



Valentin Lapchevskiy

AI-Native Product Engineer — full-stack + LLM / агентные системы, end-to-end

Собираю продукты целиком — backend, frontend и LLM-слой — оркестрируя коддинг-агентов с производительностью команды.

КОНТАКТЫ

Аргентина · UTC-3

Remote · оверлап с US-часами

v.lapchevskiy@gmail.com

Telegram @gen4sp

Upwork — Top Rated

App in the Air ↗

ГЛАВНОЕ

- **7M+** установок · featured by Apple
- **\$300K+** · 61 проект · Top Rated
- **5** AI-систем + research (2024–26)
- Agent-native подход

КАК Я РАБОТАЮ

- Превращаю размытое ТЗ в **готовые системы** — беру неопределённость на себя.
- **Сначала система**: карта пути и трейд-оффов до кода.
- **Неугомонный прототипировщик** — довожу идеи до рабочих экспериментов, быстро.
- Спокоен при меняющихся требованиях; **минимум оверхеда**, remote-friendly.

НАВЫКИ

AI / LLM

Agentic systems (tool-calling, OODA), Vercel AI SDK, OpenAI · Anthropic · Gemini, RAG / pgvector, multimodal (vision, STT), structured outputs, cost-aware LLM engineering (cost ledgers, per-tier guards), LLM eval &

КРАТКО

AI-native продуктовый инженер: собираю системы целиком — backend, frontend и LLM/агентный слой — оркестрируя коддинг-агентов с производительностью команды. 15 лет строю продукты: co-founder **App in the Air** (**7M+ установок**, featured by Apple) и независимый инженер, **Top Rated** на Upwork с **\$300K+ за 61 проект** (~100% success). Текущий фокус: production-системы на LLM-агентах, мультимодальные LLM-пайплайны, интерпретируемость ИИ.

ИЗБРАННЫЕ ПРОЕКТЫ

B2B-платформа для коллабораций брендов и креаторов

2026 – наст.

Контракт · senior full-stack / AI-инженер · под NDA

- Проектирую **workflow-движок на Temporal** и **control plane для AI-провайдеров** (cost ledger, rate-limiting, circuit breakers, webhook-replay) для мультидоменной платформы.
- Выстроил **domain-driven монорепо** (переносимое platform-ядро + бизнес-домены) с P0–P5 smoke-гейтами и полным observability (Grafana / Loki / Alloy).
- Строю **skills-based LLM-генерацию** (18+ доменных skill'ов — обогащение бренда, генерация брифов/сценариев, антифрод) с guardrails и quality-eval.

Agentic-система поддержки

2026

Контракт · AI-инженер · B2B-компания (GPU-инфраструктура), под NDA

- Построил end-to-end **систему поддержки на LLM-агентах** (Vercel AI SDK, OODA tool-calling loop) для Telegram / Email / Web — классификация intent, извлечение полей, авто-тикетинг; собрано за < 2 месяцев (3 прототипа → production).
- Реализовал **мультимодальный ввод** (vision + Whisper STT) и контекстную память поверх PostgreSQL/pgvector + Redis.
- Внедрил **самоулучшающийся eval-loop** (45+ реальных кейсов → LLM-правка промптов → regression-рейтинг в CI) и production-инфра (Docker, Langfuse, retry-очереди, graceful shutdown).

observability (Langfuse), agent-driven dev; interpretability, tensor networks.

BACKEND

Node.js / TypeScript, Fastify / Express, Temporal.io, Redis / BullMQ, PostgreSQL (Drizzle, pgvector), MongoDB.

FRONTEND

React, Vue / Nuxt, TypeScript, Tailwind, design systems, FSD.

MOBILE

React Native, Flutter, Electron.

DATA / ML

Python, PyTorch, scikit-learn, data pipelines, scraping.

INFRA

Docker, CI/CD, observability (Prometheus / Grafana / Jaeger / OTel), GCP / AWS.

ЕЩЁ

Solidity / Web3; creative tech (TouchDesigner, GLSL/HLSL).

ОБРАЗОВАНИЕ И ЯЗЫКИ

Магистратура, когнитивные науки

НИУ ВШЭ (2018–2019, не окончена).

Бакалавр, прикладная информатика

Российский университет кооперации (2000–2005).

