TokuTel AI Engineer Take-Home Assignment

Thank you for interviewing with Toku for the AI Engineer role. As the next step, we would like you to complete a short take-home assignment. The purpose is to understand how you approach problem-solving, design retrieval-augmented systems, and enforce policies within AI applications.

Objective

Build a retrieval-augmented assistant grounded on the provided synthetic dataset (in the `data/` folder). Your assistant should answer customer questions accurately and enforce the company policies defined in the dataset. The dataset has been created specifically for this exercise and is not publicly available.

Data Provided

- plans.csv — fictional telecom/communication plans.
- kb.yaml — internal policies, escalation rules, SLAs, and discount rules.
- transcripts.json — synthetic examples of customer support interactions.
- faq.jsonl — seed FAQ pairs.
- eval_prompts.txt — evaluation questions to run your system against.

What You Need To Do

- Design an indexing and retrieval approach using the provided dataset (embeddings, chunking, or any IR method).
- Implement answer generation with explicit citations pointing back to the files and rows/IDs used.
- Apply policy rules such as masking PII, enforcing escalation levels, and adhering to SLAs.
- Evaluate your system on the provided eval_prompts.txt and save the outputs.
- Document your design decisions, trade-offs, and potential improvements in a 1–2 page ARCHITECTURE.md.

Constraints

- Use only the provided data files; do not rely on external internet knowledge bases.
- You may use any language or frameworks you are comfortable with.
- Ensure your code is reproducible and can be run locally with clear instructions.
- Include citations in the format: [plans.csv#row=4], [kb.yaml#features_matrix], [transcripts.json#t-002].

Evaluation Criteria

- Grounding and correctness: answers must be traceable to the dataset.
- System design clarity and trade-offs explained in ARCHITECTURE.md.

- Code quality, structure, and readability.
- Handling of edge cases and application of policy rules.
- Creativity in approach and potential for production-readiness.

Expected Time and Deliverables

- Expected time: approximately 4–6 hours.
- Deliver your code in a GitHub repository (public or private with access granted).
- Provide a README with setup and run instructions.
- Provide a short design document (ARCHITECTURE.md).
- Include evaluation outputs for the provided prompts.

Submission and Questions

Please upload your solution to GitHub and share the repository link with us via email. If you have any questions or need clarification, contact Haniyeh at Haniyeh.abdi@toku.co. We look forward to reviewing your work and discussing your approach in the next stage.