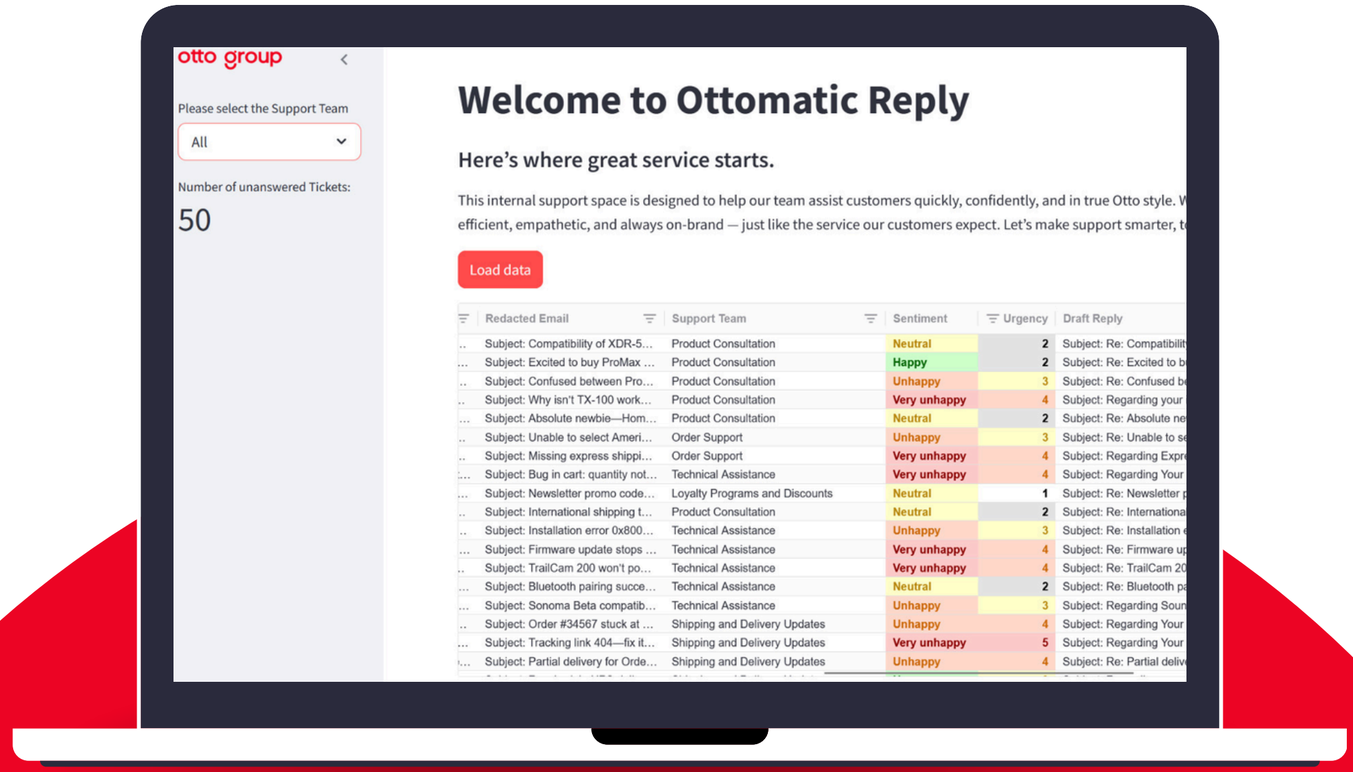
otto group one.O

Ottomatic Reply
Email Concierge

Group 2 – Hackathon 2025

Automating first-line email triage so agents spend time on solutions, not sorting.



