

# SALEH NOMAN

Data Scientist · Инженер по данным и автоматизации · Веб-скрапинг / ETL

📍 Москва, Россия    📞 +7 900 159-71-98    ✉️ salehnoman22@gmail.com    🔗 linkedin.com/in/salehnoman    📁 github.com/SalehNoman-MISIS

## ★ О СЕБЕ

Выпускник магистратуры Data Science (красный диплом с отличием, GPA 5.0/5.0). Создаю сквозные системы машинного обучения, веб-скрапинга и ETL. В магистерской диссертации разработал гибридный движок (правила + ML), определивший ATS-систему каждой компании Fortune 500 со 100% охватом. Руководил сбором данных о вакансиях в Workgaze и создал официальную библиотеку шаблонов скрапинга для RTILA Studio. Английский — свободный, арабский — родной, русский — средний.

## ★ СИЛЬНЫЕ СТОРОНЫ

Python и данные	95%	Машинное обучение	85%
Веб-скрапинг (Playwright, Selenium)	93%	Автоматизация (RTILA, n8n)	84%
Разработка с ИИ (Claude / LLM)	92%	Full-stack (Flask/FastAPI/React)	82%
ETL и дата-инжиниринг	88%	SQL и базы данных	80%

## <> КЛЮЧЕВЫЕ НАВЫКИ

**Программирование:** Python, JavaScript, Java, C#, SQL

**Данные и ML:** pandas, NumPy, scikit-learn, XGBoost, CatBoost, LightGBM, SHAP, feature engineering, кросс-валидация

**ИИ и LLM:** Разработка с помощью Claude и GPT SDK, агентные инструменты, prompt engineering, RAG-пайплайны

**Скрапинг и автоматизация:** Playwright, Selenium, Scrapling, BeautifulSoup, RTILA Studio, обход антибот-защиты, пагинация, n8n

**Веб и full-stack:** Flask, FastAPI, React, Node.js / Express, Supabase, REST API, Alpine.js

**Дата-инжиниринг:** ETL, нормализация, дедупликация, валидация, отслеживание актуальности; API Directus и Google Sheets

**Базы данных и инструменты:** PostgreSQL, MongoDB, Firebase, SQLite, Git, pytest, GitHub Actions CI/CD, PyInstaller

## 📁 ОПЫТ РАБОТЫ

### ● Менеджер по агрегации данных и автоматизации (ETL Lead)

сен. 2025 – наст. время

#### Workgaze — удалённо

- Полный цикл скрапинга и ETL для вакансий из нескольких источников: поиск → извлечение → очистка → нормализация → загрузка в Directus.
- Разработал платформу скрапинга (FastAPI + Playwright) с агентным конструктором источников, генерирующим валидные конфиги по любому URL; 200+ тестов.
- Внедрил проверки качества — дедупликацию, нормализацию времени, валидацию схем, отслеживание актуальности — с планированием, логами и оповещениями.

### ● Специалист по автоматизации данных и поддержке приложений

янв. 2023 – 2026

#### RTILA Corporation — удалённо

- Создал официальную библиотеку шаблонов скрапинга и ботов для RTILA Studio: Amazon, AliExpress, Etsy, LinkedIn, Google Maps и др.
- Реализовал устойчивое извлечение (CSS/XPath, пагинация, обход антибот-защиты); вывод в Google Sheets, MySQL, Directus и вебхуки.
- Составлял руководства, обучал команды поддержки и вёл еженедельные релизы с управлением изменениями и анализом влияния.

### ● Специалист технической поддержки

сен. 2019 – 2023

#### Yemeni Socialist Party — Сана

- Обновил ОС и ПО в организации, обеспечил защиту оборудования и данных, достиг 100% решения проблем.
- Повысил эффективность ~50% за счёт обновлений; обучил сотрудников, увеличив производительность ~40%.

### ● Преподаватель английского языка (волонтёр)

2018 – 2020

#### Сана, Йемен

- Обучал старшеклассников грамматике и технике экзаменов; 100% сдали, ~60% на отлично.

