## Core Financial-Grade Requirements (and Solutions)

| Domain | Challenge | Solution |
| --- | --- | --- |
| **Data Residency** | Customer data must stay in-region (e.g., UK/EU) | Use **Azure OpenAI hosted in UK/EU region**, ensure **Cognitive Search, Cosmos DB** also geo-bound |
| **PII Handling** | Cases may include names, accounts, financial data | Pre-process prompts with **PII redaction** (regex, Presidio) before passing to GenAI |
| **Auditability** | All AI decisions must be explainable | Log full prompt/response pairs in **Cosmos DB or Azure SQL** with timestamps, user ID, versioning |
| **Human-in-Loop** | No GenAI decisions fully autonomous initially | Implement **review workflow** via **Power Apps**, include override options and approval steps |
| **RBAC & Identity** | Access to models and data must be role-based | Enforce **Azure AD + Managed Identities**; apply **AAD security groups** to control actions |
| **Model Governance** | Model/Prompt drift, hallucinations must be tracked | Use **Azure ML + MLflow** to version models/prompts and run performance benchmarks over time |

## 📐 GenAI-Augmented Routing Workflow (Bank-Ready)

plaintext

CopyEdit

[Customer raises case (via email/form/chat)]

↓

[Case Preprocessor]

→ Clean text, redact PII (Presidio, regex)

↓

[Azure OpenAI – GPT-4 (Prompted)]

→ Generate: Summary, Case Category, Risk Tags, Urgency

↓

[Routing Reasoning Agent (LangChain + Azure Cognitive Search)]

→ Ingests SOPs, escalation policies, historical routing outcomes

↓

[Routing Decision]

→ Logs decision with reason (GPT explanation)

↓

[Power Apps Approval]

→ CSM reviews and approves routing OR

→ Auto-assigned if low-risk

↓

[CRM Update + Task Assignment via Logic Apps]

## 🧰 Bank-Tailored Azure Tech Stack

| Stack Layer | Azure Service | Notes |
| --- | --- | --- |
| **GenAI** | Azure OpenAI (GPT-4, GPT-3.5) | UK/EU data boundary, rate-limited, version-controlled |
| **RAG** | Azure Cognitive Search + Blob Storage | Index SOPs, case guides, escalation paths |
| **Preprocessing** | Azure Functions + Presidio (open-source) | PII redaction before prompting |
| **Audit/Logging** | Azure Cosmos DB OR Azure SQL | Log prompts/responses, user ID, timestamp, response confidence |
| **Prompt Routing Logic** | LangChain agents deployed via Azure Functions or Container Apps |  |
| **Case Workflow** | Power Automate / Logic Apps | Auto or semi-auto case handoff, SLA tracker |
| **Approval UI** | Power Apps | For human-in-loop CSM validation |
| **Security** | Azure AD, Azure Key Vault, Private Link | All endpoints protected, scoped, encrypted |

## 🧩 Example Financial Prompts

### 📌 1. Case Classification (with Risk Awareness)

plaintext

CopyEdit

Classify this banking case into one of the following categories:

[Account Access Issue, Transaction Reversal, Loan Inquiry, Fraud Alert, KYC Update].

Also assign a sensitivity tag: [Standard, Confidential, High Risk].

Case:

“My online banking is locked again and I saw a login from an unknown device.”

Answer:

Category: Fraud Alert

Sensitivity: High Risk

### 📌 2. Routing Logic with Policy Input

plaintext

CopyEdit

Given the case and routing policy:

Policy:

- Fraud cases go to “Fraud Ops Team”

- KYC cases go to “Compliance Team”

- Standard support goes to “CSM General Queue”

Determine:

- Routing Team

- Reason for assignment

Answer:

Routing: Fraud Ops Team

Reason: Customer account was flagged for suspicious access, indicating potential fraud.

### 📌 3. Checklist Generation (Regulatory Context)

plaintext

CopyEdit

Based on the SOP and case summary, list steps for resolution.