The Team

Nicola Dale
Stellenbosch University

Alexander Mitte
TU Dresden

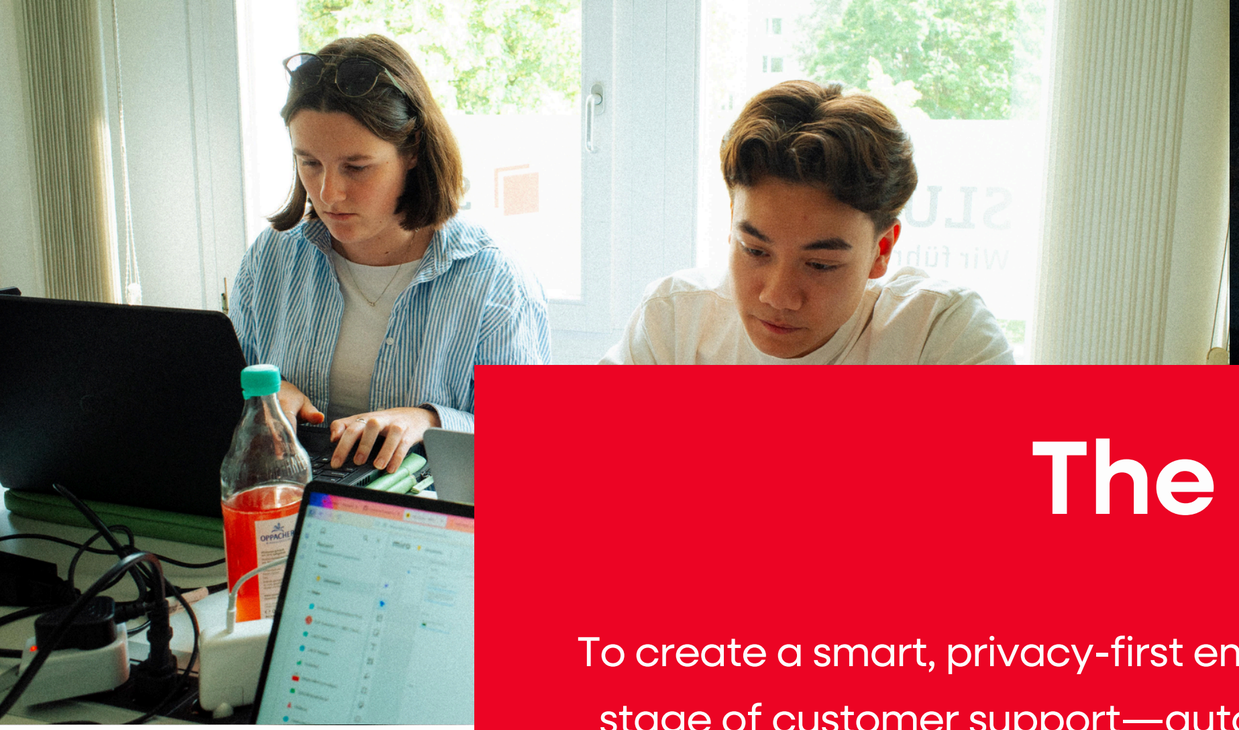
Sean Kiley
Stellenbosch University

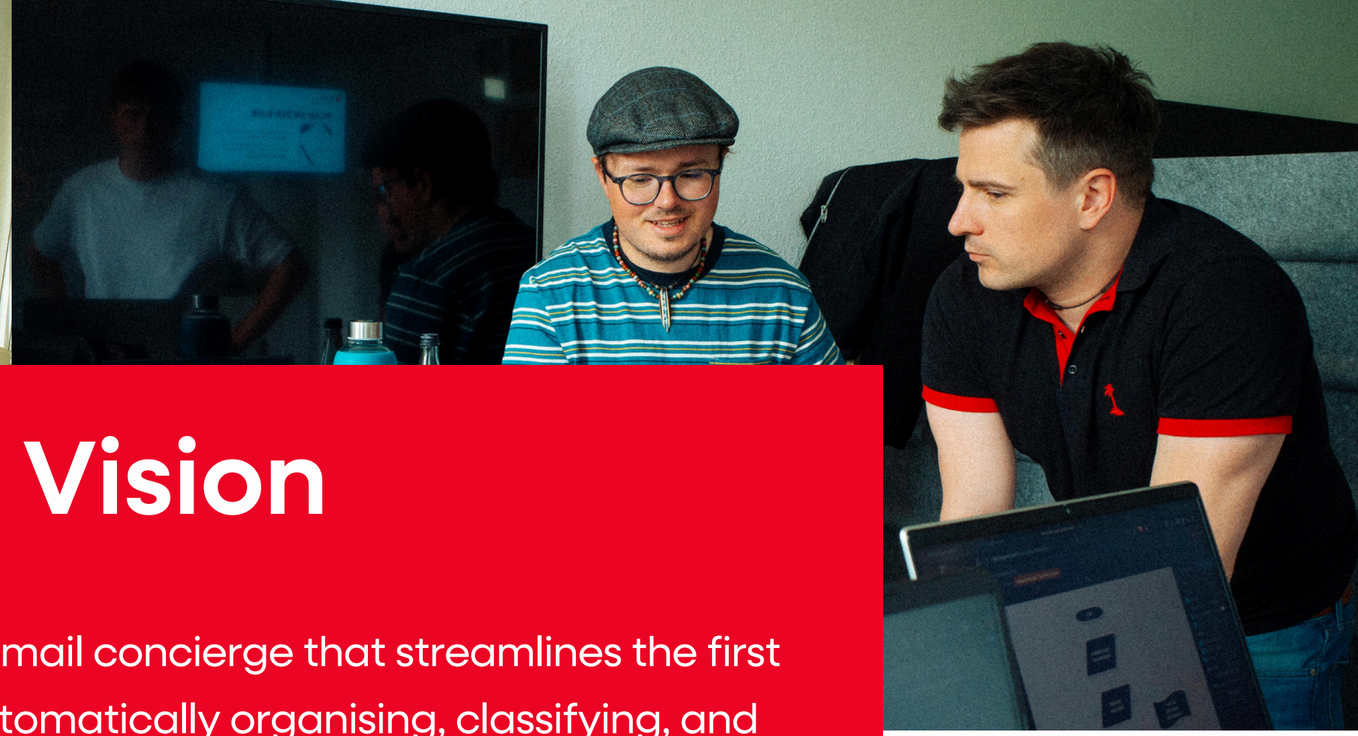
Kevin Hoang
TU Dresden

Jordan Luck
Stellenbosch University

Overview

otto group one.O





The Vision

To create a smart, privacy-first email concierge that streamlines the first stage of customer support—automatically organising, classifying, and prioritising incoming messages so teams spend less time sorting and more time solving, all powered by the precision of rule-based logic and the flexibility of AI.

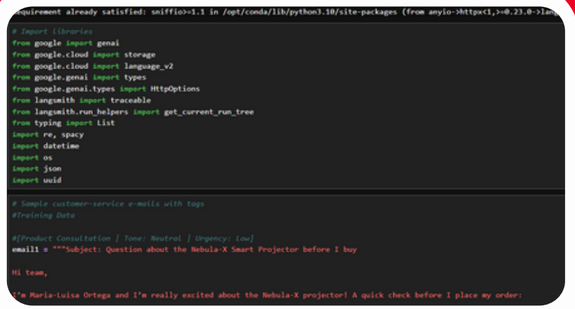
Solution Design

Our solution design leverages natural language processing to automate and enhance email support through five core processes: classifying emails by topic, masking personal information for privacy, detecting sentiment to guide tone, identifying intent to inform response selection, and scoring urgency to prioritise action. Together, these components form an intelligent concierge system that ensures efficient, accurate, and human-aligned responses to customer emails.

