

# VPBank Technology Hackathon 2025

## General Brief

Please fill up this table and use this document as a template to write our proposal.

Challenge Statement	Hack Challenge 4: AI-Powered Banking Process Redesign
Team Name	Nhóm 247 (Update to SoftAI)

## Team Members

Full Name	Role	Email Address	School Name (if applicable)	Faculty / Area of Study	LinkedIn Profile URL
Nguyễn Văn Tú	Team lead + AI dev	<a href="mailto:nvantu2305@gmail.com">nvantu2305@gmail.com</a>	University of Science, VNUHCM	Information Technology	<a href="https://www.linkedin.com/in/nguyen-van-tu/">https://www.linkedin.com/in/nguyen-van-tu/</a>
Dương Trung Nghĩa	Backend + AI dev	<a href="mailto:trungnghia.24904@gmail.com">trungnghia.24904@gmail.com</a>	University of Science, VNUHCM	Information Technology	<a href="https://www.linkedin.com/in/nghia-zun/">https://www.linkedin.com/in/nghia-zun/</a>
Phạm Thành Nhân	Frontend developer	<a href="mailto:ptnhanit230104@gmail.com">ptnhanit230104@gmail.com</a>	University of Science, VNUHCM	Information Technology	<a href="https://www.linkedin.com/in/nhan-pham-53a891323/">https://www.linkedin.com/in/nhan-pham-53a891323/</a>
Lương Xuân Thanh	DevOps + Architecture Solution	<a href="mailto:lxthanh235@gmail.com">lxthanh235@gmail.com</a>	University of Science, VNUHCM	Information Technology	<a href="https://www.linkedin.com/in/luongthanh/">https://www.linkedin.com/in/luongthanh/</a>
Đinh Nguyễn Lan Anh	DA + BA	<a href="mailto:anhdn122@uef.edu.vn">anhdn122@uef.edu.vn</a>	University of Economics and Finance	Financial Technology	<a href="https://www.linkedin.com/in/ddnlanh/">https://www.linkedin.com/in/ddnlanh/</a>

## Content Outline

Section	Page No.
<b>Addition Resources</b>	3
<b>Solutions Introduction</b>	4
<b>Impact of Solution</b>	6
<b>Deep Dive into Solution</b>	11
<b>Architecture of Solution</b>	23

## Addition Resources

[1]. VPFlow – Figma prototype demonstratation video: [https://youtu.be/hY\\_JsGxNueI](https://youtu.be/hY_JsGxNueI)

- Highly recommend watch after read the ***Solutions Introduction*** section

[2]. VPFlow – Figma: [VPFlow Figma Link](#)

## Solutions Introduction

### What is VPFlow?

**VPFlow** is an **AI-powered, centralized platform** built to help banks - especially in complex operational environments like **VPBank** - to **comprehend, visualize, optimize, and continuously improve** internal workflows. Instead of static flowcharts and disconnected SOP files, VPFlow turns our internal processes into **interactive, intelligent diagrams**, which we call **Dynamic Swimlane diagram (footnote)**.

Powered by **Generative AI, LLMs, Knowledge-augmented generation** and **AWS infrastructure**, **VPFlow** bridges the gap between traditional operations and intelligent automation.

#### What makes VPFlow different?

- **From SOPs to diagrams:** VPFlow auto-generates swimlane diagrams from unstructured documents, SOPs, and process narratives using fine-tuned LLMs (e.g., Deekseek R1).
- **Real-time interaction:** Each diagram node is clickable and explorable. Users can ask questions about each step and receive natural language explanations via AI agent.
- **Interactive agent tools** that enable users to engage directly with workflows via an **intuitive graphical user interface**
- **Bottleneck detection & AI suggestions:** System identifies pain points based on user feedback, time metadata, and process logic. GenAI offers improvement recommendations instantly.
- **Knowledge-rich database:** By integrating knowledge graphs with vector embeddings, the platform constructs a dynamic, LLM-ready context layer - enabling deep understanding of each workflow component.
- **Powered by AWS:** With AWS Textract, SageMaker, Lambda, Neptune and more, **VPFlow** ensures enterprise-grade scalability, reliability, and cost-efficiency.

