

# SALEH NOMAN

Data Scientist · Data & Automation Engineer · Web Scraping / ETL

Moscow, Russia +7 900 159-71-98 salehnoman22@gmail.com linkedin.com/in/salehnoman github.com/SalehNoman-MISIS

## ★ PROFILE

Data Science M.Sc. graduate (Honors / Red Diploma, GPA 5.0/5.0) who builds end-to-end machine-learning, web-scraping, and ETL systems. Master's thesis delivered a hybrid rule-based + ML engine that detected the Applicant Tracking System of every Fortune 500 company with 100% coverage. Leads multi-source job-data ingestion at Workgaze and built the official scraping-template library for RTILA Studio. Fluent English, native Arabic, working Russian.

## ★ TOP STRENGTHS

|                                     |     |                                  |     |
|-------------------------------------|-----|----------------------------------|-----|
| Python & Data                       | 95% | Machine Learning                 | 85% |
| Web Scraping (Playwright, Selenium) | 93% | Automation (RTILA, n8n)          | 84% |
| AI-Assisted Dev (Claude / LLMs)     | 92% | Full-stack (Flask/FastAPI/React) | 82% |
| ETL & Data Engineering              | 88% | SQL & Databases                  | 80% |

## <> CORE SKILLS

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

**Data & ML:** pandas, NumPy, scikit-learn, XGBoost, CatBoost, LightGBM, SHAP, feature engineering, cross-validation

**AI & LLMs:** AI-assisted development with Claude & GPT SDKs, agentic tooling, prompt engineering, RAG-style pipelines

**Scraping & Automation:** Playwright, Selenium, Scrapling, BeautifulSoup, RTILA Studio, anti-bot, pagination, n8n

**Web & Full-stack:** Flask, FastAPI, React, Node.js / Express, Supabase, REST APIs, Alpine.js

**Data Engineering:** ETL, normalization, deduplication, validation gates, freshness tracking; Directus & Google Sheets APIs

**Databases & Tools:** PostgreSQL, MongoDB, Firebase, SQLite, Git, pytest, GitHub Actions CI/CD, PyInstaller

## 📁 EXPERIENCE

- Data Aggregation & Automation Manager (ETL Lead)** Sep 2025 - Present  
**Workgaze — Remote**
  - Own end-to-end scraping & ETL for multi-source job listings: discovery → extraction → cleaning → normalization → load to Directus.
  - Built an AI-assisted scraping platform (FastAPI + Playwright) with an agentic source builder that generates validated scraper configs from any URL; 200+ tests.
  - Implemented QA gates — dedup, timestamp normalization, schema validation, freshness tracking — with scheduling, logging, and alerting.
- Data Automation Specialist & Application Support** Jan 2023 - 2026  
**RTILA Corporation — Remote**
  - Built the official library of scraping templates and standalone bots for RTILA Studio, covering Amazon, AliExpress, Etsy, LinkedIn, Google Maps, and more.
  - Engineered resilient extractions (CSS/XPath, pagination, anti-bot); wired outputs to Google Sheets, MySQL, Directus, and webhooks.
  - Authored troubleshooting guides, trained support teams, and directed weekly releases with change management and impact analysis.
- IT Support Specialist** Sep 2019 - 2023  
**Yemeni Socialist Party — Sana'a**
  - Upgraded OS and software institution-wide, secured equipment and data, and reached a 100% troubleshooting success rate.
  - Improved work efficiency ~50% through system updates; trained employees, boosting productivity ~40%.
- Volunteer English Language Teacher** 2018 - 2020  
**Sana'a, Yemen**
  - Taught high-school English grammar and exam technique; 100% of students passed finals, ~60% with excellent grades.

## <> SELECTED PROJECTS

### ATS Detection System — M.Sc. Thesis

Python scikit-learn XGBoost Playwright SHAP

- Scraped and parsed the career pages of all 500 Fortune 500 companies via a resilient Playwright/Selenium/BeautifulSoup cascade with HTML caching.
- Designed a 4-tier detection cascade ending in a RandomForest over 1,651 features with SHAP explainability.
- Reached 100% detection (500/500) across 30 ATS providers; 90 passing tests, k-fold cross-validation, active-learning loop.

## AI-Assisted Job-Scraping Platform

Python FastAPI Playwright Alpine.js

- Multi-source scraper ingesting job boards from YAML / CSS-selector configs with pagination and stealth fetch modes.
- Agentic “source builder” drives a live browser via Claude/GPT SDKs to auto-generate and validate scraper configs from any URL.
- Normalized and loaded listings into Directus with dedupe and freshness tracking; 200+ automated tests.

## Contact Data Enrichment Platform

Python Flask React Playwright Scraping

- Full-stack app extracting emails, phones, and social links from CMS collections or URL lists, with language detection.
- Google-Maps website finder, SMTP verification, and JSON/CSV/TSV/JSONL export over a REST API.
- Packaged as a licensed Windows app with a signed per-client licensing system for commercial distribution.

## Browser Automation Agent Platform

Node.js Express Playwright

- Visual Flow Builder (26 step types: variables, conditions, loops, error handling) chaining seven reusable browser agents.
- Single n8n/HTTP endpoint (saved-flow, inline-steps, legacy modes) with in-memory session management and real-time SSE logs.

## Backlink Marketplace Price Aggregator

Python anti-detect browsers Google Sheets API

- Checks a domain across five marketplaces, parses availability and pricing, and writes the cheapest option to Google Sheets.
- Credentials encrypted at rest (PBKDF2 + Fernet); one-click Windows installer built via GitHub Actions CI.

## Gender-Prediction ML Model

Python CatBoost LightGBM 1-month course project

- End-to-end pipeline: ingestion, cleaning, feature engineering, model training, evaluation.
- CatBoost/LightGBM with k-fold cross-validation and threshold tuning for macro-F1, benchmarked against baselines.

## EDUCATION

### M.Sc., Data Science — Honors (Red Diploma), GPA 5.0/5.0

NUST MISIS, Moscow

2024–2026

### B.Sc., Information Technology (92%)

University of Modern Sciences

2017–2021

### Russian Language & Literature

MIPT, Moscow

2023–2024

### Advanced English Diploma

Seeds Education

2016–2017

## CERTIFICATIONS

- 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)

## LANGUAGES

Arabic — Native ●●●●●

English — Fluent (C1) ●●●●●

Russian — Intermediate (B1) ●●●●●

Japanese — Elementary (A1) ●●●●●

**Additional builds:** Google Maps business-data scraper (Python) · tutoring & booking website (React + Supabase) · service-provider marketplace app (Android) · retail POS / accounting system · Python + C++ BST/AVL benchmarking study.