# Release 3 Documentation

**Project: Bosch UML diagram translation**

**Team Members: Aamish Samotra, Akshita Arora, Lida Wei, Shambhavi Singh**

---

## Release Objective

The primary objective of Release 3 was to enhance the quality, accuracy, and contextual detailing of generated cloud networking diagrams using fine-tuning and prompt engineering methods. This release focused on developing structured synthetic datasets and integrating them into the fine-tuning process to improve model understanding of networking elements and their relationships.

---

## Key Focus Areas

### 1. Creation of Synthetic Datasets

- Developed two distinct types of datasets in tabular format:
    - *Component-Level Dataset*: Focused on detailed attributes of networking elements such as ports, security groups, IP protocols, NAT rules, and routing configurations.
    - *Relationship-Level Dataset*: Focused on defining rules and communication between VPCs, subnets, and other networking components, capturing dependencies, flow of traffic, and firewall permissions.

---

### 2. Fine-Tuning the Model

- Integrated the synthetic datasets to fine-tune the language model for better diagram generation capability.
- Fine-tuning targeted improvements in:
    - Correct placement of networking components.
    - Better context in security configurations.
    - Capturing rules and dependencies between components.

---

### 3. Diagram Generation Using Prompt Engineering

- Generated multiple AWS sample diagrams using structured and scenario-driven prompts.
- Refined prompt templates for better consistency, clarity, and accuracy.
- Focused on representing complex networking scenarios like multi-tier architectures, traffic routing, and access control setups.

---

### 4. Evaluation & Testing

- Compared results of pure prompting vs. fine-tuned model outputs.
- Identified that fine-tuning improved component recognition while prompting worked better for simpler cases.
- Discussed hybrid strategy with the client for future phases.

---

## Deliverables in Release 3

- Component-Level Synthetic Dataset (Tabular Format).
- Relationship-Level Synthetic Dataset (Tabular Format).
- Fine-tuned Model with updated datasets.
- Sample AWS Architecture Diagrams generated using prompting and fine-tuned model.
- Documentation of approach, dataset structure, and output evaluations.

---

## Challenges Faced

- Handling complex networking relationships like multi-region setups and VPC peering required richer data context.
- Maintaining consistency between different diagrams generated using similar inputs.
- Balancing the trade-off between fine-tuning time/resource consumption and prompt flexibility.

---

## Key Outcomes of Release 3

- Improved diagram generation quality with better descriptive details.
- Standardized datasets for scalable model fine-tuning.
- Developed a clear separation of datasets for individual components vs. inter-component relationships.
- Established a direction for implementing scenario-based datasets for next releases.

# Next Steps for Release 4

- Expand relationship-level dataset with real-world networking scenarios.
- Continue hybrid approach combining fine-tuning and advanced prompt engineering.
- Map dataset fields to specific diagram outputs for explainability and control.
- Prepare for potential deployment framework in the cloud for automated diagram generation.