VPFlow offers a **comprehensive experience** for both **new and experienced banking professionals**, streamlining complex operational processes into accessible visual models.

The platform leverages **Large Language Models (LLMs)** - such as a fine-tuned version of **Deekseek R1** for diagram generation-and combines **knowledge graphs** with **vector embeddings** to build an LLM-friendly, context-rich database.

Meanwhile, **AWS services** power the system's **intelligence, scalability, and adaptability**, making VPFlow a truly enterprise-grade solution for process intelligence.

**Our Mission:** To provide a dynamic and interactive platform that makes the integration of AI technologies-such as Large Language Models, knowledge graphs, and process mining-seamless and

accessible for banks. VPFlow empowers institutions to transform static operational knowledge into living, visual workflows that evolve through automation, collaboration, and continuous improvement.

- **Bankers:** to navigate complex internal processes visually, ask questions in plain language, and propose improvements.
- **Experts/Process Owners:** to receive user feedback, simulate changes, and continuously optimize operations - faster, smarter, and with full visibility.

## Why AWS Services?

### ***AWS is not just infrastructure - it's the intelligence backbone of VPFlow***

VPFlow leverages AWS not merely for hosting, but as a modular engine that powers the platform's scalability, real-time capability, and enterprise readiness:

- **Serverless & real-time:** AWS Lambda and EventBridge handle SOP ingestion, diagram generation, and user interaction in real time - without infrastructure overhead.
- **AI model orchestration:** SageMaker and Bedrock power the extraction of workflows from unstructured documents, bottleneck detection, and contextual responses using LLMs.
- **Semantic workflow intelligence:** Amazon Neptune fuses graph relationships and vector embeddings to enable business-aware search and logic validation across processes.
- **Continuous learning:** User feedback feeds into SageMaker's reinforcement learning loop, allowing the system to adapt and improve optimization suggestions over time.
- **Secure and compliant:** IAM, KMS, and CloudTrail ensure strict access control, encryption, and auditability - essential for financial-grade compliance.
- **Flexible integration:** Services like S3, DynamoDB, and DataSync support hybrid deployments and easy integration with core banking systems.

In short, AWS allows VPFlow to scale intelligently, respond instantly, learn continuously, and comply rigorously - meeting both technical and regulatory demands of modern banks.

*For Better Comprehend, please watch the demonstration:*

[https://youtu.be/hY\\_JsGxNuel](https://youtu.be/hY_JsGxNuel)

## Impact of Solution

### How does our solution benefit the society / the target audience?

*We are not merely building a tool - we are creating a “second brain” for the bank.*

#### 1. Redefining Internal Operational Capabilities in Modern Banking

Our solution is not just a diagramming platform, but rather an Operating System for Workflow Intelligence - a centralized system that enables banks to comprehend, monitor, optimize, and continuously automate their internal processes. From a raw SOP file, the system automatically:

- Extracts and visualizes the process in a swimlane format, identifying roles, conditions, and business flows.
- Integrates Generative AI to analyze logic, detect bottlenecks, simulate what-if scenarios, and measure operational impacts.
- Transforms each process step into a dynamic interaction point, where employees can click to read, inquire, give feedback, or suggest improvements.

As a result, the entire operation becomes transparent, measurable, learnable, and continuously improvable - surpassing traditional Business Process Management (BPM) systems.

#### 2. High Practical Relevance: Addressing Real-World Bottlenecks Identified in the Field

This solution was developed through extensive field interviews and direct observations at banks such as ACB, VPBank, and Techcombank - where we identified pressing challenges such as:

- Fragmented processes without metadata or version control, leading to time-consuming information retrieval.
- Delays in CIC (Credit Information Center) approvals due to opaque, manual underwriting flows and logic errors.
- Unsynchronized CRM data, resulting in misrouted or unprocessed requests.
- Lack of a systematic feedback loop, rendering processes “frozen” and disconnected from operational reality.

Our system addresses these issues at the root by converting operational knowledge into live, queryable data, eliminating dependence on static documents or manual meetings.

#### 3. Laying the Foundation for AI-Driven Digital Banking

Our solution is not only designed for current operations but also serves as an infrastructure for the future of digital banking, enabling:

- Neo-banking capabilities, where each process step can self-analyze, flag anomalies, and support FinTech onboarding, identity verification, and anti-fraud operations.
- Live compliance monitoring systems, capable of real-time process control and auto-generating audit logs.
- RPA 2.0, where repetitive tasks are automatically identified and proposed for automation without the need for coding.

