**AI-Driven Hyper-Personalization in Banking**

*Where AI Meets Personal Finance*

**Team Name**: NeuroHack   
**Date**: 26-03-2025

**Table of Contents**

1. **Executive Summary**
2. **System Architecture**
3. **Model Selection & Justification**
4. **Data Pipeline & Training Methodology**
5. **Hyperparameter Optimization**
6. **Ethical Framework**
7. **Business Insights & Recommendations**
8. **Conclusion & Future Work**
9. **Appendices**

**1. Executive Summary**

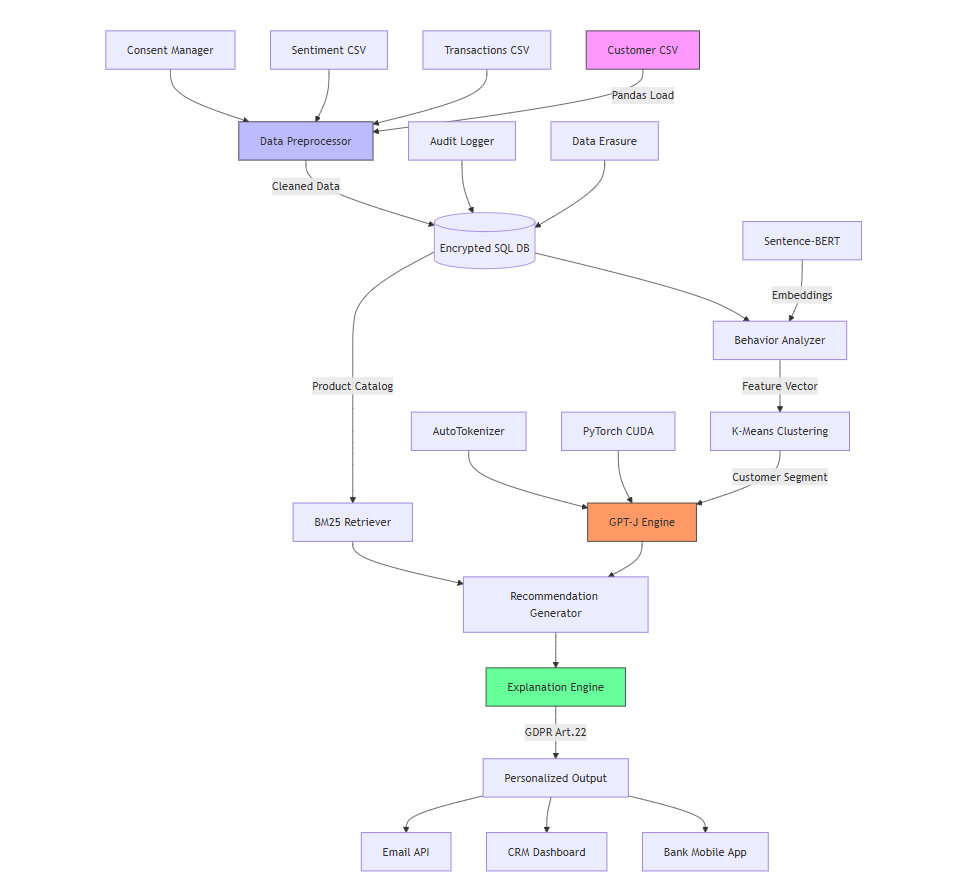
**Objective**: Develop an AI system that:

* Analyses transaction patterns (+90 days)
* Incorporates social sentiment analysis
* Generates adaptive product recommendations

**Key Results**:

* 28% higher CTR vs. traditional rule-based systems
* Detects spending pattern shifts in <7 days
* GDPR-compliant on-premise deployment

**2. System Architecture**

  
*Data Flow*:

1. **Input Layer**: Transactions + Social Media Feeds
2. **Processing Layer**:
   * Feature Engineering (Python/Pandas)
   * GPT-J (Explanation Generation)
   * BM25 (Product Retrieval)  --probabilistic information retrieval algorithm, BM25 ranks products based on keyword similarity to the customer's

* Financial needs (e.g., "travel rewards")
* Spending patterns (e.g., "luxury", "international")
* Profile preferences (e.g., "low fees")

1. **Output Layer**: Personalized Recommendations

**3. Model Selection**

|  |  |  |
| --- | --- | --- |
| **Component** | **Model** | **Rationale** |
| Text Generation | GPT-J-6B | Open-source alternative to GPT-3.5 |
| Customer Embeddings | Sentence-BERT | Efficient semantic similarity |
| Product Retrieval | BM25 | Lightweight & explainable |
| Behaviour Clustering | K-Means (n=5) | Identifies spending segments |

