов и выполнена оценка точности генерации семантических конструкций (по метрикам Pass@k и CodeBLEU), а также проведен комплексный анализ производительности и времени отклика системы при работе на персональных компьютерах.

Все основные теоретические положения, экспериментальные данные и программные реализации, представленные в диссертации, получены автором лично.

Опубликование результатов диссертации

По материалам выполненных исследований опубликована 1 научная работа в сборнике тезисов международной конференции.

Апробирование результатов диссертации

Основные результаты диссертационной работы докладывались и обсуждались на XV Международной научной конференции «Информационные технологии и системы» (ITS-2025), (Минск, 19 ноября 2025 г.).

Краткое содержание работы

Заключение

Основные результаты работы заключаются в следующем:

а) **Проведен системный анализ** методов автоматической генерации программного кода. Выявлено, что глобальные большие языковые модели (LLM), такие как GPT-4 и Llama 2, демонстрируют высокую эффективность для императивных языков, но оказываются практически неприменимы для декларативного языка SC-код в режиме «Zero-shot». Основной причиной является феномен «низкой ресурсности» (Data Scarcity) — отсутствие достаточного объема примеров SC-кода в обучающих выборках. Обоснована необходимость разработки специализированного локального решения, обеспечивающего конфиденциальность и автономность разработки.

б) **Разработана и апробирована методика формирования обучающего корпуса данных.** Для преодоления дефицита данных предложен гибридный подход, сочетающий интеллектуальный парсинг открытых репозиториев экосистемы OSTIS и генерацию синтетических примеров на основе формальной грамматики (EBNF) языка SCs. Сформирован уникальный датасет объемом более 250 000 строк кода, содержащий 70 000 пар инструкций и ответов, что позволило решить проблему «холодного старта» обучения нейросети.

в) **Реализован метод эффективного дообучения (Fine-Tuning)** модели в условиях ограниченных вычислительных ресурсов. С использованием техники низкоранговой адаптации (LoRA) с использованием облачных вычислительных мощностей проведено дообучение базовой модели Llama-2-7B. Доказано, что обучение адаптеров (менее 1% от общего числа параметров) достаточно для снижения функции потерь (Loss) с 2.4 до 0.22 и формирования у модели устойчивых навыков генерации графовых структур.

г) **Предложен алгоритм оптимизации модели для локального инференса.** Применение методов асимметричного 4-битного квантования (Q4_K_M) и гибридной декомпозиции матриц позволило сократить потребление оперативной памяти процессом модели с 14.8 ГБ до 4.6 ГБ. Это обес-

печило возможность эксплуатации системы на стандартных ноутбуках (в том числе на архитектуре Apple Silicon) без использования серверных GPU.

д) **Создан программный комплекс интеллектуальной поддержки.** Разработана микросервисная архитектура, включающая локальный сервер инференса на Python/FastAPI и плагин для интегрированной среды разработки JetBrains PyCharm на языке Kotlin. Реализовано бесшовное взаимодействие компонентов, обеспечивающее генерацию кода в режиме реального времени с учетом контекста (Fill-In-the-Middle).

е) **Проведена экспериментальная оценка эффективности.** Сравнительные испытания показали, что разработанная модель обеспечивает точность генерации синтаксически корректного кода (Pass@1) на уровне 82,5%, что в 6,5 раз превосходит показатели базовой модели. Среднее время генерации подсказки составляет 1,3 секунды, что удовлетворяет эргономическим требованиям к инструментам разработчика.

Практические рекомендации и внедрение

Практическая значимость работы заключается в создании готового к использованию инструмента, который снижает порог вхождения для новых специалистов в технологию OSTIS и ускоряет процесс формализации знаний.

Результаты работы прошли успешную апробацию в производственных условиях на базе ООО «Интелиджент Семантик Системс» (г. Минск). Опытная эксплуатация подтвердила работоспособность предложенных алгоритмов и возможность сокращения трудоемкости кодирования типовых задач на 30–40%.

Для дальнейшего внедрения результатов исследования предлагается:

– **В учебном процессе БГУИР:** Использовать разработанный плагин в лабораторных работах по курсу «Интеллектуальные системы» для помощи студентам в освоении синтаксиса SC-кода и снижения количества синтаксических ошибок.

– **В сообществе Open Source:** Опубликовать обученные веса модели (LoRA adapters) на платформе Hugging Face и исходный код плагина в открытом доступе для развития экосистемы OSTIS.

Перспективы дальнейшей разработки темы

Дальнейшее развитие системы видится в следующих направлениях:

а) **Реализация RAG (Retrieval-Augmented Generation):** Интеграция механизма векторного поиска по базе знаний проекта. Это позволит модели

учитывать контекст не только текущего файла, но и всей семантической сети системы, избегая дублирования идентификаторов.

б) **Поддержка SCg-кода:** Расширение функционала модели для генерации XML-описаний графического представления (GWF), что позволит создавать визуальные графы по текстовому описанию.

в) **Интеграция с LSP:** Перенос логики плагина на протокол Language Server Protocol (LSP), что позволит использовать интеллектуальные подсказки не только в PyCharm, но и в VS Code, Sublime Text и других редакторах.

В заключение автор выражает глубокую благодарность научному руководителю, доценту кафедры интеллектуальных информационных технологий БГУИР **Захарову Владимиру Владимировичу**, за всестороннюю поддержку, ценные советы и помощь в определении вектора исследования.

Особую признательность автор выражает руководителю практики от предприятия, инженеру-программисту ООО «Интелиджент Семантик Системс» **Чайковскому Сергею Владимировичу**, за помощь в сборе обучающих данных, технические консультации по платформе OSTIS и содействие в апробации разработанного плагина.

Также автор благодарит коллектив кафедры ИИТ БГУИР за создание творческой научной атмосферы и за неоценимую помощь в техническом оформлении диссертации, позволившую сосредоточиться на сути исследования.

Список опубликованных работ

1–А. РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОГО ПЛАГИНА С ЛОКАЛЬНОЙ МОДЕЛЬЮ ГЕНЕРАЦИИ КОДА ДЛЯ ЯЗЫКА SC-КОД / Е. О. Дубинина // Информационные технологии и системы 2025 (ITS-2025) : материалы XV Международной научной конференции, Минск, 19 ноября 2025 г.