The entire system is built on an AWS cloud-native stack, including:

- Bedrock/SageMaker: for diagram generation, semantic analysis, and process-related Q&A.
- Neptune/DynamoDB: for storing workflow diagrams as analyzable graphs.
- Textract + S3: for parsing input SOP documents.
- Lambda + API Gateway: a flexible serverless architecture ensuring fast responses and high security.

This architecture ensures full scalability across the bank without disrupting existing systems, and seamless integration with core internal platforms (CRM, DMS, Core Banking, eKYC, ticketing, etc.).

#### 4. Social Impact – Enhancing Efficiency Across the Financial Sector

At the societal level, the system brings about:

- A more transparent and equitable work environment for bank employees.
- Reduced processing time for customer services, especially in lending, identification, and account opening.
- A cost-efficient digital transformation platform for small and mid-sized banks in Vietnam.

In summary: This is not merely a technical solution but a strategic enabler that supports the Vietnamese banking sector in transitioning to an intelligent, adaptive, and continuously optimized operating model.

### Why our solution is a good solution? How is it better than existing solutions in the market / competitors?

*Because it goes beyond the limitations of current tools and lays the foundation for intelligent banking operations.*

#### 1. Aligned with the Practical Needs of the Vietnamese Banking Sector

Based on field surveys and interviews with operational staff at major Vietnamese banks such as ACB, Techcombank, and VPBank, our research identified several systemic challenges:

- Business processes (SOPs, guidelines, forms) are scattered across departments, lacking centralization and metadata, resulting in difficulties in retrieval and consistency maintenance.
- Employees have limited visibility into the relationships between process steps and lack intuitive tools to understand roles and workflows holistically.
- Operational bottlenecks (e.g., delayed CIC checks, internal handovers, CRM input errors) are not systematically tracked or proactively flagged.
- Process updates and improvement proposals are typically manual and lack tools for simulation or performance assessment prior to implementation.

These issues cannot be effectively addressed using existing tools such as Microsoft Visio, Lucidchart, or traditional BPMN systems, which mainly provide static diagramming capabilities and lack business logic comprehension or advanced analytical features.

#### 2. Existing Tools are "Isolated" - Our Solution is an "Ecosystem"

Most existing solutions in the market (Lucidchart, Visio, Signavio, etc.) are standalone diagramming tools designed primarily for presentation or static business flow design. They lack business context understanding, real-time data integration, continuous improvement support, and adaptability to the complex banking environment.

In contrast, our solution:

- Understands the semantics of SOPs and disparate documents to auto-generate process diagrams.
- Links process steps with real-time data, metadata, and business roles.
- Enables users to interact, give feedback, and suggest improvements directly on the diagram interface.
- Integrates seamlessly into existing bank ecosystems (CRM, Core Banking, eKYC, etc.).

#### 3. Core Technological and Functional Differentiation

Our solution leverages Generative AI (GenAI) to analyze SOPs, operational guidelines, and user feedback to:

- Auto-generate swimlane process diagrams instead of relying on manual diagramming.
- Detect logical errors such as missing handlers, infinite loops, or ambiguous conditions.

- Recommend process improvements at bottlenecks using metadata such as time delays, wait steps, or user feedback.
- Simulate the impact of process changes (e.g., merging steps, changing handlers, altering approval conditions) before deployment.
- Incorporate frontline feedback into each process step to enable continuous process optimization.

These features are not fully integrated or user-friendly in current market tools - especially under the decentralized operations and domain-specific requirements of the Vietnamese banking industry.

#### 4. Positioned as a "Smart Control Layer" Spanning the Entire Organization

Rather than replacing existing systems, we introduce a new control layer that:

- Empowers any employee to search, understand, and improve active processes.
- Reuses operational data to auto-generate training materials, suggest automation, or flag process errors.
- Converts the entire business workflow into a living graph, enabling leadership to visualize bottlenecks, access controls, and illogical flows - all in real time.

