AI ENGINEER / СИСТЕМИ АВТОНОМНИХ АГЕНТІВ / ВІДДАЛЕНО

Євген Шарий



AI software engineer, який проєктує агентні системи, яким бізнес довіряє гроші і клієнтів.

AI-інженер із діючими клієнтськими системами у продакшені з 2025 року. Перетворюю ймовірнісні моделі на надійну інженерну систему: розробляю серверну архітектуру, інструментів, інтеграцію телефонії, транзакційні машини станів у Postgres та автоматичний аудит кожного діалогу. Бекенд-платформи і фронтенд довкола них теж мої: адмін-портали, дашборди і мобільні застосунки на Next.js, React Native і Postgres. Мої голосові агенти обробляють живі дзвінки та формують замовлення для бізнесів у логістиці й авіації. Вісім років у комерційному маркетингу до інженерії гарантують головне: кожна система проєктується під юніт-економіку, конверсію та безпеку даних.

ЗАРАЗ Супроводжую голосові системи трьох бізнесів і платформу маркетингу у продакшені. Розробляю з агентними AI-інструментами розробника, з власними pre-tool хуками й багаторівневим тестуванням. Відкритий до позиції AI Engineer (віддалено, команди США/ЄС).

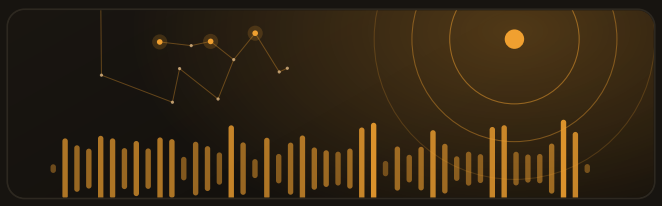
ПОШТА
yevgensharyi@gmail.com

TELEGRAM
t.me/yevgensharyi

САЙТ
yevhensharyi.site/uk/

01 Проєкти

Робота для клієнтів і власні продукти: AI-агенти, голосові системи, бекенд-платформи і мобільні застосунки. Частина вже працює, частина в розробці, дещо внутрішнє. Клієнти під NDA, тому напрямки описано без назв компаній. Кожне інженерне рішення нижче готовий захистити на співбесіді.



01 Логістична компанія в США, далі два авіаційні бізнеси / Лист. 2025 - дотепер

● У ПРОДАКШЕНІ

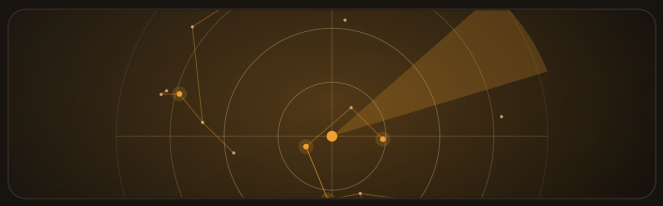
Провідний архітектор та розробник

Голосова система продажів

Автоматичний передзвін за заявкою до 60 секунд, прийом вхідних дзвінків, кваліфікація ліда та створення замовлення в бекенді компанії. ElevenLabs, Twilio, Edge Functions та Postgres, панель для сейлзів і LLM-суддя для аудиту транскриптів. Реалізовано для логістичної компанії у США, далі масштабовано на оператора бізнес-джетів та авіаційну платформу.

01 Друге покоління системи: бізнес-логіка з візуальних воркфлоу перенесена на типізовані Edge Functions, стан централізовано в Postgres, тестове середовище відтворює продакшен.
02 Для оператора бізнес-джетів розгорнуто за 8 днів без зовнішніх оркестраторів; для передплатного авіаційного сервісу адаптовано за 1 день з прайс-листом у RAG і контролем паритету.
03 Серверний пріоритет даних: тарифи, знижки та статуси замовлень контролює база даних; агент читає їх через інструменти тільки на читання, тому маніпуляція ціною з боку співрозмовника неможлива.
04 Ідемпотентне виконання: кожен виклик функції має унікальний ключ, тому мережеві повтори від голосового провайдера не створюють повторних транзакцій.
05 Безпека вихідного зв'язку: списки DNC, часові вікна контакту та атомарне блокування запису в Postgres проти одночасного набору одного контакту.
06 Контроль якості: окрема модель інспектує кожен дзвінок і фіксує порушення лише з цитатою з транскрипту; термінові випадки йдуть команді в Telegram.
07 Тримовна підтримка: англійська, українська та російська, з точками відкату перед кожним релізом.