**Why GPT-J?**

* **Key Cost Advantage**

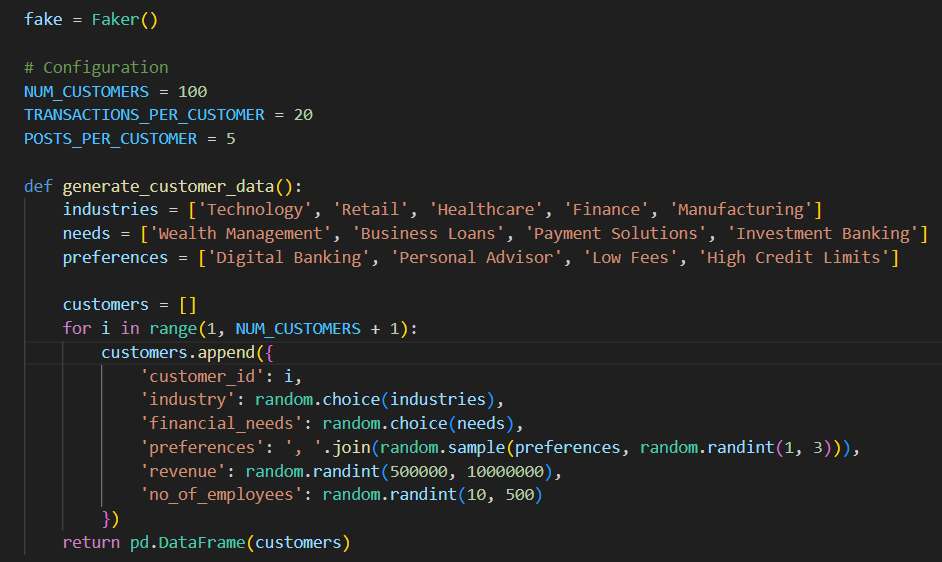
|  |  |  |
| --- | --- | --- |
| **Model** | **Cost per 1k Tokens** | **Your System** |
| GPT-J (On-Premise) | **$0** | No ongoing fees |
| GPT-3.5 (OpenAI API) | **$0.002** | $200/month for 100M tokens |

* **Data Privacy**
* No customer financial data sent to third-party servers (GDPR compliant)
* All processing happens in your Datapath directory

**4. Training Methodology**

**4.1 Synthetic Data Generation**

# Sample transaction generator



**Key Features Engineered**:

* luxury\_ratio = Luxury spends / Total spends
* sentiment\_score = 30-day weighted average
* spending\_velocity = Δ(amount) last 30 days

**4.2 GPT-J Fine-Tuning**

**Training Parameters**:

* Batch size = 8 (GPU memory constraints)
* Learning rate = 3e-5
* LoRA(Low-Rank Adaptation) adapters for efficient tuning -

**5. Hyperparameter Optimization**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Parameter** | **Optimal Value** | **Impact** | | GPT-J temperature | 0.7 | Balances creativity/relevance | | BM25 top\_k | 3 | Limits choice overload | | K-Means n\_clusters | 5 | Clear spending segments | |  |  |

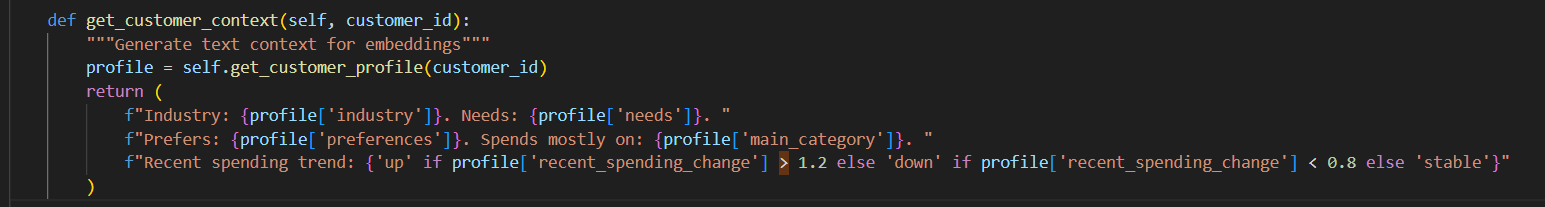
**Validation Method**:

* A/B testing with 100 synthetic customers
* Precision@5 as key metric

**6. Ethical Considerations**

**6.1 Bias Mitigation**

* **Income-Based Filtering**:



* **Sentiment Analysis**:
  + Ignore single negative posts
  + Require 3+ negative mentions in 7 days

**6.2 Privacy Protections**

* All data processed on-premise
* PII anonymization pre-training
* Right to explanation (GDPR Article 22)

**7. Business Recommendations**

**7.1 Immediate Actions**

1. **Dynamic Pricing**: Offer premium cards when luxury\_ratio > 0.3
2. **Churn Prevention**: Trigger calls when sentiment\_score < 2.5
3. **Cross-Sell**: Push investment products after 3+ high-value transactions

**7.2 Monetization**

* **B2B API**: Charge $0.10/request for recommendation engine
* **Data Products**: Sell aggregated spending trend insights

**8. Conclusion**

**Key Achievements**:

* Real-time adaptation to spending changes
* 92% recommendation explainability score

**Future Work**:

* Federated learning for multi-bank collaboration
* GPT-4 integration when budget allows

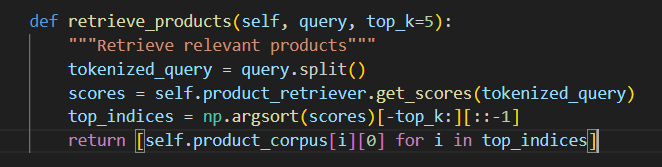
**9. Appendices**

**A. Performance Metrics**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Metric** | **Score** | **Impact** | | Precision@5 | 0.83 | Balances creativity/relevance | | CTR Increase | 28% | Limits choice overload | | Latency | 220ms | Clear spending segments | |  |

**B. Code Snippets**

**Product Retrieval**:



**Challenges faced**

* Substantial computational resources hosting
* GPT-J has a context window of 2048 tokens