SOP:

Fraud Alert → Steps: ID verification, IP review, temporary freeze, alert customer.

Case:

Customer flagged for suspicious login attempts from foreign IPs.

Checklist:

1. Initiate identity verification

2. Temporarily freeze online access

3. Review login history

4. Notify customer via secure message

## 🧾 Compliance Features Checklist (for Banks)

| Requirement | Covered in Design |
| --- | --- |
| GDPR & UK DPA Compliance | ✅ Azure OpenAI EU-hosted, data masking pre-prompt |
| Explainability / Justification Logs | ✅ GenAI rationale captured with every routing |
| Data Sovereignty | ✅ All services deployed within UK/EU Azure regions |
| Role-Based Access & Segregation | ✅ Azure AD + Conditional Access + |
| Key Vault |  |
| PII Sanitization | ✅ Presidio/NLP-based scrubber before LLM |
| SOC2 / ISO27001 controls | ✅ Azure platform services meet certifications |

## 🔄 Pilot-to-Production Path for Bank Clients

| Phase | Deliverables |
| --- | --- |
| **Discovery & Design** | Functional mapping, SOP indexing, prompt templates |
| **MVP Pilot** | Case summary + routing with human-in-loop |
| **Phase 2** | Checklist generation, SLA flagging, risk tagging |
| **Full Rollout** | Multi-team, multi-channel support, direct CRM integration |
| **Ops** | Monitoring dashboards (Azure Monitor), fallback logic, retraining ops (if any) |

## 📎 Assets I Can Build for You

Would you like me to deliver any of the following next?

1. **Azure Reference Architecture Diagram (Visio-style)**
2. **Enterprise Governance Checklist** for banks using GenAI
3. **Prompt Library** tailored to financial use cases
4. **Deployment Guide** for your engineering teams
5. **Slide Deck** to present this architecture to client stakeholders?

**GenAI-Based Case Routing and Fulfilment Automation for Banking Client (Azure-Native)**

### 🏛️ Client Context

* Industry: Banking (regulated)
* Cloud: Azure
* Use Case: Automate case triage, routing, and fulfilment to improve SLA, reduce manual work, and meet compliance standards.

### 🔧 Solution Summary (GenAI-First, Azure-Native)

#### 1. Objectives

* Auto-classify and route customer service cases (email, form, chat)
* Generate fulfilment checklists based on case category and internal SOPs
* Provide explainability, audit logging, and RBAC compliance

#### 2. GenAI-Optimized Workflow

Customer Submits Case (Email/Form/Chat)

↓

Azure Function → Preprocessing (Redact PII using Presidio)

↓

Azure OpenAI (GPT-4)

→ Case Summary

→ Case Type Classification

→ Risk/Urgency Tagging

↓

LangChain + Azure Cognitive Search

→ Fetch SOP/KB from Blob

→ Route Decision (with rationale)

↓

Power Apps (Human-in-Loop Approval UI)

↓

Logic Apps → CRM Update + Task Checklist + Notifications

### 🔢 Azure Tech Stack

|  |  |  |
| --- | --- | --- |
| Layer | Service | Purpose |
| GenAI | Azure OpenAI (GPT-4) | Summarization, classification, checklist generation |
| Retrieval | Azure Cognitive Search | Fetch SOPs, escalation rules |
| Prompt Orchestration | LangChain (Azure Functions) | Multi-step logic, contextual prompts |
| PII Redaction | Presidio, Regex (Azure Functions) | Data sanitization before LLM prompt |
| Audit Logging | Azure Cosmos DB / SQL | Store input, output, prompt, timestamp, user ID |
| Integration | Power Automate / Logic Apps | Connect CRM, Notifications, Timer Escalation |
| UI Review | Power Apps | Approve/reject routing decisions manually |
| Security | Azure AD, Key Vault, Private Link | Role control, key storage, secure API access |

### 🌐 Example Prompts

#### Case Classifier

Classify this case: [Account Locked, Loan Query, KYC, Fraud, Transaction Issue]

Also assign: Risk Level [Low, Medium, High]

Case: "My account got locked after 3 failed attempts, and I got a message about fraud."