Problem 01	Problem 02	Problem 03	Problem 04	Problem 05
Classifying emails by topic	Masking personal information	Recognising sentiment	Recognising sentiment	Scoring urgency

otto group one.O

Project Plan

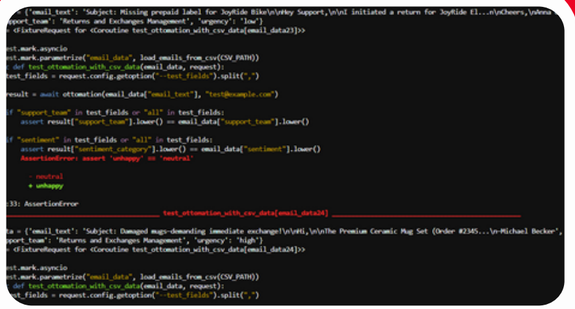


Tuesday

- Getting emails into Notebook
- Creating basic test data
- Categorising each email by Support Team
- Built simple draft reply based on each email
- Sentiment Analysis: Output sentiment category & scores

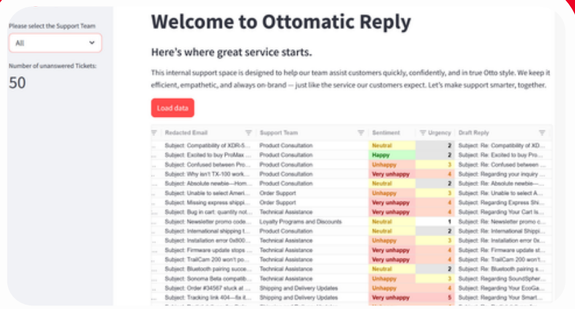
Wednesday

- Basic PII Reduction
- Enhance PII reduction to catch all relevant data
- Get urgency from Sentiment Analysis and Support Team Category
- Add Tracing with Langsmith to follow
- Define output format for different functions



Thursday

- Creating more test data
- Creating automatic Unit Testing and scores to understand how good the pipeline is
- Enhance dashboard with additional filters, formatting...
- Insert retrieved file from cloud storage into Frontend
- Craft basic Frontend to show the data

