Indoha Guy Clovis

Self-taught data engineer and ML researcher
Kigali, Rwanda · Second-year bachelor's student

> Building practical systems at the intersection of software engineering, data, intelligent automation, and real-world business operations.

GitHub profile

About

I'm Indoha Guy Clovis, an undergraduate student and independent software builder based in Kigali, Rwanda. My work connects **software engineering, data systems, AI integration, and business operations**.

I build tools to make complex work more structured, reliable, and measurable. My projects range from a logistics operations platform to self-hosted AI-driven business automation. I'm also developing **Foundry**, a research-oriented project focused on environments, evaluations, and trajectories for training and testing AI agents.

I am particularly interested in **data engineering, applied machine learning, AI agent infrastructure, model evaluation, and reinforcement-learning environments**. I want to contribute to teams that build reliable AI systems and the datasets, tooling, and feedback loops behind them.

Open to: undergraduate internships, early-career software and data engineering roles, AI coding/evaluation projects, and research-engineering collaborations. Opportunities must match my current experience and eligibility.

Selected Projects

01 / InDataFlow

Logistics operations software · Data modeling · Business workflow engineering
Status: Active project
Website: indataflow.com

InDataFlow addresses the operational complexity of freight forwarding and logistics, where shipment information, documentation, approvals, and updates can become fragmented across people and tools.

Problem
Logistics teams need a consistent view of shipments, responsibilities, documents, and milestones. Fragmented records and communication can make follow-up and accountability difficult.

Approach

- Organize shipment activity around structured records and operational workflows.
- Support document handling, verification steps, milestones, and shipment visibility.
- Create clearer handoffs between internal operators and customers.
- Design software around the way logistics businesses actually work—not just generic CRUD screens.

What it demonstrates

- Translating domain-specific business processes into software models.
- Thinking about structured operational data, traceability, and workflow state.
- Designing applications for reliability, visibility, and human coordination.

- Balancing technical implementation with business needs.

---

02 / AutoEvolve

Self-hosted AI sales and marketing automation · LLM integration · Human-in-the-loop workflows
Status: Existing project

AutoEvolve explores how small teams can run structured research, outreach, and marketing processes with AI assistance while retaining control over their workflows and data.

Problem
Sales and marketing work often involves repetitive research, writing, coordination, and review across disconnected tools. AI can help, but automation needs safeguards, transparent steps, and human oversight.

Approach

- Build workflows for prospect research, contact enrichment, outreach, and campaign preparation.
- Integrate language models into multi-step processes rather than treating them as standalone chat interfaces.
- Keep human review and approval as explicit parts of the workflow.
- Explore self-hosted operation and flexibility across AI providers.

What it demonstrates

- Applied LLM integration and automation design.
- Workflow orchestration and tool-driven application logic.
- Building AI-assisted software around user control and practical constraints.
- Product thinking for small-business use cases.

---

03 / Foundry

Agent training environments · Trajectory generation · Evaluation infrastructure
Status: Research in development

Foundry is my planned research platform for creating realistic tasks, running AI agents against tool-using environments, recording their decisions, and evaluating their outcomes. Its purpose is to bridge the gap between application development and the infrastructure needed to improve agent behavior.

Research direction

- **Environment design:** represent a task's starting state, available tools, constraints, and desired outcome.
- **Agent execution:** expose controlled tools and record model actions and environment observations.
- **Verifiable evaluation:** check objective task completion and constraint adherence using deterministic tests where possible.
- **Trajectory delivery:** store and export structured sequences of observations, actions, tool results, and feedback.
- **Experimentation:** compare agents or models using reproducible tasks and transparent metrics.

Proposed first benchmark: logistics operations
An agent receives a simulated freight-forwarding task—for example, checking shipment documents, identifying missing information, updating a shipment record, and choosing the correct next operational step. A verifier checks the final state and rule compliance rather than relying only on a model's judgment of success.

Initial milestones

1. Define task, environment-state, tool-call, and trajectory schemas.
2. Build one deterministic simulated logistics environment with synthetic records.
3. Implement task generation, execution logging, and a rule-based verifier.
4. Run baseline agents and publish reproducible evaluation results.

5. Export a documented trajectory dataset and test a small supervised fine-tuning experiment.
6. Explore reinforcement-learning methods after establishing reliable baselines.

What this project is intended to demonstrate: Python engineering, data pipelines for ML, evaluation design, experimentation, and understanding of agent training workflows. Claims of research results will be added only after experiments have been completed.

Engineering Focus

Area	How I apply or intend to develop it
Software systems	Design applications around data integrity, well-defined workflows, APIs, and maintainable architectures.
Data engineering	Model structured operational information and develop toward validated, reproducible data pipelines and datasets.
AI integration	Connect language models to controlled tools and multi-step processes with human oversight.
ML evaluation	Develop verifiable tasks, baselines, metrics, and experiment reports through Foundry.
Business understanding	Translate practical domain constraints into product requirements and engineering decisions.

Technical stack

Languages: Python, Javascript, SQL, C++
Backend & APIs: FastApi
Data & storage: Apache spark, dbt, Apache Airflow, Ray, Postgress SQL, Redis **AI/ML:** Focusing on agentic ai training, Harbor, ATIF, OpenCode, OpenEnv, Hugging Face TRL
DevOps & infrastructure: Docker, Github Actions, Nginx, Caddy, Terraform, Lambda labs, S3

Research & Engineering Notes

I plan to publish short, reproducible technical write-ups that focus on engineering decisions and measured outcomes—not only product screenshots.

- **Designing verifiable tasks for logistics agents** — proposed Foundry design note.
- **What makes a good agent trajectory dataset?** — proposed technical exploration.
- **Turning freight workflows into a reliable data model** — proposed InDataFlow case study.
- **Human approval as an engineering constraint in AI workflows** — proposed AutoEvolve case study.

These are planned articles, not published research or completed experiments.

Education

Bachelor's degree — Software engineering
African Leadership University, Rwanda
Year: Second year · **Expected graduation:** Sept 2028

What I'm Looking For

I'm interested in opportunities where I can contribute and grow in:

- **AI coding and evaluation:** code quality review, debugging, task creation, test design, and model output assessment.
- **Data / backend engineering:** pipelines, structured data, APIs, workflow systems, and reliability.
- **AI agent infrastructure:** controlled tools, simulations, verifiers, evaluation harnesses, and trajectories.
- **Research engineering internships:** experiments, reproducibility, datasets, and ML tooling.

I especially value teams that care about **clear problem definitions, reproducible results, technical rigor, and useful software**.

Contact

If you're building ML infrastructure, evaluation systems, data-intensive products, or applied AI tools, I'd be glad to connect.

Location: Kigali, Rwanda

Last updated: November 2026. Project descriptions reflect current direction; proposed milestones and research are labeled separately from completed work.