→ Classification: Fraud | Risk Level: High

#### Routing Decision

Using SOP:

- Fraud: Route to Fraud Ops

- KYC: Route to Compliance

- Loans: Route to Lending Desk

Determine routing + reason

→ Routing: Fraud Ops | Reason: Suspicious login detected

#### Checklist Generator

Given: Fraud Alert case type

Generate steps from policy:

→ 1. Freeze account

→ 2. ID verification

→ 3. Notify Risk Team

### 📓 Compliance Controls for Banking

|  |  |
| --- | --- |
| Control Area | Solution |
| PII Handling | Presidio-based redaction before prompt |
| Audit Logging | Store prompt/response with userID, time in Cosmos DB |
| RBAC | Enforce via Azure AD, scoped tokens |
| Data Residency | Host all services in UK/EU Azure region |
| Explainability | GenAI reasoning saved and displayed in UI |
| Model Governance | Prompt versioning + rollback with MLflow |

### ✅ Rollout Phases

|  |  |
| --- | --- |
| Phase | Milestone |
| Phase 1 | Case Summary + Classification via GPT |
| Phase 2 | Routing logic via LangChain + SOP lookup (RAG) |
| Phase 3 | Fulfilment planner + task checklist generation |
| Phase 4 | UI + Feedback loop (Power Apps) + CRM sync |
| Phase 5 | SLA violation prediction + multilingual support |

### 📅 Workshop Discovery Checklist

* Functional: Volume of cases? Types? Who owns routing logic?
* Technical: CRM system? API availability? SOP format (PDF, Confluence)?
* Compliance: PII policies? Data residency requirements? Audit expectations?
* Governance: Who signs off routing? How often should prompts be reviewed?

Prepared by: GenAI Architecture & DS Strategy Lead Target: ELT, Data, Tech, and Compliance Stakeholders

## 7. Discovery Questions for Planning Workshop

### ✅ ****Functional Questions****

1. **What are the primary case types** received (e.g., Fraud, KYC, Loan)?
2. How many **cases per day/week/month** are handled across CSMs?
3. What is the **average SLA** and current breach rate?
4. Is there a **standard set of fulfilment steps** or checklists for each case type?
5. What does the current **triage process** look like (decision points, escalation criteria)?
6. Who currently **owns routing decisions** — CSMs, Ops, or automated rules?
7. What level of **human-in-loop control** is expected post-routing?

### ✅ ****Technical Questions****

1. What system(s) store case data? (e.g., **Salesforce, ServiceNow, custom CRM**)
2. Is there API access to:
   * Ingest case data?
   * Update case routing/team assignment?
   * Retrieve knowledge documents or SOPs?
3. What is the **format and source** of the SOPs and internal routing guides?
4. Are historical cases **labeled with routing team and resolution time**?
5. Can case submissions be parsed in **real-time or near-real-time** (stream vs batch)?
6. Are we supporting **multilingual cases** or just English initially?
7. What is the average **length and structure** of incoming case text?

### ✅ ****Compliance & Data Governance Questions****

1. Is **PII or sensitive data** present in case notes or forms?
2. Are there internal **policies for redacting or masking PII** before processing?
3. Does the GenAI solution need to be hosted **entirely within UK/EU Azure regions**?
4. What are the **audit logging requirements** (prompt, response, user ID, time)?
5. How frequently will **model or prompt reviews** occur (monthly, quarterly)?
6. Are we permitted to use Azure OpenAI models (GPT-4) for routing decisions, or must models be isolated?
7. What **data retention rules** apply for prompts, case logs, and responses?