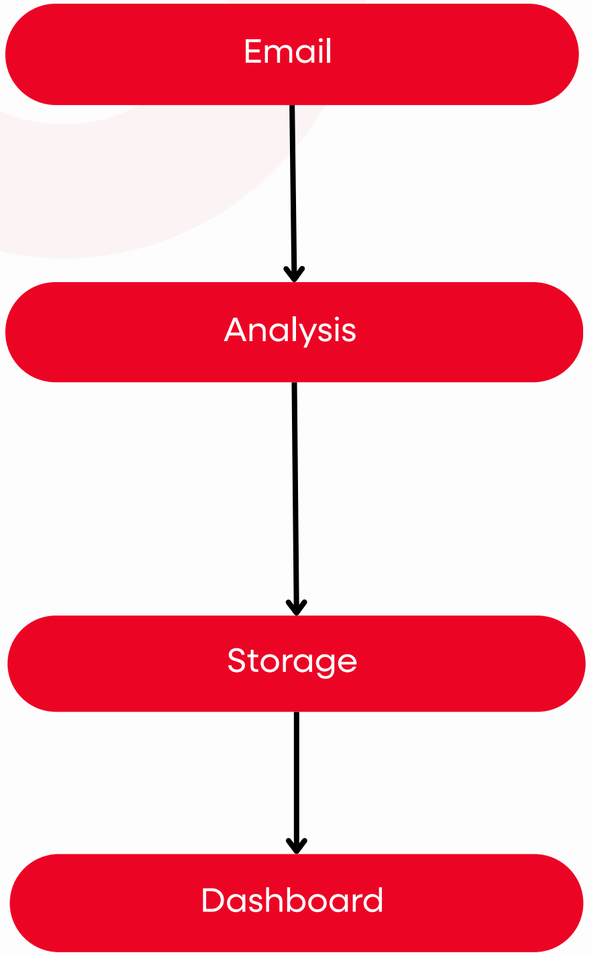


Friday

- Presentation

otto group one.O

Architecture Diagram



1. Email (Input Source)

- Source Data: Uses 60 generated emails from ChatGPT for testing.
- Nature: Raw, semi-structured data with varying content, tone, and urgency (focus on unstructured body).

2. Analysis (Intelligent Processing)

- Core intelligence: AI/ML models process email content.
  - PII Redaction: Masks sensitive personal information using Regex AND spaCy's Named Entity Recognition (NER).
  - Classification (Gemini): Categorises emails into 10 specific support groups.
  - Sentiment Analysis (Google Cloud NL API): Determines emotional tone and magnitude.
  - Urgency Scoring: Assigns a priority score from Level 1 (highest) to Level 5 (lowest).
    - This is calculated based on a combination of the email's classified category and its sentiment.
  - Draft Reply Generation (Gemini): Creates brand-aligned, PII-free preliminary response templates.

3. Storage (Data Persistence)

- Saves all original and analysed email data for persistence and access.
  - Google Cloud Storage (GCS): Comprehensive analysed email data is converted to a JSON file and uploaded to a GCS bucket, serving as the central archive for all processed insights.

4. Dashboard (Streamlit UI)

- Interface: Built using Streamlit, serving as the intuitive front-end for human agents.
- Functionality:
  - Displays a clear table showing each email's: Timestamp, address, original email, redacted email, support team category, sentiment, urgency, and draft reply.
  - Enables quick understanding, prioritisation, and efficient review/editing of responses.