ElevenLabs Agents · Claude · Supabase · Deno Edge Functions · Postgres · Next.js · Twilio · Pipedrive



02 Чартерна платформа для операторів, брокерів і власників літаків, США / Трав. 2026 - дотепер

● У ПРОДАКШЕНІ

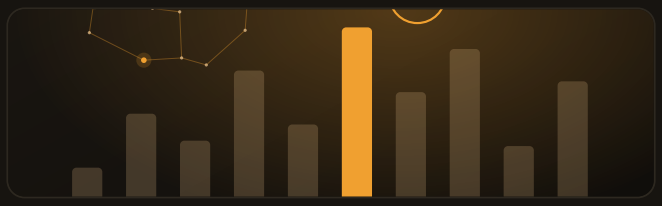
Модуль пілотів: бекенд і фронтенд

Пошук і аутріч пілотів

Мультитенантна чартерна платформа для операторів, брокерів і власників літаків; модуль пілотів у ній зробив я, бекенд і фронтенд. Вакансія розбирається на критерії, кандидати збираються з реєстрів, job-бордів і сервісів збагачення контактів, кожен перевіряється за реєстром FAA і ранжується під вакансію, далі йде ланцюжок листів з власного домену компанії. Відповіді та відгуки з job-бордів збираються автоматично, AI читає їх і визначає інтерес і кваліфікацію пілота.

01 Спільна база пілотів компанії: вердикт FAA з причиною, наліт, type ratings, контакти, позначки do-not-contact і вердикти по вакансіях в одній картці.
02 Пошук: критерії вакансії, прогони пошуку, верифікація за FAA, збагачення контактів через Apollo.io і LinkedIn, ранжування під роль.
03 Аутріч: кампанії з окремого домену з DKIM, SPF і DMARC; відповіді повертаються в систему, модель класифікує інтерес і кваліфікацію.
04 Відгуки: вакансія публікується на job-борді через браузерний контейнер, відгуки збираються і верифікуються, креденшели job-борду зашифровані.
05 Безпековий огляд перед релізом: прибрано eval() на виводі моделі, ендпоінт з персональними даними без автентифікації і відкритий поштовий релей; приймальні тести і dry run перед кожним модулем.

Python · FastAPI · PostgreSQL · Celery · OpenAI · Playwright · Next.js · Apollo.io · FAA registry · Mailgun



03 Портфель продуктів технологічного холдингу в США / Вер. 2026 - дотепер

● У ПРОДАКШЕНІ

Архітектор платформи та Lead Backend

Маркетингова платформа

Мультитенантна платформа, де агенти готують стратегії, статті, метадані для маркетплейсів, рекламу та відповіді клієнтам для портфеля продуктів холдингу; фахівці редагують у панелі керування, серверні шлюзи випускають назовні лише верифіковане. На цій же платформі працює done-for-you AI marketing agency: AI-агенти і маркетологи ведуть SEO, Google Ads, соцмережі, email і відгуки 24/7, а клієнт бачить кожен пост, лід і результат наживо у власному дашборді. У продакшені з жовтня 2026.

