LUMEN-RA

AI Integration Report

How Artificial Intelligence Was Used Across the Data Science Pipeline

Team 39 · Women Techsters Fellowship 2025 · SDG 5 — Gender Equality · February 2026

Project	Lumen-Ra — Digital Allyship Toolkit
Team	Team 39 — Data Science & Engineering
Document Type	AI Tools Integration Report
Phases Covered	Phase 1 (Warehouse) · Phase 2 (Ad Hoc) · Phase 3 (Dashboards) · Phase 4 (ML)
Date	February 2026

1. Introduction

Across the four phases of the Lumen-Ra Data Science and Engineering project, the team made deliberate and transparent use of Artificial Intelligence tools to accelerate development, improve quality, and overcome challenges that would otherwise have slowed progress significantly. This document provides an honest account of where, why, and how AI was used throughout the project.

Using AI tools in a data science project is not a shortcut — it is a professional practice. Senior data scientists, engineers and analysts in industry use AI assistants daily to write faster, debug more efficiently, understand unfamiliar concepts, and produce higher quality outputs. The team approached AI the same way: as a powerful tool that still required human judgment, domain knowledge and decision-making at every step.

2. AI Tools Used Across the Project

Tool	Phase	Primary Use
Claude (Anthropic)	All phases	Structural guidance, documentation drafting, code scaffolding, data analysis and insights

ChatGPT (OpenAI)	Phase 2, 3 & 4	Roadmap generation, to-do list planning, dataset structure brainstorming, ad hoc question framing, ML concept explanations
GitHub Copilot	Phase 4	Python code validation and autocomplete in Google Colab notebooks
Luca Studio (AI-assisted)	Phase 2	AI-assisted dashboard layout suggestions and chart type recommendations

3. Phase-by-Phase AI Usage

Phase 1 — Data Warehouse & Schema Design

The data warehouse phase presented the team's first major challenge: how do you build a data pipeline for a platform that does not yet have real users? The team used Claude to think through the Medallion Architecture — discussing the purpose of Bronze, Silver and Gold layers, what transformations were appropriate at each stage, and how to structure the schema to support future analytics without over-engineering it.

When designing the five-table schema on dbdiagram.io, Claude was used to review the entity relationships and flag potential issues with the foreign key structure before any data was generated. This saved the team from discovering referential integrity errors after hundreds of rows of synthetic data had already been created. Claude also helped draft the rationale for why Mockaroo was chosen as the data generation tool and what the dependency order for table generation should be.

In short, AI was used in Phase 1 as a senior technical reviewer — someone to sanity check decisions, suggest best practices, and explain why certain architectural choices matter in production environments.

Phase 2 — Ad Hoc Analysis

The ad hoc analysis phase required the team to move from raw data to business questions quickly. ChatGPT and Claude were both used to help frame the four core business questions in language that would resonate with non-technical stakeholders. Rather than just presenting a line chart of user counts, the team used AI to help articulate the underlying question: is the platform gaining momentum, or is user interest fading over time?

AI was also used to interpret the scatter plot findings — specifically to help describe what a cluster of high-progress, low-score users means in practical terms for a learning platform and what interventions would be appropriate for that user segment. The dashboard was built in Luca Studio, which itself uses AI-assisted suggestions for layout and chart type selection based on the data being visualized.

Phase 3 — Dashboards

The dashboard phase involved building two interactive analytics dashboards — one focused on platform performance and one on user retention and impact. The team used Claude as a structural guide throughout this phase, using it to plan the layout architecture, determine which KPIs and chart types were most appropriate for each business question, and define how the components should relate to one another on the page.

Claude provided a scaffold — a clear structural plan specifying which sections the dashboard needed, what each visual should communicate, and how the HTML, CSS and JavaScript components should be organized. The team then used that scaffold to write, refine and customize the actual code, applying the Lumen-Ra color palette and ensuring the dashboards functioned correctly across different browsers.