otto group one.O

Google Cloud Storage

Streamlit

spaCy

vertex.ai

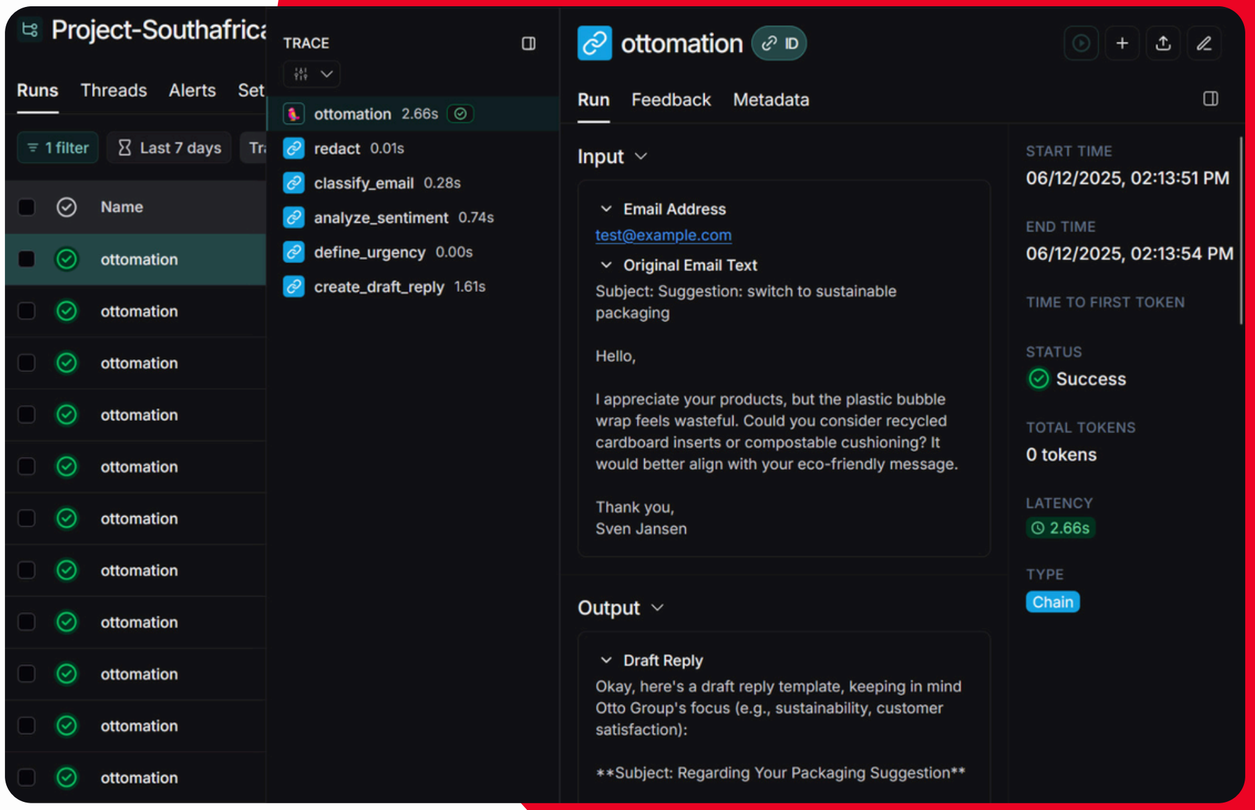
Leveraging LangSmith

We use LangSmith for enhanced development and monitoring of our AI pipeline.

How We Use LangSmith (Our Observability Platform):

- **Tracing**: Visualises the data flow and execution path through our email processing pipeline.
- **Debugging:** Helps identify and resolve issues within AI calls and data transformations.
- **Monitoring**: Tracks the performance of LLM interactions and overall pipeline for quality assurance.

In short, LangSmith enables us to build, debug, and maintain our AI-driven E-Mail Concierge system effectively.