### ✅ ****Operational & Governance Questions****

1. Who owns and maintains the **routing SOPs** and escalation paths?
2. How are updates to SOPs communicated — and can this sync with GenAI logic?
3. Will **Power Apps or another UI** be needed for CSM approvals?
4. What KPIs will be used to track success? (e.g., SLA compliance, routing accuracy, CSM time saved)
5. Who signs off on routing **automation thresholds** (e.g., <80% confidence = human review)?
6. Who will manage the **prompt library and vector knowledge base** post-deployment?
7. Do we require a **fallback ML-based system** in parallel during early adoption?

Topic 2:

Certainly. Here's a comprehensive set of **discovery questions for planning workshops** focused on the **AI Servicing Agent – Answering Enquiries** use case for a **banking client hosted on Azure**, involving GenAI.

This solution involves building a **GenAI-powered agent** to assist Customer Success Managers (CSMs) in responding to frequent or repetitive queries — leveraging Retrieval-Augmented Generation (RAG), knowledge bases, and email/chat data.

## 🧩 Discovery Questions for Planning Workshop

**Use Case**: ✅ 2. AI Servicing Agent – Answering Enquiries  
**Goal**: Automatically respond to customer queries or assist CSMs with AI-generated draft replies.

### ✅ ****Functional Questions****

1. What **types of customer queries** are most common (e.g., status requests, document clarifications, account-related FAQs)?
2. What % of queries are **repetitive or templated** vs complex/escalated?
3. Are responses expected to be **automated** (sent directly) or **suggested** (with CSM approval)?
4. What’s the **targeted response time** for different query categories?
5. Is the aim to **deflect queries**, **assist drafting**, or **auto-resolve L1 tickets**?
6. Are there pre-existing **response templates** or macros CSMs use today?
7. Are customers served via **email, secure portal, chat**, or a mix?

### ✅ ****Technical Questions****

1. What system stores **customer interaction history**? (CRM, shared inbox, ticketing system?)
2. Is there an existing **knowledge base** or document set? If so:
   * What format (PDF, HTML, Confluence)?
   * How often is it updated?
   * Who owns it?
3. Are **previous queries and responses** stored in a structured way?
4. Can email threads or chat logs be **extracted programmatically** (e.g., via Outlook Graph API, Salesforce API)?
5. Do queries contain sensitive fields like **account numbers, personal IDs, or documents**?
6. What languages do customer queries arrive in? Do we need **multilingual LLM capability**?
7. How many queries do CSMs handle per day/week/month?

### ✅ ****Compliance & Risk Questions****

1. What **level of automation** is acceptable for a banking client? Full automation or draft + approval?
2. Should **disclaimers** be added to AI-drafted messages (e.g., “This message was AI-assisted”)?
3. Can **GenAI use internal data** (e.g., customer emails) for prompt context?
4. Where must the GenAI system be hosted? UK/EU only?
5. Is **response content** subject to compliance review (e.g., legal signoff on regulated answers)?
6. What level of **audit logging** is required (e.g., prompt, generated response, approver)?
7. What **hallucination prevention mechanisms** are required (RAG, prompt guards, confidence thresholds)?

### ✅ ****Governance & Operational Questions****

1. Who owns the **official source of truth** for answers — Product, Ops, Legal?
2. How often do policies or procedures (e.g., payment timelines, KYC processes) change?
3. Should **CSMs have override/edit ability** on AI-generated responses?
4. Who will own the **GenAI knowledge base** (indexing, versioning, retraining)?
5. Will we need an **internal feedback loop** to improve response quality (CSM corrections → re-ingest)?
6. What are the target KPIs:
   * Time saved per case?
   * Reduction in first-touch resolution time?
   * Deflection %?
7. What does the **post-response escalation** workflow look like (if GenAI fails to answer)?

### 🛠️ Optional UX/Feature Discovery Questions

* Should CSMs have access to a **sidebar assistant** inside the CRM for response suggestions?
* Do we want to **summarize historical context** for CSMs before drafting a response?
* Should the AI agent include **attachment or document recommendations** in replies?
* Will there be different agents/personas per product line or geography?