#### 5. Comparative Assessment with Market Alternatives

Evaluation Criteria	Microsoft Visio / Lucidchart	SAP Signavio / Bizagi	Proposed Method
Auto-generation of process diagrams from SOP text	No	Yes, with advanced configuration	Yes, using GenAI
Logic error detection in process flows	No	No	Yes, AI-powered warnings
Improvement suggestions based on internal data	No	Partial (via expert analysis)	Yes, using AI and staff feedback
What-if simulation of process changes	No	Yes, but complex and hard to access	Yes, simplified for business users
Attachment of documents at each process step	Manual only, not standardized	Yes, in enterprise platforms	Yes, with flexible file/metadata linking
Suitability for Vietnamese banking operations	No	Yes, but expensive and complex to deploy	Yes, field-tested and lightweight

#### 6. Full Leverage of AWS Capabilities to Realize the Solution

We utilize AWS services not just for technical support, but as strategic components in our solution architecture:

- Textract: Automatically reads and extracts logic with from original SOP documents, minimizing manual input.
- Bedrock / SageMaker: Powers GenAI core functionalities such as diagram generation, logic analysis, and conversational interfaces.
- DynamoDB / Neptune: Builds process graph structures for bottleneck detection and scenario simulation.



- Lambda + API Gateway: Implements a serverless architecture for cost efficiency, instant scalability, and secure integration with internal systems.

We do not simply “use AWS” - we architect the solution around AWS’s infrastructure strengths, enabling performance, scalability, and ease of integration.

## Our solution’s competitive advantage / unique selling point?

*A “living command center” for banks - where processes are not only visualized, but also learned, improved, and self-adaptive.*

Our solution offers a unique combination of **business process comprehension via GenAI**, **interactive swimlane visualization**, and **continuous process optimization**, specifically tailored to the operational context of Vietnamese banks. While existing platforms (e.g., Lucidchart, Signavio, Camunda) primarily focus on diagramming or high-level BPM modeling, our solution distinguishes itself in several critical areas:

### 1. Auto-generation of Process Maps from Unstructured Text - No Drawing or Configuration Required

Current tools like Visio, Lucidchart, or Camunda require users to have domain knowledge and manually draw process diagrams. This is time-consuming, not scalable across staff, and prone to inconsistencies. Our solution:

- Takes input as SOPs, internal guidelines, or PDF documents.
- Uses GenAI to automatically identify roles, flows, conditions, and related documentation.
- Generates interactive swimlane diagrams - transforming static documents into living data.

This reduces process mapping time by up to 80% and democratizes process understanding for non-technical staff.

### 2. AI-Powered Logic Analysis and Process Performance Optimization

Unlike current static diagrams, our system enables:

- Clicking each step to view the responsible party, associated documents, and input/output conditions.
- Interacting at each process point - users can access documents or directly ask questions via AI assistant.
- Capturing in-process feedback from frontline staff - no more scattered suggestions lost in email threads.

As a result, the process map becomes a smart operational blueprint that can learn and self-improve over time.

### 3. Step-Level Feedback Integration and Logic-Driven Optimization

We train GenAI not just to read SOPs, but to understand business logic, enabling:

- Logic validation - detecting design flaws (e.g., missing handlers, infinite loops).
- Bottleneck detection based on metadata, time duration, and user interactions.
- Improvement suggestions such as step merging, role consolidation, or automation opportunities.

This promotes continuous, practical, and staff-driven process optimization.

### 4. What-if Simulation - Model Process Changes Before Deployment

We integrate simulation capabilities to:

- Merge steps → quantify time savings.
- Remove approval layers → assess associated risk increases.
- Shift tasks to RPA → estimate ROI of automation.

Each change is quantified and validated before implementation, enabling data-driven decisions rather than intuition-based ones.

## **5. Built for Banking - Scalable to Any Process-Driven Organization**

Our features are designed specifically for banking use cases, including:

- Credit underwriting
- eKYC and CRM workflows
- Customer complaints and account opening

However, the platform is fully extensible to:

- Insurance (claims, verification)
- FinTech (onboarding, fraud detection)
- Large enterprises (HR workflows, internal control)

This is more than a product - it's a "process graph + GenAI + feedback loop" platform for any organization seeking efficient, transparent, and adaptive operations.

