



Гура Денис Владимирович

Мужчина

+7 (999) 8152292

splinter.rat.pro@gmail.com — предпочитаемый способ связи

telegram: @Splinter_Rat

Проживает: Москва, м. Саларьево

Гражданство: Россия, есть разрешение на работу: Россия

Не готов к переезду, готов к редким командировкам

Желаемая должность и зарплата

Тимлид, техлид, руководитель направления

400 000 ₽ на руки

Специализации:

- Директор по информационным технологиям (CIO)
- Руководитель группы разработки
- Руководитель проектов
- Технический директор (CTO)

Тип занятости: полная занятость, частичная занятость, проектная работа/разовое задание

Формат работы: удалённо, гибрид

Желательное время в пути до работы: не более полутора часов

Опыт работы — 29 лет 4 месяца

Июнь 2025 —
настоящее время
1 год 4 месяца

ООО «Онтохаб»

Москва, www.ontohub.ru/

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения

Технический директор (СТО)

Руководство техническим направлением компании-разработчика ИИ-решений, специализирующейся на системах онтологической памяти и интеграции больших языковых моделей с корпоративными базами знаний.

Проектирование архитектуры ИИ-платформ когнитивной автоматизации: системы управления знаниями на основе графовых моделей, семантический поиск, многоуровневое хранение и извлечение контекста.

Исследование и прототипирование подходов к интеграции онтологических моделей с LLM-системами для задач промышленной автоматизации, энергетики и управления предприятиями.

Разработка методологии внедрения ИИ в корпоративные процессы: от аудита данных и оценки зрелости до пилотного развёртывания и масштабирования.

Прототипирование высокопроизводительных NLP-пайплайнов с GPU-ускорением для обработки русскоязычных текстов.

Проектирование микросервисной инфраструктуры ИИ-платформы на Rust: система оркестрации задач, интеграция ИИ-агентов с внешними сервисами через стандартизированные протоколы (MCP), распределённая обработка данных.

Достижения:

- На внутреннем технологическом стенде спроектирован, реализован и введён в непрерывную эксплуатацию комплекс ИИ-инфраструктуры, включающий: систему внешней памяти для ИИ с семантическим поиском по корпусу 9000+ документов (время отклика менее 50мс); микросервисную платформу из 35+ взаимосвязанных сервисов на языке Rust — современном SOTA в балансе производительности, безопасности памяти и надёжности промышленных систем. Сервисы взаимодействуют по трём каналам: синхронно через REST API, асинхронно через брокер сообщений NATS с гарантированной доставкой (JetStream), и дуплексно в реальном времени через WebSocket — полный спектр коммуникационных паттернов от запрос-ответ до событийно-управляемой архитектуры и потоковой обработки данных
- Внедрена система самодокументирования сервисов на основе двух взаимодополняющих стандартов: OpenAPI 3.1 для синхронных HTTP-взаимодействий и AsyncAPI 3.0 для асинхронных каналов (NATS, WebSocket), обеспечивающая полное покрытие всех межсервисных контрактов единой верифицируемой спецификацией, генерируемой непосредственно из кода
- Разработан и протестирован на внутреннем стенде прототип NLP-движка с GPU-ускорением, обеспечивающий глубокий морфологический и синтаксический анализ русского языка на уровне, не имеющем качественных аналогов среди существующих открытых решений. Быстродействие — 57-кратное превосходство над spaCy при сопоставимом и более высоком качестве разбора
- Проведено сравнительное тестирование открытых LLM-моделей, включая новейшие архитектуры Mixture-of-Experts (Qwen3-Next). Выявлены оптимальные конфигурации для промышленного развёртывания в условиях ограниченных вычислительных ресурсов: модели с 80 млрд параметров при активации лишь 3 млрд обеспечивают качество, сопоставимое с флагманскими решениями, прикратно меньших требованиях к GPU-инфраструктуре. Все исследованные модели — с открытыми весами под лицензией Apache 2.0, что исключает санкционные и лицензионные риски, обеспечивает полный контроль над данными и возможность локального развёртывания в закрытом контуре без зависимости от внешних API
- Выстроена методология четырёхтактного цикла разработки с участием ИИ-агентов,кратно сократившая время реализации задач
- Проведено исследование графовых моделей зависимостей для билингвальной (русский/английский) обработки текстов