01 Захист від подвійних публікацій: публікація під row lock, виклик зовнішніх API поза транзакцією.
02 Транзакційний бюджет: ліміти витрат перевіряються в тій самій транзакції, що й затвердження; сценарії перевірені стрес-тестами на одночасний доступ.
03 Мультитенантна ізоляція: права перевіряє сервер на кожному запиті; спроба доступу до чужого продукту повертає 404.
04 Захист токенів: ключі рекламних кабінетів шифруються окремим системним ключем; інтерфейс лише записує токен і не читає його назад.
05 Санітизація недовірених даних: відгуки, контент конкурентів і відповіді моделей проходять схематичну валідацію; моделі не можуть обирати фінансові дії.
06 Культура постачання: конектори в mock-режимі локально і в тестах, релізи через чергу перевірок з детермінованим оновленням.

Python · FastAPI · SQLAlchemy · PostgreSQL · Celery · Claude · Next.js · shadcn/ui · GitLab CI



04 Голосовий AI-продукт для малого бізнесу в США / Трав. 2026 - дотепер

● У ПРОДАКШЕНІ

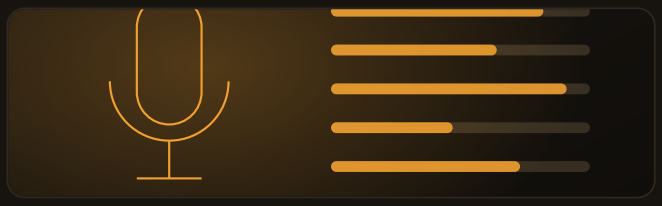
Архітектор бекенду й телефонії

AI-рецепціоніст за URL сайту

Перетворює сайт компанії на голосового оператора за лічені секунди: конвеєр сканує сторінки, видобуває правила бізнесу в pgvector, конфігурує маршрутизацію Twilio і відповідає клієнтам без окремого агента й номера під кожного користувача.

01 Динамічний пул номерів: вхідні дзвінки маршрутизуються через спільний пул, профіль бізнесу прив'язується на старті сесії.
02 Двоетапне вилучення даних: асинхронний скрапер збирає факти, валідаційна модель відсікає галюцинації, резервні моделі доводять сесію до дзвінка.
03 Ізоляція клієнтів: окремі векторні розділи в pgvector, ліміти викликів, явна згода абонента, жорсткий ліміт бюджету.
04 Модульні інтерфейси: бекенд оркестрації, панель моніторингу, типізовані API-контракти для фронтенду.

ElevenLabs · Claude · OpenAI · Supabase pgvector · Twilio · Next.js



05 Власний мобільний продукт, команда з двох інженерів / Лип. - серп. 2026

● У РОЗРОБЦІ

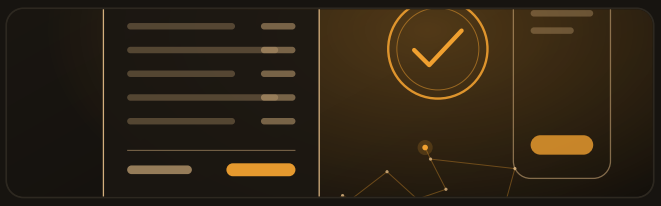
Lead Developer & Systems Engineer

Голосовий тренажер співбесід реального часу

Усне технічне інтерв'ю під реальні вакансії для IT-фахівців: субсекундна затримка спілкування, структурована оцінка відповідей і детальний звіт після сесії.

01 Стримінг через WebSocket напряму до Gemini Live з калібруванням детекції мовлення, відновленням сесії та українськими фонетичними підказками.
02 Контейнерний релей на Cloud Run тримає з'єднання до 45 хвилин замість обривів за таймаутами edge-функцій.
03 Безпека: RLS у кожній міграції, захист від SSRF при парсингу вакансій, тести на стійкість до prompt injection.
04 Кросплатформний застосунок на Expo React Native з підписками через RevenueCat.

Expo · React Native · TypeScript · Gemini Live · Vertex AI · Supabase · Cloud Run · RevenueCat



06 Власний мобільний продукт / Квіт. 2026 - дотепер

● У ПРОДАКШЕНІ

Власник продукту і розробник

Мобільний застосунок для рахунків фрилансерів

Mobile-first застосунок для рахунків фрилансерів: пропозиція, договір і рахунок в одному потоці, надсилання з телефону, відстеження відкриттів і нагадування про прострочення. Наявний рахунок імпортується з PDF або фото за один виклик моделі. Дві візуальні мови: застосунок для фрилансера і чистий PDF для клієнта.