English (professional) · Russian (native) · Spanish (B1, в процессе).

Аудируемая система корпоративной памяти

2026

Контракт · AI R&D · B2B-компания (GPU-инфраструктура), под NDA

- Спроектировал и прототипировал **аудируемую систему корпоративной памяти** — неизменяемый raw-слой + пересоздаваемый structured-слой, дающие сквозной provenance от любого AI-вывода к первоисточнику.
- Построил **LLM-извлечение knowledge-graph** (сущности / связи / факты) с подключаемыми доменными «призмами» и type-guard'ами поверх PostgreSQL/pgvector, отдача через MCP — передан как прототип для дальнейшей доработки.

Viracle — AI-аналитика виральности видео

2025 – 2026

Founder / Solo Engineer

- Спроектировал **event-driven backend** (Fastify, Temporal.io, 4 микросервиса), который скрейпит, анализирует и оценивает короткие видео (TikTok / Reels / Shorts).
- Построил **мультимодальный анализ видео** на Gemini structured outputs (хук, пейсинг, СТА, ниша, sentiment) и **7-уровневый Python ML-пайплайн** (корреляции, кластеризация, детекция аномалий → контент-рекомендации).
- Построил production-grade инфру: PostgreSQL/Drizzle (20+ таблиц), Redis Streams, полный observability (Prometheus / Grafana / Jaeger), биллинг и auth.

Noracle — AI-SaaS интерпретации снов

2026

Founder / Solo Engineer · Живое демо ↗

- Запустил **живой монетизированный** Telegram-AI-продукт — интерпретация снов + развивающийся психо-портрет пользователя (оплата Telegram Stars; **D30-retention 38%** среди платящих).
- Собрал **многоэтапный LLM-пайплайн** — агентный роутинг режимов, мультипровайдерная генерация картинок, per-tier cost-guards, structured outputs (~\$0.02/сон).
- Выполнил **zero-downtime идемпотентную миграцию** 3 485 пользователей / 6 340 записей с cold-backup.

Независимое AI-исследование — интерпретируемость и эффективные архитектуры

2025 – 2026

Самостоятельно

- 200+ экспериментов по извлечению **точной, верифицируемой логики** из обученных сетей (lossless NN → формула → Verilog), в отличие от приближённых методов (SHAP/LIME).
- Результаты: рекуррентная булева ячейка на **40 параметрах** решает плотный parity-128 (100% на held-out, 5/5 сидов) там, где GRU остаётся на уровне случайности; точное извлечение **ANF → Verilog** со 100% совпадением формулы и модели; булев копроцессор в **0.08% параметров** даёт замороженному GPT-2 многоразрядное сложение, недоступное ему zero-shot.
- Строгая методология (multi-seed, поправка Бонферрони, 3 раунда верификации) с задокументированным логом опровергнутых гипотез.

GRAB-A-WORD-II — real-time мультиплеер, словесная игра

2024

Founder / Solo Engineer

- Построил соло **full-stack real-time мультиплеер** (Nuxt/Vue, Socket.io, Redis/BullMQ) с локализацией RU/EN и Telegram-интеграцией.
- Создал модульную дизайн-систему (Storybook, 26 компонентов) и словарный пайплайн (Apify + OpenAI) в монорепо из 10 модулей.

ОПЫТ

Независимый full-stack инженер и продуктовый консультант

2015 – наст.

Upwork · Top Rated

- **\$300K+ заработано за 61 проект / 4 446 часов, ~100% success;** клиенты преимущественно US, Canada, Europe.
- 30+ MVP и proof-of-concept; глубокое участие в продуктовых решениях и техническом консультировании.

Co-founder · продукт и full-stack инженер

2011 – 2014

Empatika — App in the Air

- Co-основал travel-ассистент и три года вёл продукт — пивот с геочата на помощь в аэропортах, рост **0 → 600K пользователей** при маркетинге меньше \$15K.
- Featured by Apple — **Best New Apps в 120 странах**, Editors' Choice, предустановка в Apple Store; сегодня у приложения **7M+ установок**.

Creative Technologist · Visual Programmer

2008 – 2015

INTY · TNT Broadcast Network

- Интерактивные инсталляции и real-time визуальные системы (TouchDesigner, Ventuz, GLSL, Arduino).