Технологии: Rust, Python, CUDA, LLM (GPT, Claude, LLaMA, Qwen, Mistral), RAG, LangChain, LlamaIndex, NATS JetStream, WebSocket, REST API, OpenAPI 3.1, AsyncAPI 3.0, SurrealDB, Meilisearch, MCP (Model Context Protocol), NLP, графы знаний, онтологическое моделирование, микросервисная архитектура, событийно-управляемая архитектура (EDA), GPU-ускорение, Docker, Linux

Июль 2023 —
Май 2025
1 год 11 месяцев

Бэлл Интегратор, ООО

Москва, bellintegrator.ru

Информационные технологии, системная интеграция, интернет

- Системная интеграция, автоматизация технологических и бизнес-процессов предприятия, ИТ-консалтинг

Ведущий системный аналитик

Проектирование и запуск пилотной ИИ-платформы для автоматизации работы с большими языковыми моделями (LLM), локальными и удалёнными. Bell Integrator — крупнейший российский ИТ-интегратор, 3000+ сотрудников, разработчик B2B-решений для банковского, промышленного и государственного секторов.

Достижения:

- Спроектировал архитектуру RAG-системы: обоснован выбор стека реализации (LangChain, LlamaIndex, векторные базы данных), разработана система автоматического обновления и администрирования RAG-источников с полиморфной конфигурацией по типу данных
- Внедрил высокопроизводительную систему доставки сообщений на базе NATS через WebSocket, совместно с командами backend и frontend довёл до промышленного уровня отказоустойчивости
- Спроектировал и внедрил ролевую модель доступа на базе Casbin (RBAC/ABAC), обеспечившую гранулярное управление правами в многопользовательской ИИ-среде
- Провёл комплексное сравнительное тестирование GPU-платформ AMD Instinct MI300X и NVIDIA H100 для задач инференса LLM. Выявлено кратное преимущество AMD: 2.4x по объёму памяти (192 ГБ HBM3 против 80 ГБ), 1.6x по пропускной способности (5.3 ТБ/с против 3.35 ТБ/с), до 3x по соотношению цена/производительность — стратегически значимый результат в контексте технологической независимости и оптимизации бюджетов на GPU-инфраструктуру
- Реализовал систему удалённого запуска процессов дообучения и бенчмаркинга моделей на GPU-облаках (Runpod.io), построил платформу тестирования открытых LLM
- Выполнил подбор, тестирование и интеграцию Autocomplete-движка для внутренней IDE с подключением к LLM
- Обеспечил менторинг и техническое руководство командой разработки

Технологии: Python, LLM (GPT, LLaMA, Mistral), RAG, LangChain, LlamaIndex, NATS, WebSocket, Casbin (RBAC/ABAC), AMD Instinct MI300X, NVIDIA H100, CUDA, ROCm, Runpod.io, векторные базы данных, Docker, Linux

Июнь 2021 —
Июнь 2023
2 года 1 месяц

Независимый консультант, исследователь

Независимая исследовательская и консалтинговая деятельность в области архитектуры распределённых систем и технологий искусственного интеллекта.

- Системное исследование ландшафта больших языковых моделей в период их взрывного развития (GPT-3, Codex, ChatGPT): сравнительный анализ архитектур, методов дообучения, стратегий промышленного развёртывания
- Архитектурный консалтинг: проектирование и оптимизация микросервисных архитектур для заказчиков, аудит существующих решений, рекомендации по выбору технологического стека
- Исследование методов обработки естественного языка (NLP) применительно к русскоязычным текстам: морфологический анализ, извлечение сущностей, построение семантических графов
- Изучение подходов к организации корпоративных баз знаний: онтологическое моделирование, графовые базы данных, методы семантического поиска
- Глубокое освоение языка Rust как основы для построения высокопроизводительных и надёжных ИИ-инфраструктур

Технологии: Rust, Python, NLP, трансформерные архитектуры, GPT-3/Codex/ChatGPT, графовые базы данных, семантический поиск, микросервисная архитектура

ООО Квадком Саратов

Москва, quadcom.ru

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения

Тимлид

Управление разработкой проекта по подряду от ГБУ Инфогород ДИТ Москвы (через субподряд)

Управление командой разработки проекта - в разные периоды более 20 человек

Рефакторинг архитектуры проекта, перепроектирование системы с целью повышения надёжности, отказоустойчивости и скорости работы, снижения накладных расходов на разработку и эксплуатацию, кратного снижения аппаратных требований.