AI was also used to help define and validate the measures used inside the dashboards — ensuring that KPI calculations such as return rate, average quiz score and completion rate were correctly structured and consistent across both dashboards before they were finalized.

AI was also used to explore the underlying data — feeding in the cleaned values and asking for patterns that might not be immediately obvious. The finding that text modules outperform video by 4.7 points, and that learners returning for a fourth session score 8.4 points higher on average, both emerged from AI-assisted data exploration before being incorporated into the dashboards as highlighted insights.

Phase 4 — Machine Learning

The machine learning phase saw the most varied and hands-on use of AI tools across the project. Before any technical work began, the team used ChatGPT to help generate a structured project roadmap after an initial brainstorming session. Rather than starting from a blank page, the team described the problem — building a behavioural prediction pipeline for a learning platform — and used AI to help translate that into a sequenced, realistic plan covering EDA, feature engineering, modelling and output integration. This gave the team a clear direction from day one.

AI was also used to generate to-do lists at the start of each working session, breaking the broader roadmap down into specific, actionable tasks. This kept the work organized and ensured nothing was missed as the team moved through the different stages of the pipeline.

When it came to the data itself, AI was used to help think through the structure of the synthetic dataset before it was generated in Mockaroo — discussing what fields would be most useful for machine learning, what realistic value ranges should look like, and how the weekly behavioral features should be represented in the data. This front-loaded thinking made the actual data generation process faster and produced a more ML-ready dataset from the start.

GitHub Copilot assisted with Python code in the Google Colab notebooks — autocompleting Pandas operations for weekly aggregation and validating the Scikit-learn Logistic Regression implementation to catch errors before they caused silent failures in the model output. Having AI validate the code in real time gave the team confidence that the pipeline was running correctly even when working with an unfamiliar library.

Claude was used to review the feature engineering logic and suggest the confidence score as a behavioral proxy that would avoid the subjectivity problems of self-reported survey data. When the ML pipeline documentation was written, Claude helped translate the technical process into plain English that a non-technical stakeholder could follow without a machine learning background.

4. AI in Documentation

The full project documentation — including the technical documentation, the README files for each phase folder, and this AI integration report — was produced with significant Claude assistance. The team provided all of the facts, data, context and decisions. Claude helped structure them into coherent, professional documents with appropriate headings, flow and language.

This is a legitimate and increasingly standard use of AI in professional environments. The accuracy of the documentation is the team's responsibility — every number, every architectural decision, and every insight described came from the team's actual work. AI was the writing partner, not the analyst.

5. Responsible AI Use — What We Did and Did Not Do

The team approached AI integration with clear boundaries. AI was used to assist, accelerate and improve — not to replace the team's judgment or fabricate outputs.

What AI Was Used For	What AI Was NOT Used For
Generating a project roadmap from the team's brainstorm	Deciding the project direction or scope
Breaking the roadmap into actionable to-do lists	Prioritizing tasks — the team made those calls
Thinking through dataset structure before generation	Generating or fabricating the actual data
Validating Python code in Google Colab notebooks	Writing the core ML pipeline logic without human oversight
Providing structural scaffolding for dashboard layout	Deciding what the dashboards should measure or show
Helping define and validate dashboard measures and KPIs	Choosing which KPIs mattered — that required business context
Writing documentation from facts provided by the team	Generating findings or insights the data did not support

6. Conclusion

AI was a meaningful contributor to the quality and speed of this project. Without it, the production-grade dashboards would have taken significantly longer to plan and build, the machine learning pipeline would have been harder to organize and validate, and the documentation would have been less comprehensive.

At the same time, the value of this project came from the team. The data architecture was designed by people who understood the platform's needs. The business questions were asked by people who understood the SDG 5 mission. The roadmap was built by people who understood what was achievable in the time available. The insights were validated by people who understood what the numbers actually meant for real learners going through a digital allyship programme.

AI made the team faster, sharper and more polished. The team made the project meaningful.

Lumen-Ra · Team 39 · Women Techsters Fellowship 2025

SDG 5 — Gender Equality · AI Integration Report · February 2026