otto group one.O

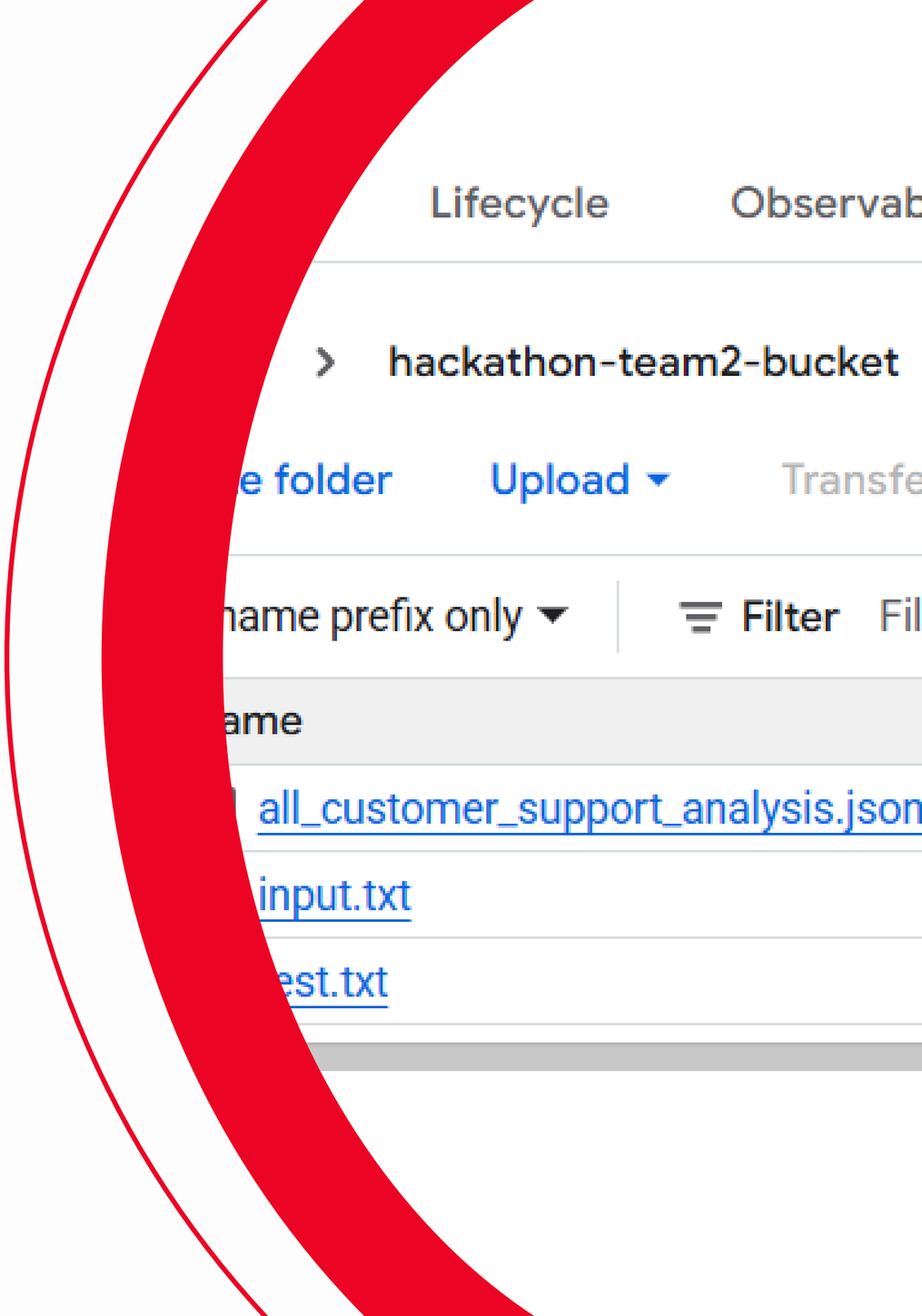
Testing & Storage

1. Testing Data

- **Source**: 60 unique customer service emails generated by ChatGPT.
- **Use**: Rigorous testing and refinement of the GEMINI AI models.

2. Storage (Google Cloud Storage - GCS)

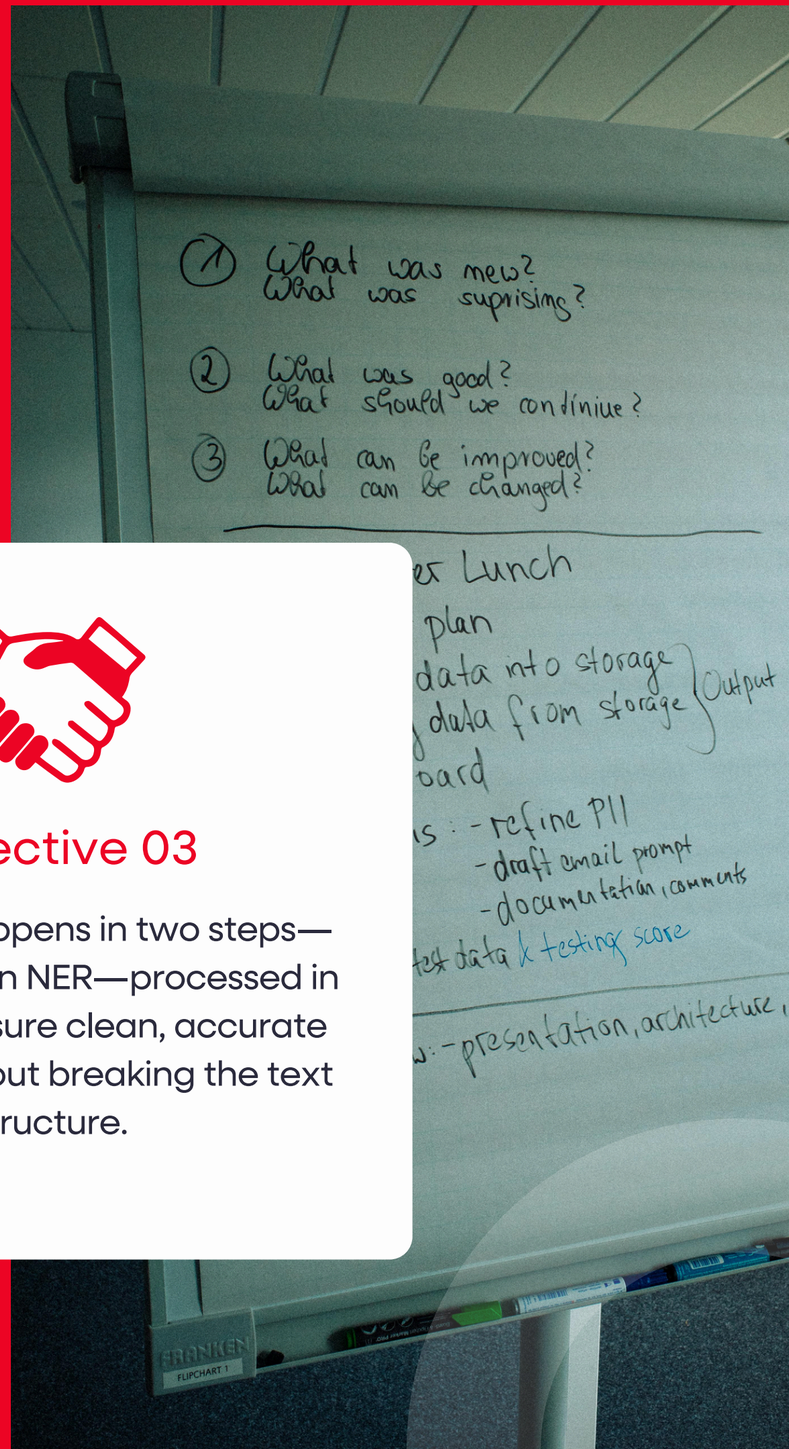
- **Method**: All comprehensive analysed email data is converted into a single JSON file.
- **Location**: Uploaded to a dedicated Google Cloud Storage (GCS) bucket.
- **Role**: Serves as the primary, scalable, and durable archival repository for all processed insights.