Управление документированием системы

Взаимодействие со смежными подразделениями, представителями заказчиков и внешних информационных систем

Достижения

Полная пересборка команды проекта. От старого состава осталось менее 15% участников, все ключевые сотрудники ушли до моего прихода либо находились в процессе увольнения. Я разработал и реализовал систему подбора кадров необходимой квалификации, построив пайплайн анализа рынка труда с участием кадровой службы и лучших экспертов нашего и смежных отделов, за счёт чего удалось восстановить необходимый штат сотрудников и продолжить разработку проекта.

Тонкая настройка команды, устранение всех значимых конфликтов между сотрудниками, урегулирование конфликтов между стороной заказчика и нами, медиация в согласовании позиций сторон заказчика, подрядчика и субподрядчика (наша сторона). Результат - слаженная работа людей, которые без такой настройки не могли договориться. Главное достижение - отношения между сотрудниками, различными уровнями руководства и прочими стейкхолдерами проекта - в той мере, в какой проходили с моим участием - были избавлены от конфликтов почти полностью. До этого нестроения и конфликты были частыми и мешали работе.

Архитектура проекта была изменена с использования большого количества сложно отлаживаемых самописных микросервисов на использование высоконадёжных и хорошо промышленно оттестированных компонентов, что позволило построить надёжную распределённую систему с хорошей наблюдаемостью (observability) и сосредоточиться на реализации требуемой заказчику функциональности, вместо постоянного багфикса самописных решений

Технологии

Для обработки потоковых данных в системе микросервисов внедрён Benthos, используемый в разных ролях от задачи. В качестве кеша и imdb - DragonflyDb, кратное повышение быстродействия подсистемы. Для обеспечения распределённых транзакций - dtm.

ООО "Мегапейдж"

Москва, megapage.ru/

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения
- Системная интеграция, автоматизация технологических и бизнес-процессов предприятия, ИТ-консалтинг

Начальник отдела внутренних разработок

Работы и обязанности

Управление работой отдела (5-7 человек в разное время)

Анализ крупных гетерогенных клиент-серверных систем

Согласование ТЗ с внутренними и внешними заказчиками компании

Управление разработкой внутреннего ПО и согласование характеристик этого ПО с контрагентами компании

Обеспечение взаимодействия отдела со смежными подразделениями для достижения результата

Подготовка стека решений для рефакторинга унаследованных систем

Проектирование и разработка систем взамен унаследованных

Достижения

Налажено взаимодействие между отделами, устранены неурядицы между сотрудниками и начальством. Климат в переговорах стал ощутимо более здоровым. Удалось добиться бесшовного взаимодействия со смежными отделами, что облегчило разработку новых подсистем - для этого было критично важно получать и передавать актуальную информацию, что ранее было усложнено человеческим фактором - непониманием и отказом идти на контакт. Поставлено на поток документирование разрабатываемых решений - до моего прихода документация была фрагментарной и критично не полной, для ряда подсистем вообще отсутствовала.

Технологии

Для программирования встраиваемых автомобильных контроллеров внедрён и протестирован Rust. Для обработки логов - Vector.dev.

Январь 2018 —
Июнь 2019
1 год 6 месяцев

-

Москва

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения

Разработчик ПО, консультант

Выполнял заказы и предоставлял консультации в сфере разработки ПО. Анализировал, строил и оптимизировал архитектуры систем для заказчиков.

Используемые языки и инструменты:

C/C++

Golang

Lua

Draw.io (Diagrams.net)

PlantUML

Достижения:

Изучил и внедрил в свою практику ряд методик, показавших полезность в управлении командой и переговорном процессе: КПТ/ДПТ, методологию ведения переговоров Джима Кэмпа, тактику ведения переговоров с корпоративными клиентами и малым бизнесом

Август 2014 —
Декабрь 2017
3 года 5 месяцев

Частная компания

Москва

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения

Ведущий разработчик, руководитель направления

Занимался разработкой высоконагруженного сервера приложения

Используемые языки и инструменты:

C/C++

FreeMarker template engine

Java

Lua

Tarantool

Достижения

Удалось получить следующие характеристики сервера:

RPS - 200,000 - 250,000 (тысяч)

Latency - 99% менее 8-9ms

Хост-система - Debian Linux 8.0, x86-64

Приведённая производительность соответствует характеристикам хост-машины
https://ru.hetzner.com/hosting/produkte_rootserver/ex40

Проведён глубокий анализ существующих OpenSource-решений на основе публичных рейтингов и репозиторий кода, а также ручное и автоматизированное тестирование библиотек хранения и обработки данных, выдачи Web-контента.