## <> ИЗБРАННЫЕ ПРОЕКТЫ

### Система определения ATS — магистерская диссертация

Python    scikit-learn    XGBoost    Playwright    SHAP

- Скрапинг и разбор карьерных страниц всех 500 компаний Fortune 500 устойчивым каскадом Playwright/Selenium/BeautifulSoup с кэшем.
- 4-уровневый каскад с финальной моделью RandomForest на 1651 признаке и объяснимостью SHAP.
- 100% охват (500/500) по 30 ATS-провайдерам; 90 тестов, k-fold CV, цикл active learning.

## Платформа скрапинга вакансий с ИИ

Python FastAPI Playwright Alpine.js

- Многоисточниковый скрапер, читающий job-борды из YAML / CSS-конфигов с пагинацией и stealth-режимами.
- Агентный конструктор источников управляет браузером через SDK Claude/GPT и генерирует валидные конфиги по любому URL.
- Нормализация и загрузка в Directus с дедупликацией; 200+ автотестов.

## Платформа обогащения контактных данных

Python Flask React Playwright Scrapling

- Full-stack приложение: извлекает email, телефоны и соцсети из коллекций CMS или списков URL с определением языка.
- Поиск сайтов через Google Maps, SMTP-верификация и экспорт JSON/CSV/TSV/JSONL через REST API.
- Поставляется как лицензируемое Windows-приложение с подписанными лицензиями на клиента.

## Платформа браузерных агентов автоматизации

Node.js Express Playwright

- Визуальный Flow Builder (26 типов шагов: переменные, условия, циклы, обработка ошибок) с семьёю агентами.
- Единый endpoint n8n/HTTP (сохранённый флоу, inline-шаги, legacy) с управлением сессиями и логами в реальном времени (SSE).

## Агрегатор цен биржи ссылок

Python анти-детект браузеры Google Sheets API

- Проверяет домен на пяти площадках, разбирает наличие и цены и пишет лучший вариант в Google Sheets.
- Шифрование учётных данных (PBKDF2 + Fernet); установщик Windows «в один клик» через GitHub Actions CI.

## ML-модель предсказания пола

Python CatBoost LightGBM проект курса, 1 месяц

- Сквозной пайплайн: сбор, очистка, feature engineering, обучение, оценка.
- CatBoost/LightGBM с k-fold CV и настройкой порогов под macro-F1, сравнение с базовыми моделями.

## 🎓 ОБРАЗОВАНИЕ

### Магистр, Data Science — красный диплом с отличием, GPA 5.0/5.0

НИТУ МИСИС, Москва

2024–2026

### Бакалавр, Информационные технологии (92%)

University of Modern Sciences

2017–2021

### Русский язык и литература

МФТИ, Москва

2023–2024

### Диплом продвинутого уровня английского

Seeds Education

2016–2017

## ✔ СЕРТИФИКАТЫ

- Google IT Support Professional Certificate — Coursera (Sep 2022)
- IT Support Fundamentals — Coursera (Feb 2022)
- The Bits and Bytes of Computer Networking — Coursera (Mar 2022)
- Operating Systems and You: Becoming a Power User — Coursera (May 2022)
- System Administration and IT Infrastructure Services — Coursera (Jul 2022)
- IT Security: Defense Against the Digital Dark Arts — Coursera (Sep 2022)
- Information Systems Auditing, Controls & Assurance — HKUST (Dec 2023)
- Foundations of User Experience (UX) Design — Coursera (Oct 2022)
- Start the UX Design Process: Empathize, Define & Ideate — Coursera (Nov 2022)
- English Teacher Training Program — I Learn Academy (2017)

## 🌐 ЯЗЫКИ

Арабский — родной ●●●●●

Английский — свободный (C1) ●●●●●

Русский — средний (B1) ●●●●●

Японский — начальный (A1) ●●●●●

**Прочие проекты:** скрапер Google Maps (Python) · сайт репетиторства с бронированием (React + Supabase) · маркетплейс исполнителей (Android) · POS/учётная система · исследование деревьев BST/AVL на Python и C++.