Would you like this workshop checklist converted into a:

* ✅ **Custom Notion/Confluence page**
* ✅ **Stakeholder deck (with workshop flow and swimlanes)**
* ✅ **Excel tracker to assign roles/owners per area**

Let me know which format you prefer!

Here is the **updated solution for ✅ 2. AI Servicing Agent – Answering Enquiries** reimagined from an **ML-first to a GenAI-first approach**, customized for a **banking client on Azure**, considering **compliance, PII handling, explainability, audit trails**, and **human-in-loop workflows**.

## ✅ Updated Solution: ****GenAI-Powered Servicing Agent for Customer Enquiries (Banking | Azure-Native)****

### 🎯 ****Use Case Objective****

Enable a **secure, compliant, and intelligent assistant** that supports CSMs by:

* Auto-suggesting accurate, contextual responses to customer queries
* Reducing response time and manual drafting effort
* Ensuring compliance by using internal knowledge and SOPs

## 🧠 Replacing ML with GenAI

| Traditional ML Component | GenAI Enhancement |
| --- | --- |
| Query classifier (ML model) | Zero/few-shot classification using **GPT-4 with examples** |
| FAQ Matching (TF-IDF/BM25) | Retrieval-Augmented Generation (RAG) using **Azure Cognitive Search + LangChain** |
| Template filler (rule-based) | Natural language **GPT completions** from prompt templates |
| Sentence generation model | **GPT-4 response drafts** with tone/style customization |
| Response ranking (scoring model) | Human feedback loop via **Power Apps** with override + rating |

## 🔁 Updated Process Flow (Azure, GenAI-centric)

[Customer Email/Chat Submitted]

↓

[Azure Function]

→ Preprocess + redact PII (Presidio)

↓

[Azure OpenAI (GPT-4)]

→ Generate: Summary, Category, Draft Response

→ Uses system prompt tuned for tone, compliance

↓

[RAG via Azure Cognitive Search]

→ Index KBs, SOPs, past answers

→ Retrieved documents fed to GPT for grounded response

↓

[Power Apps UI (Human-in-loop)]

→ CSM reviews, edits, or approves response

↓

[Logic App Workflow]

→ Send to CRM/email/chat

→ Log prompt + response + approver

## 🧰 Azure-Native Tech Stack (GenAI-first)

| Layer | Azure Service | Function |
| --- | --- | --- |
| **GenAI** | Azure OpenAI (GPT-4) | Drafts replies using context & tone prompt |
| **RAG Engine** | Azure Cognitive Search + Blob + LangChain | Retrieves relevant KBs, SOPs |
| **Preprocessing** | Azure Functions + Presidio | Redacts PII, formats queries |
| **Approval UI** | Power Apps | Human-in-loop control for compliance |
| **Workflow Engine** | Logic Apps / Power Automate | Sends message, tracks status |
| **Audit Logging** | Azure Cosmos DB / SQL / Log Analytics | Stores input, output, approver ID, time |
| **Security** | Azure Key Vault + Azure AD + Private Link | Keys, access, secure endpoint management |

## 🧾 Prompt Engineering: Key Patterns

### 1. ✉️ Response Drafting Prompt

You are a banking CSM assistant. Based on the query below, draft a professional response that:

- Is concise, accurate, and friendly

- Refers to relevant product or policy if applicable

- Avoids speculation

Query: "Can you confirm if my card replacement is on the way?"

Answer:

Hello [Customer Name],

Thank you for reaching out. Your replacement card was dispatched on [Date] and should arrive within 5 working days.

Let us know if you don't receive it by then.

Kind regards,

[CSM Team]

### 2. 📄 RAG Integration Prompt (Grounded Answers)

Given the internal policy below, generate a customer-facing answer.