**In summary:** Our solution is not merely a process visualization tool - it is an intelligent, banking-specialized platform that connects operations, compliance, and technology into a cohesive ecosystem that is easy to deploy, scalable across

## Deep Dive into Solution

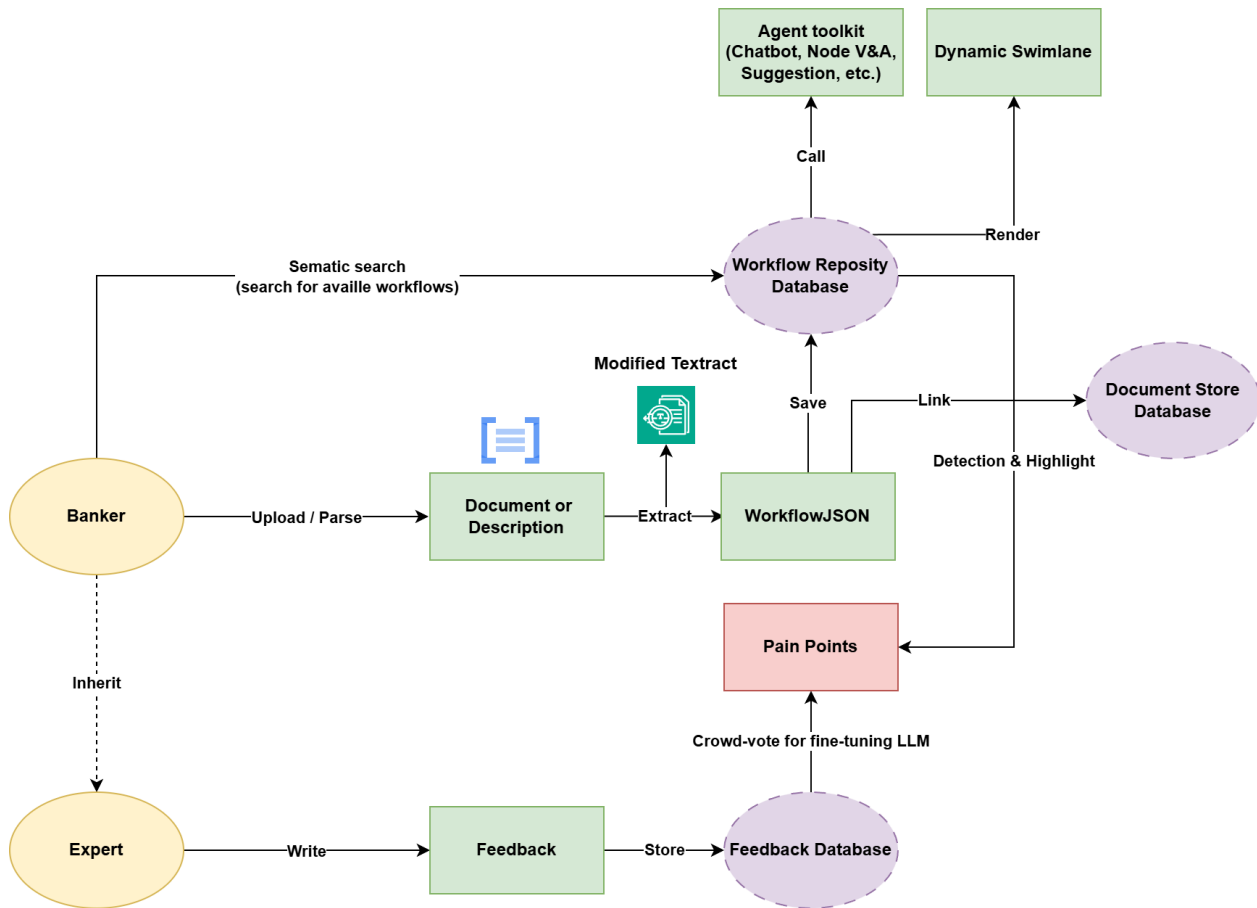


Image 1: VPFlow – High Level Data Flow Diagram

### EPIC 1 – Workflow Management

This EPIC aims to provide bankers with the ability to upload, extract, and manage workflow processes from multiple document formats (e.g., SOPs, DOCX, PDF), while supporting version control over time. Additionally, it enables advanced multilingual search to retrieve workflows based on keywords or business context.