01 Пропозиція, договір і рахунок мають одну модель даних, тому зміна в пропозиції перетікає в рахунок.
02 Імпорт з фото або PDF: один виклик зі структурованим виводом, валідація zod до збереження.
03 Supabase: auth, база, storage і edge functions з RLS; підписки через RevenueCat.
04 PDF рендериться на пристрої через expo-print, графіки у victory-native, анімація в Reanimated і Skia.

React Native · Expo · TypeScript · Zustand · React Hook Form · zod · Supabase · RevenueCat · Reanimated · Skia



07 Власний продукт / Серп. 2026

● НА ТЕСТУВАННІ

Solo Developer

Telegram Mini App з внутрішньою економікою

Інтерактивний 3D-застосунок з оплатою в Telegram Stars, автоматичною інтерпретацією розкладів і генерацією медіа на орендованих GPU.

01 Сліпе тестування моделей: робочі LLM обрано за порівняльним тестом, з автоматичним перемиканням на резервну.
02 Ідемпотентність платежів: покупки прив'язані до charge id провайдера; ідентифікатор користувача лише з криптографічного контексту Telegram.
03 Zero-trust база: політики RLS і обмеження доступу в тих самих міграціях, що й таблиці.
04 Медіа-пайплайн: арт через Gemini з локальним резервом FLUX ControlNet; короткі відео через Remotion і headless ffmpeg на орендованих GPU.

TypeScript · Deno · grammY · React · three.js · Supabase · Vertex AI · ComfyUI · Vercel



08 Агенція нерухомості, Київ / Квіт. 2026 - дотепер

● У ЩОДЕННІЙ РОБОТІ

Архітектор та розробник систем

Автономний робочий простір агенції

Агентний AI-інструмент розробника, підключений до щоденної роботи невеликої агенції: CRM клієнтів, об'єктів і угод, яку агент читає та оновлює, пошук по порталах і щоденний моніторинг, порівняльна оцінка продажу й оренди, договори, пропозиції власникам і презентації для клієнтів, згенеровані з даних. Використовується брокерами щодня з квітня 2026 року.

01 CRM для агента: JSON-таблиці плюс Markdown-картки зі схемою зв'язків; кожна зміна валідується (зв'язки існують, ID унікальні, цифри перераховані).
02 Гібридний збір даних: API порталів плюс Playwright для неіндексованих оголошень; оцінка порівнює кілька методик.
03 Генератори: договори, пропозиції власникам, презентації об'єктів та інвесторські з окупністю, у HTML і PDF.
04 Контроль публікацій: автоматичні чернетки, публікація лише після затвердження людиною.

agentic AI developer tool · Python · Playwright · SQLite

02 Інженерна методологія

Дисципліна розробки автономних систем та агентного програмування.

> Агентні AI-інструменти розробника реалізують код у межах визначених мною архітектурних меж. Я проєктую системну архітектуру, формулюю специфікації, налаштовую автоматичні pre-tool хуки та особисто відповідаю за надійність у продакшені. Швидкість зростає у рази, але відповідальність за результат належить інженеру.

ПРОЦЕС

01 Скоуп та юніт-економіка

Виявлення вузьких місць у бізнес-процесі. Розрахунок бюджетів затримок, витрат на токени та конверсії до написання коду.

02 Специфікації та межі

Фіксація інтерфейсних контрактів, машин станів і приймальних критеріїв до підключення генеративних моделей.

03 Контрольована розробка

Агентний AI-інструмент розробника пише код в ізоляції pre-tool хуків. Кожна гілка проходить автоматизований безпековий аудит.

04 Експлуатація у продакшені

Запуск на реальному трафіку. Безперервний аудит транскриптів окремою моделлю та миттєві сповіщення про збої.