otto group one.O

Masking Personal Information

To safely use customer emails with LLMs, we implemented a robust PII redaction system. This ensures sensitive data is removed before processing, protecting user privacy and meeting company compliance standards.



Objective 01

We use regex patterns to automatically detect and redact emails, phone numbers, card details, and order IDs, replacing them with “[REDACTED]” to protect user privacy.

Objective 02

Using spaCy’s NER model, we identify and redact names, locations, and other sensitive entities that regex alone might miss.

Objective 03

Redaction happens in two steps—regex first, then NER—processed in reverse to ensure clean, accurate masking without breaking the text structure.

otto group one.O

Classification

Support Teams

We fine-tuned a system prompt to classify emails into one of ten support categories, covering everything from presales inquiries to post-purchase complaints. This ensures accurate routing and faster response times.

Gemini Prediction

The Gemini 2.0 Flash model is used to process each incoming email. It applies the classification logic asynchronously using Google Vertex AI, ensuring scalability and low-latency predictions.

Data Cleaning

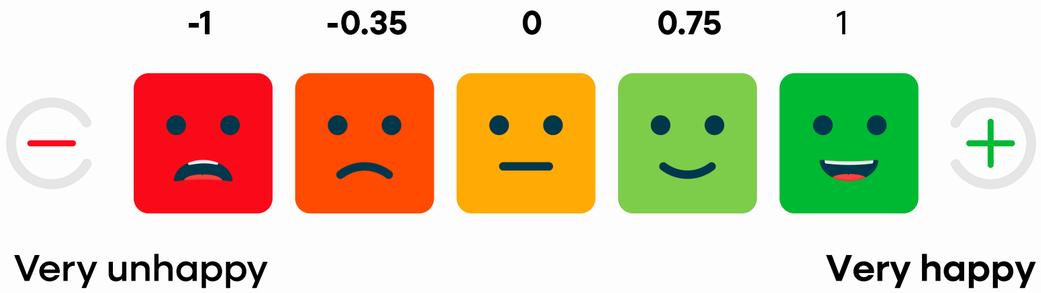
Every response is stripped of noise and returned as a clean label, allowing for real-time analytics, prioritisation, and dashboard integration across support teams.

otto group one.O

Recognising Sentiment

We used a ML model to assign a sentiment score and magnitude ranging from very unhappy to very happy.

We then categorize these scores into five distinct intervals: very unhappy, unhappy, neutral, happy, or very happy based on generated test data.



Learnings:

Hard to define intervals without properly validated test data, risk of overfitting

Customers often only contact customer Service when they are unhappy (motivation)

Scoring Urgency

Each email is scored out of 5 based on a combination of its category and sentiment:

- Category weighting:
  - Aftersales (e.g., Shipping, Returns) → +3
  - Presales/General Support → +2
  - Feedback/Promotions → +1

- Sentiment modifier:
  - "Very unhappy" → +2
  - "Unhappy" → +1
  - Neutral/Happy → +0

➡ Final Score = Category + Sentiment

Scores range from 1 (low urgency) to 5 (critical priority), helping the system route emails efficiently and fairly.

01 Not urgent

02

03

04

05 Very urgent

We first explored using an LLM to grade urgency, but found the results to be inconsistent and somewhat random. We transitioned to a more reliable rule-based system.

We discovered that not all problems within a given support category share the same level of urgency. This suggests a need for more detailed categorization.