#### 1.1 Upload / Search Workflow

**Purpose:** Users can upload a new workflow document or search existing workflows using business-related keywords in natural language across multiple languages.

**Key Capabilities:** Supports **semantic keyword search** using vector embeddings (e.g., pgvector, FAISS), allowing for **free-text input** with **multilingual normalization**, including diacritic removal, slang

correction, and typo tolerance. Each workflow stores **rich metadata**, such as process title, actors, department, frequency, and objective. Results can be **filtered** by tags like department, role, or last modified date.

**Architecture:**

- Text is embedded using a multilingual language model (BGE-M3), if Vietnamese detected ViDense will be used for State-of-the-art performance.
- Indexing engine supports both full-text and semantic queries.
- Optional tag-based filtering by department, role, or last modified date.

**Output:** Delivers a **ranked list of relevant workflows**, with options to **preview summaries** or open full visualizations (linked to EPIC 2.1).

## 1.2 Parse SOP / PDF / DOCX

**Purpose:** Users can upload standard operating procedure documents in unstructured formats (PDF, DOCX, etc.), and the system will automatically extract a structured JSON representation of the workflow, including: Steps, Actors, Dependencies, Conditional logic (if/else, approval, exception handling)

**Key Capabilities:**

- The parsing process leverages a **multi-stage GenAI pipeline** that includes:
  - **OCR** for scanned documents,
  - **Layout-aware parsing** to detect tables, forms, or flows,
  - And **LLM-based entity extraction** to identify multilingual elements such as steps, actors, conditions, and references.
- The final output is a **visualization-compatible JSON**, ready for confirmation and editing before rendering.

**Architecture:**

- Document Preprocessing → Chunking → Prompt-based multilingual extraction → Post-processing → JSON schema validation.
- Uses prompt templates trained on domain-specific language (e.g., enterprise SOPs, regulatory documents across supported languages).

**Output:** A **structured, editable JSON file** representing the workflow, with the option to **review and confirm** before visualization.

## 1.3 Version Control & History

**Purpose:** Enables comprehensive version management for workflows—users can **track changes**, **restore previous versions**, or **compare edits** over time with full traceability.

**Key Capabilities:** Each workflow change **auto-generates a version ID**, and a **Git-like diff engine** highlights differences in diagram structure. Annotations can include who made changes, when, and why (if enabled via comments).

**Architecture:**

- Workflows stored in JSON format with hashed version IDs.
- Delta storage or full snapshot depending on change granularity.
- Timeline UI to navigate and visualize previous versions.

**Output:** An intuitive **version timeline panel**, with tools to **compare**, **revert**, or **audit changes**, supporting full accountability and rollback functionality.

## EPIC 2 – Workflow Visualization

This EPIC provides an interactive, user-friendly interface to visualize, interact with, and analyze business workflows. The focus is on clarity (via swimlanes), interactivity (node insights), and optimization (pain point detection).

### 2.1 Interactive Diagram (Swimlane)

**Purpose:** Renders each workflow as an interactive swimlane diagram, organizing tasks by actors/roles (e.g., customer, staff, supervisor).

**Key Capabilities:** Diagram engine (e.g., Mermaid.js, D3.js, or custom Canvas) supports:

- Lane separation by actor
- Sequential and parallel flows
- Gateways (if/else)
- Real-time updates when workflow JSON is modified.

**Architecture:** JSON → AST → Renderer engine (Mermaid parser or D3-based DSL). Diagram components are clickable (linked to EPIC 2.2).

**Output:** A responsive swimlane diagram with visual actor separation and structural clarity, ready for analysis and interaction.

### 2.2 Clickable Node

**Purpose:** Each task or node in the diagram is interactive. Clicking on it reveals:

- Responsible actor(s): Example

- Estimated duration
- Linked forms or files
- SLA (Service Level Agreement), deadlines, notes

**Key Capabilities:**

- Tooltip + side panel on click.
- Rich metadata display from workflow JSON.
- Support for role-based access (e.g., editors see edit button, viewers see only details).