ПРИНЦИПИ

AI там, де він справді потрібен

Детерміноване правило б'є виклик моделі щоразу, коли правила достатньо. Моделі дістаються частини, де проблема саме в мові.

Спершу система, потім екран

Таблиці, стани і сценарії відмов першими. UI іде останнім і успадковує обмеження.

Останнє слово за сервером

Ціни, ліміти і статуси замовлень є обмеженнями бази даних. Агент їх читає і ніколи не встановлює, хоч би що сказав абонент.

Промпт і інструменти під моїм контролем

Структурований вивід замість вільного тексту, промпти, заземлені на схему, замість здогадок, підтвердження перед будь-якою деструктивною дією. Жодного ланцюга абстракцій, який я не зможу відлагодити.

Будувати для реального світу

Повтори, ключі ідемпотентності, таймаути, dead-letter черги і паритет staging. Вебхуки приходять двічі, з'єднання рвуться, а система має працювати далі.

Триматися ремесла

Я читаю код, який пише агент, переглядаю кожен diff і зберігаю повторювані процедури як версіоновані skills, щоб наступний проєкт починався з того, що спрацювало.

03 Досвід

Лист. 2025 - дотепер

AI Engineer, Technical Lead із систем автономних агентів

Технологічний холдинг у США: логістичний софт, голосові AI-продукти, корпоративні рішення

Відповідаю за повний життєвий цикл AI-систем холдингу та його клієнтів: від проєктування архітектури та інженерії даних до щоденної експлуатації та безпекового аудиту. Керую розробкою голосових агентів, мультитенантних платформ і валідаційних пайплайнів.

01 Спроєктував і запустив у продакшен голосову систему продажів для логістичної компанії у США, що обробляє живі вхідні та вихідні дзвінки клієнтів.
02 Масштабував голосову архітектуру на два авіаційні бізнеси (оператор бізнес-джетів за 8 днів, SaaS-платформа за 1 день) зі 100% збереженням функціонального паритету.
03 Створив з нуля мультитенантну платформу генерації маркетингу: автономні агенти, інтерфейс затвердження та серверні транзакційні шлюзи. У продакшені з жовтня 2026.
04 Провів безпековий аудит діючої голосової інфраструктури, виявив та усунув ендпоінти без автентифікації і відкрите збереження токенів.
05 Побудував пайплайн рекрутингу пілотів: верифікація ліцензій за базою FAA, приймальні тести та безпечний поштовий аутріч.
06 Провів технічну експертизу пропозицій підрядників щодо парсингу авіаційних реєстрів за даними FAA; висновки запобігли запуску нежиттєздатної розробки.
07 Спроєктував архітектуру SaaS-рецепціоніста: конвеєр вилучення даних із сайтів, векторна база в pgvector та панель адміністрування.
08 Оцінював строки реалізації нових AI-продуктів та вів щотижневу технічну звітність для топменеджменту.

2024 - 2025

Незалежний розробник AI-систем

Малий бізнес, партнерські дизайн-студії, власні SaaS-продукти

Розробка відмовостійкої автоматизації та окремих AI-продуктів. Повнофункціональні застосунки на TypeScript та Python поверх Supabase із дотриманням Row-Level Security з першої міграції.

01 Побудував надійні інтеграції між CRM, Telegram, поштовими сервісами та наскрізною аналітикою.
02 Спроєктував воронки кваліфікації лідів, що автоматизують шлях клієнта від першого контакту до оплати.
03 Розробив модуль білінгу для партнерської студії: рев'ю pull request, миттєве OCR-розпізнавання рахунків за один виклик моделі.
04 Створив браузерного голосового асистента з анімованим аватаром на Gemini Live з виконанням інструментів на сервері.
05 Зібрав робочий прототип платформи нерухомості на Twenty CRM, LangGraph та власному MCP-сервері.