Применена методология генеративного программирования для создания на основе расширенного FreeMarker генератора исходного кода, реализующего функциональность ORM с быстроедействием, близком к предельно возможному в данной архитектуре.

Исследованы пределы вертикального масштабирования с целью максимального использования мощности одного сервера и дальнейшей кластеризации инстансов - горизонтального масштабирования.

Построена каскадная архитектура потока данных от первичных источников через уровни кэширования к раздаче контента с обратной связью и трассировкой операций.

Сентябрь 2003 —
Июнь 2014
10 лет 10
месяцев

-

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения

Фрилансер

Выполнение различных проектных работ по заказам

Большая часть выполненных за это время работ подпадает под те или иные варианты NDA (соглашение о неразглашении), потому здесь не приводятся.

Используемые языки и инструменты:

SDK различных специализированных программно-аппаратных платформ - ассемблеры, компиляторы, компоновщики, профайлеры, библиотеки, среды моделирования.

Assembler

C/C++

Delphi

Общие и специализированные генераторы конечных автоматов и автоматов с магазинной памятью на основе DSL. Некоторые - на базе открытых свободных решений, некоторые самописные.

Январь 2002 —
Август 2003
1 год 8 месяцев

ООО «Семиком», дочернее предприятие УкрИНТЭИ

Киев

Преподаватель

Курсы:

- Информационные технологии;
- Базы данных и язык SQL;
- Языки программирования C/C++

Январь 1996 —
Октябрь 1998
2 года 10
месяцев

ООО «Коминформсистема»

Киев

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения

Программист-разработчик

Должность – программист в области телекоммуникаций, баз данных и обработки неструктурированных текстовых данных

Операционные среды - DOS, OS/2

Используемые языки программирования – Clipper, Object Pascal/Delphi, C/C++, Assembler

Январь 1994 —
Декабрь 1995
2 года

**Минская центральная районная клиническая больница (МЦРКБ),
отдел АСУ**

Киев

Информационные технологии, системная интеграция, интернет
• Разработка программного обеспечения

Программист-разработчик

Создание оболочек доступа и ведения специализированных баз данных учёта и контроля для нужд МЦРКБ

Используемые языки программирования – Pascal/Object Pascal/Delphi, C/C++, Clipper

Образование

Высшее

1998
Высшее

**Национальный технический университет Украины "Киевский
политехнический институт", Киев**

Факультет прикладной математики, Системный анализ и управление

Навыки

Знание языков

Русский — Родной
Английский — B2 — Средне-продвинутый

Навыки

Организаторские навыки Постановка задач разработчикам Python BPMN
UML Обучение и развитие Agile Scrum Деловая коммуникация
Лидерство Управление командой Деловое общение
Agile Project Management Системный подход Стратегическое мышление
Руководство коллективом Аналитическое мышление Teamleading
Командообразование Управленческие навыки Мотивация персонала

Дополнительная информация

Обо мне

Более 30 лет в ИТ. Прошёл путь от системного программиста до технического директора ИИ-компаний. Последние годы полностью сосредоточен на архитектуре и промышленном внедрении систем искусственного интеллекта.

Строю ИИ-системы, которые работают: от выбора моделей и проектирования RAG-пайплайнов до развёртывания микросервисной инфраструктуры на Rust и вывода в непрерывную эксплуатацию. Владею полным циклом — от исследования до промышленного стенда с 35+ сервисами в продуктивном режиме.

Глубоко разбираюсь в задачах технологической независимости: открытые LLM-модели с локальным развёртыванием в закрытом контуре, альтернативные GPU-платформы (AMD Instinct) с кратным выигрышем в цене, собственный NLP-движок для русского языка без зависимости от западных библиотек.

Умею строить команды и управлять ими. Трижды пересобирав коллективы в кризисных условиях — и трижды сотрудники называли меня лучшим руководителем за время их карьеры. Владею профессиональными навыками медиации и переговоров, что позволяет выстраивать продуктивную работу на стыке интересов заказчиков, подрядчиков и внутренних команд.

Готов возглавить программу ИИ-модернизации крупного предприятия — от формирования стратегии и команды до запуска пилотных проектов и масштабирования в промышленную эксплуатацию.