Customer: “How do I upload my KYC documents?”

Policy Extract:

“To upload KYC documents, customers must use the secure portal sent by email. Upload must be completed within 7 days.”

Answer:

Please use the secure upload link we emailed you to submit your KYC documents. For your protection, this link expires after 7 days.

## 🧩 Value of GenAI-first Over ML-First

| Feature | ML-Based | GenAI-Based |
| --- | --- | --- |
| Adaptability | Fixed intent classes | Zero-shot, few-shot |
| Explainability | Feature importance (low clarity) | Human-readable justification |
| Speed to deploy | Months (labeling, training) | Days (prompt + RAG tuning) |
| Personalization | Hard-coded templates | Adaptive tone and context |
| Regulatory control | Requires rules engine | Prompt-based with human-in-loop |

## 🧱 Compliance Enhancements

| Control | Implementation |
| --- | --- |
| **PII Redaction** | Presidio + Regex filters before prompt |
| **Data Residency** | All components deployed in UK/EU Azure regions |
| **Audit Logs** | Prompt, response, RAG source, user ID logged in Cosmos DB |
| **Explainability** | Responses contain LLM-generated rationale + source |
| **Approval Flow** | All outbound comms approved via Power Apps by CSM |
| **Fallback** | CSM can override or generate manually if GenAI fails |

## 📈 Pilot Success Metrics (Target Benchmarks)

| Metric | Baseline | Target (GenAI) |
| --- | --- | --- |
| Avg response drafting time | 8–10 min | <1 min |
| First-touch resolution rate | 65% | >85% |
| Automation-ready query % | 0–10% | 40–60% (with human-in-loop) |
| CSM satisfaction | N/A | +30 NPS uplift |
| Hallucination rate (LLM) | N/A | <3% (RAG-grounded, monitored) |

## 📦 Rollout Plan (for Bank)

| Phase | Scope |
| --- | --- |
| **Phase 1** | GPT-only pilot on internal queries (non-customer-facing) |
| **Phase 2** | RAG setup + approval UI (Power Apps) |
| **Phase 3** | Rollout to 1 use case (e.g., KYC queries) |
| **Phase 4** | CRM integration (Salesforce/ServiceNow) + email send |
| **Phase 5** | Feedback loop → prompt tuning → scale to other teams |

Would you like a:

* 🎨 **Architecture diagram** like before?
* 📊 **Slide deck** summarizing this solution?
* 📋 **Prompt library template** with version control recommendations?

Let me know, and I’ll generate that immediately.

### 🔧 Solution Summary (GenAI-First, Azure-Native)

#### 1. Objectives

* Auto-classify and route customer service cases (email, form, chat)
* Generate fulfilment checklists based on case category and internal SOPs
* Provide explainability, audit logging, and RBAC compliance

#### 2. GenAI-Optimized Workflow

Customer Submits Case (Email/Form/Chat)

↓

Azure Function → Preprocessing (Redact PII using Presidio)

↓

Azure OpenAI (GPT-4)

→ Case Summary

→ Case Type Classification

→ Risk/Urgency Tagging

↓

LangChain + Azure Cognitive Search

→ Fetch SOP/KB from Blob

→ Route Decision (with rationale)

↓

Power Apps (Human-in-Loop Approval UI)

↓

Logic Apps → CRM Update + Task Checklist + Notifications

### 🔢 Azure Tech Stack

|  |  |  |
| --- | --- | --- |
| Layer | Service | Purpose |
| GenAI | Azure OpenAI (GPT-4) | Summarization, classification, checklist generation |
| Retrieval | Azure Cognitive Search | Fetch SOPs, escalation rules |
| Prompt Orchestration | LangChain (Azure Functions) | Multi-step logic, contextual prompts |
| PII Redaction | Presidio, Regex (Azure Functions) | Data sanitization before LLM prompt |
| Audit Logging | Azure Cosmos DB / SQL | Store input, output, prompt, timestamp, user ID |
| Integration | Power Automate / Logic Apps | Connect CRM, Notifications, Timer Escalation |
| UI Review | Power Apps | Approve/reject routing decisions manually |
| Security | Azure AD, Key Vault, Private Link | Role control, key storage, secure API access |

### 🌐 Example Prompts

#### Case Classifier

Classify this case: [Account Locked, Loan Query, KYC, Fraud, Transaction Issue]

Also assign: Risk Level [Low, Medium, High]

Case: "My account got locked after 3 failed attempts, and I got a message about fraud."