Our current model inherently prioritizes existing customers through the assignment of urgency based on the support team.

otto group one.O

Draft Reply Generation

This function uses Google's Gemini 2.0 Flash model to generate a brand-aligned, privacy-safe response to customer emails. After receiving the original email, it builds a prompt instructing the AI to act like a support agent from the Otto Group—explicitly excluding any personal data. The Gemini model then returns a helpful draft reply, enabling faster, consistent customer communication while keeping human interaction in the loop for final review.

```
draft_prompt = f"""You are a customer support agent.
You are generating a brand-aligned draft reply template from the Otto Group in response to a customer message.
Do not include any personal information about the customer in the template.

Here is the customer's original email:
---
{email_text}
---

Based on all this information, please write a draft reply.
"""
```

otto group one.O

Statistics

Quality Assurance and Performance

- *measured using automatic testing*

88%

Of support team classfication accuracy

57%

Sentiment accuracy

60

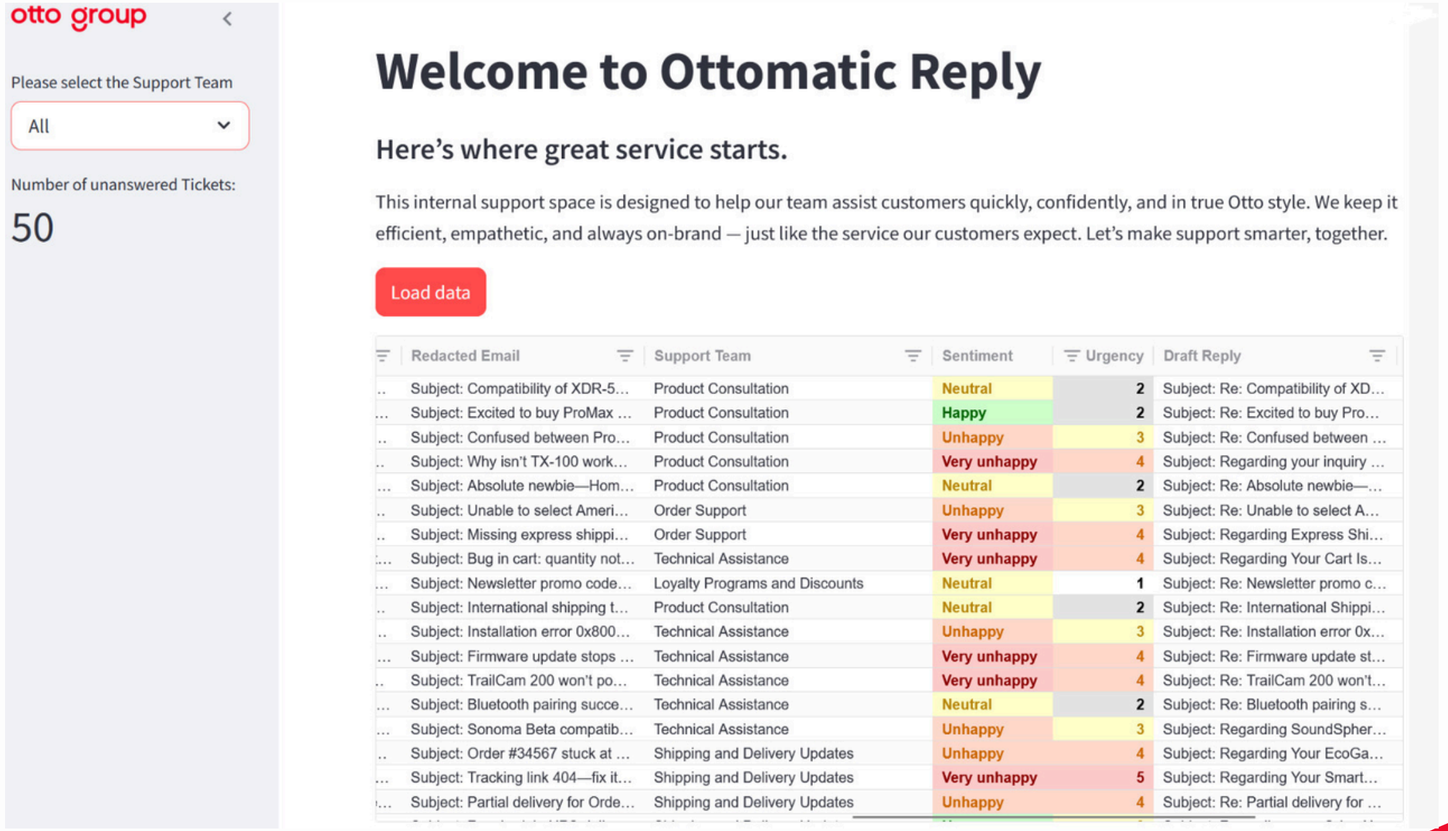
Test emails

otto group one.O

Dashboard (Streamlit UI)

Our Streamlit UI provides an intuitive interface for customer service agents.

- **Purpose**: At-a-glance view of processed email insights.
- **Displays**: Tabular data including: Timestamp, address, original email, redacted email, support team category, sentiment, urgency, and draft reply.
- **Benefit**: Enables rapid review, prioritization, and efficient response generation.



otto group one.O