2016 - 2024

Керівник digital-маркетингу / Growth Lead

Виробничі та ритейл-бренди преміум-сегмента, Київ

Керував залученням клієнтів та інфраструктурою маркетингу: платний трафік, воронки продажів, автоматизація обробки лідів і вебаналітика. Цей восьмирічний комерційний фундамент - причина, чому кожна створена мною AI-система підпорядкована юніт-економіці, довірі клієнтів та поверненню інвестицій.

01 Керував мультиканальним платним трафіком у Google Ads та Meta з наскрізною аналітикою в CRM.
02 Будував чат-боти в соціальних мережах, що автоматизували кваліфікацію та первинну маршрутизацію лідів до відділу продажів.
03 Об'єднав розрізнені канали звернень у єдиний централізований пайплайн, суттєво скоротивши час першої відповіді клієнту.
04 Організував виробництво креативів та коротких відео від розробки концепцій до фінального монтажу.

04 Навички

Підтверджений стек розробки та реальний досвід експлуатації.

AGENTIC AI

API моделей та архітектура

Claude, OpenAI Responses API та Gemini у продакшені. Вибір моделей на основі сліпих бенчмарків, автоматичні резервні контури та жорсткий контроль лімітів токенів.

Claude Opus/Sonnet | OpenAI Responses API | Gemini Live | Vertex AI

Агенти та виклик інструментів

Серверні інструменти голосових і текстових агентів із обов'язковою ідемпотентністю за ID сесії. Власні цикли виконання через WebSocket без зайвих проміжних фреймворків.

ElevenLabs Agents | Function Calling | Gemini Live WebSockets

Мультиагентні системи

Двоконтурний нагляд: незалежні моделі-судді, факт-чекінг контенту та ізоляція гілок рев'ю.

LLM-as-judge | fact-checker agents | read-only reviewers

Промпт-інжиніринг та схеми

Системні інструкції з детермінованими переходами станів та розміткою джерел. Валідація виводу за схемами JSON Schema з перевіркою полів регулярними виразами.

Structured Outputs | JSON Schema | State Mapping

RAG та векторний пошук

Ізольовані векторні бази в pgvector, семантичний пошук кандидатів у Postgres та оперативний пошук прайс-листів у RAG. Пріоритет структури даних над передчасним fine-tuning.

pgvector · OpenAI Embeddings

Оцінка якості та тестування

Автоматичний аудит кожного дзвінка суддею LLM із вимогою обов'язкового цитування транскрипту. Регресійні прогони, змагальні тести агентів та сліпі порівняння моделей.

LLM-as-Judge · Регресійні тести · Змагальні тести

Безпека та захист AI

Нульова довіра до тексту абонента, даних із вебсторінок і відповідей моделей. Аудит коду з ліквідацією eval(), незахищених ендпоінтів та prompt injection. Дотримання TCPA та DNC у коді.

Захист від prompt injection · Row-Level Security · Захист від SSRF · TCPA / DNC

AI governance

Категорії ризиків для агентів; обов'язкове підтвердження людиною будь-яких платежів, розсилок і записів у БД.

risk tiers · human approval gates

Observability

Очищення телеметрії від PII, незмінний аудит-лог подій, моніторинг затримок і диспетчеризація помилок.

PII-sanitized telemetry · audit trails · latency percentiles

Вартість та затримки

Посекундний облік витрат на телефонію та токени. Ліміти навантаження, бюджетування виводу, резервні моделі та заміри затримок вебхуків до прийняття архітектурних рішень.

Облік витрат · Rate Limits · Субсекундна затримка

Агентна розробка та MCP

Агентний AI-інструмент розробника як основне робоче середовище: кастомні команди агентів, pre-tool хуки та процедури-скіли. Сервери MCP для Supabase, Playwright, Context7, Jira та власний сервер індексації.

agentic AI developer tool · Model Context Protocol (MCP) · Хуки

ГОЛОС, ДАНІ І ПОСТАЧАННЯ

Голосовий конвеєр

Передавання мовлення в реальному часі через ElevenLabs Agents та Gemini Live: перехоплення ініціативи, налаштування VAD, утримання сесії. Телефонія на Twilio: реєстрація A2P 10DLC, toll-free верифікація та розпізнавання автовідповідачів.