→ Classification: Fraud | Risk Level: High

#### Routing Decision

Using SOP:

- Fraud: Route to Fraud Ops

- KYC: Route to Compliance

- Loans: Route to Lending Desk

Determine routing + reason

→ Routing: Fraud Ops | Reason: Suspicious login detected

#### Checklist Generator

Given: Fraud Alert case type

Generate steps from policy:

→ 1. Freeze account

→ 2. ID verification

→ 3. Notify Risk Team

### 📓 Compliance Controls for Banking

|  |  |
| --- | --- |
| Control Area | Solution |
| PII Handling | Presidio-based redaction before prompt |
| Audit Logging | Store prompt/response with userID, time in Cosmos DB |
| RBAC | Enforce via Azure AD, scoped tokens |
| Data Residency | Host all services in UK/EU Azure region |
| Explainability | GenAI reasoning saved and displayed in UI |
| Model Governance | Prompt versioning + rollback with MLflow |

### ✅ Rollout Phases

|  |  |
| --- | --- |
| Phase | Milestone |
| Phase 1 | Case Summary + Classification via GPT |
| Phase 2 | Routing logic via LangChain + SOP lookup (RAG) |
| Phase 3 | Fulfilment planner + task checklist generation |
| Phase 4 | UI + Feedback loop (Power Apps) + CRM sync |
| Phase 5 | SLA violation prediction + multilingual support |

### 📅 Workshop Discovery Checklist

* Functional: Volume of cases? Types? Who owns routing logic?
* Technical: CRM system? API availability? SOP format (PDF, Confluence)?
* Compliance: PII policies? Data residency requirements? Audit expectations?
* Governance: Who signs off routing? How often should prompts be reviewed?

Prepared by: GenAI Architecture & DS Strategy Lead Target: ELT, Data, Tech, and Compliance Stakeholders

### ✅ AI Servicing Agent – Answering Enquiries (Banking | Azure-Native)

#### Objective

Support CSMs with GenAI that:

* Auto-summarizes incoming queries
* Drafts grounded responses using internal SOPs
* Allows CSM approval before sending

#### GenAI Workflow

Customer Email/Chat → Azure Function (PII Redaction)

↓

Azure OpenAI (GPT-4) → Summary + Draft Response

↓

Azure Cognitive Search (RAG) → Fetch KB/SOP context

↓

LangChain → Prompt Rerouting + Reasoning

↓

Power Apps UI → Human Review + Edit + Approve

↓

Logic App → Send to CRM/Email, Audit Logs

#### Prompt Examples

**Draft Email Response Prompt:** "Customer asked: 'When will my card arrive?' → Response: 'Your card was dispatched on Monday and should arrive in 3–5 business days.'"

**Grounded RAG Prompt:** "Customer: 'How do I upload my KYC?' Policy: 'Use the secure upload link within 7 days.' → Response: 'Please use the emailed link to upload your KYC docs within 7 days.'"

#### Tech Stack (Same as Routing use case, +Email Connector via Graph API or SendGrid)

#### Compliance Notes

* No auto-send without review (bank policy)
* Log CSM who approved final draft
* Redact account numbers and PII before prompting
* GPT-4 hosted in Azure UK region

#### Pilot Metrics

* Reduce draft time: 10 → 1 min
* CSM satisfaction ↑ 30%
* Automation-ready queries: 40–60%