ElevenLabs · Gemini Live · Twilio · VAD і перехоплення

Автоматизація процесів

n8n як код через SDK з публікацією виключно за дозволеними ID сценаріїв. Надійні вебхуки, експоненційні повтори та обробка помилок у CRM, пошті та платежах.

n8n як код · Pipedrive · Telegram Bot API · Mailgun · Stripe

Шар даних на Postgres

Supabase Postgres із політиками Row-Level Security з першої міграції, розмежування прав, атомарне блокування рядків, валідація через CHECK-констрейнти та задачі в pg_cron.

Postgres · Supabase · Row-Level Security (RLS) · pg_cron

Надійність у продакшені

Точне відтворення продакшену в dev, фіча-прапорці без редеплою, тестові прогони перед записом та точки відкату. Автоматичні сповіщення та контури розбору діалогів.

Docker · Cloud Run · Vercel · GitHub Actions

Мультимодальні системи

Розпізнавання документів за один виклик моделі, пайплайни генерації зображень із локальним резервом, транскрипція аудіо та автоматична генерація відео на орендованих GPU.

Gemini Vision · FLUX ControlNet · LTX Video · ffmpeg

Python та TypeScript

Продакшен-код на Python (скрапінг, оцінка вартості, векторний пошук) і TypeScript (Edge Functions, адмін-панелі, n8n як код). Фундаментальна база фронтенду за програмою EPAM (2022).

Python · TypeScript · Deno · Next.js

ІНСТРУМЕНТАРІЙ ЗА РІВНЯМИ

Моделі та API

Claude Opus, Sonnet and Haiku · Anthropic web_fetch · OpenAI Responses API · GPT-4.1 family · Structured Outputs · text-embedding-3-small · Gemini Flash and Pro · Gemini Live · Gemini TTS and image models · Vertex AI · Perplexity sonar

Голос

ElevenLabs Agents · server tools and webhooks · language presets · Twilio numbers and SMS · A2P 10DLC · toll-free verification · answering-machine detection · Gemini Live WebSocket · VAD and barge-in · whisper · Kokoro · Simli

Пошук і дані

Supabase Postgres · RLS and migrations · pg_cron and Vault · Edge Functions on Deno · pgvector · ElevenLabs KB in RAG mode · SQLite

Оркестрація та агенти	n8n workflow-as-code · @n8n/workflow-sdk · n8n MCP · agent teams of the agentic AI developer tool · hooks and skills · MCP: Supabase, Playwright, Context7, Jira · LangGraph (prototype)
Застосунки та інфраструктура	TypeScript · Python · React · Next.js · Expo and React Native · Vite · grammY · Telegram Mini Apps · three.js · Vercel · Cloud Run · Docker · GitHub Actions · RunPod · Mailgun · Pipedrive · Playwright
Генеративні медіа	ComfyUI · LTX video · FLUX ControlNet via mflux · Remotion · ffmpeg

05 Бекграунд

МОВИ	Англійська: робоча письмова мова з командою в США (специфікації, звіти, листування). Українська (рідна), російська.
ОСВІТА	Магістр, кафедра електронних систем, Донбаський державний технічний університет (2008-2013). Менторська програма фронтенд-розробки, EPAM UpSkill та Strategeast (червень-листопад 2022, диплом ID 913874).
ПІДТВЕРДЖЕННЯ КОДУ	Комерційний код захищений взаємним NDA, тому репозиторії на GitHub приватні. Замість коду клієнтів на технічному інтерв'ю проводжу детальний розбір архітектури діючих систем, демонстрацію власних продуктів або інтерактивний live-build з агентним AI-інструментом розробника.

КОНТАКТИ

yevgensharyi@gmail.com

t.me/yevgensharyi yevhensharyi.site/uk/