**Architecture:** Node registry with event listeners. Contextual panel rendered on click.

**Output:** A dynamic side panel showing detailed task metadata, with navigation support between connected steps.

## 2.3 Highlight Pain Points

**Purpose:** Automatically scans the workflow and highlights **pain points**, which are steps that:

- Take excessive time.
- Involve multiple handoffs or actors.
- Frequently cause delays, confusion, or rework.

**Key Capabilities:** Pain point scoring algorithm:

- Step duration × number of dependencies × actor transitions
- Optionally incorporates execution logs or heuristics
- Visual cue (e.g., red outline or glow) on affected node(s)

**Architecture:** Pain scores are computed in a preprocessing stage and visualized using threshold-based color cues (e.g., red glow, warning icons).

**Output:** An annotated diagram highlighting pain points. Hovering over a node reveals a quick explanation and score.

## 2.4 Pain Point Explanation

**Purpose:** AI explains why a step is marked as a **pain point** and optionally suggests ways to improve it.

**Key Capabilities:** Uses LLM with custom prompt template:

- “Explain why this step is problematic.”
- “Why is this handoff causing issues?”
- “How can we improve this part of the process?”

May reference prior logs or organizational KPIs if available.

**Architecture:** Node metadata and flow context are sent to a templated prompt, and results are rendered inline via tooltip or collapsible panel.

**Output:**

- Natural language explanation (e.g., “This step involves 3 departments and lacks automation”).
- Optional links to AI suggestions (integrated with EPIC 3.4).

## EPIC 3 – GenAI-powered Analysis & Assistant

This EPIC leverages Generative AI to analyze workflows at multiple levels — from high-level assessments to step-specific insights — enabling users to understand, validate, and improve complex processes through natural language interaction.

### 3.1 Global AI Assistant

**Purpose:** Empowers users to ask high-level, natural-language questions about the entire workflow diagram — such as “Which step is the most complex?”, “What documents or forms are needed?”, or “Is this process logical and efficient?”

**Key Capabilities:**

- Converts swimlane structure into a knowledge graph (actors, steps, dependencies)
- Uses LLM + RAG (Retrieval-Augmented Generation) to ground answers in business metadata, rules, and past executions
- Provides answers in natural language with optional highlights or summaries

**Architecture:**

- Swimlane JSON → Knowledge Graph (nodes, actors, relations)
- Query Parser → Prompt Generator with injected metadata
- RAG pipeline: vector search on past data → LLM synthesis (e.g., GPT-4, Claude)
- Frontend shows annotated results with highlight overlays

**Output:** Users receive actionable insights across the entire workflow, often with clickable suggestions or links to deeper drill-downs.

### 3.2 Node-Level Q&A

**Purpose:** Users can select a specific step (node) and ask targeted questions about that task’s context, conditions, and responsibilities.

Examples:

- “What triggers this task?”

- “Who is responsible here?”
- “What risks are associated with this step?”

**Key Capabilities:** Generates dynamic prompts based on the selected node’s metadata — including connected steps, roles, conditions, and historical execution data. Answers are focused and informative.

**Architecture:**

- Node selection triggers metadata extraction (actor, conditions, in/out edges)
- Prompt template filled dynamically with node-local context
- Optional RAG augmentation using internal process logs
- UI renders answers in a collapsible side panel

**Output:** Clear, targeted responses related to the chosen step, with options to explore deeper or simulate changes.

### 3.3 Smart Error Detection

**Purpose:** The system automatically scans the diagram to identify structural or logical issues, such as:

- Tasks without assigned actors
- Gateways missing conditions
- Infinite or invalid loops
- Disconnected or redundant nodes

**Key Capabilities:** Converts workflows into DAGs (Directed Acyclic Graphs) for rule-based static analysis. Uses LLM to explain detected issues in plain, accessible language and marks them visually.

**Architecture:**

- Workflow JSON → DAG conversion
- Rule engine checks for graph violations and anti-patterns
- Findings passed through explanation prompt → LLM output
- Frontend highlights faulty nodes with tooltips and fix suggestions

**Output:** Combined visual/textual error reports, with friendly explanations and fix recommendations. Admins can preview or auto-apply corrections.

### 3.4 Suggest redesign

**Purpose:** The AI recommends structural improvements to optimize efficiency — for instance by merging redundant steps, parallelizing independent paths, or balancing workloads across actors.



**Key Capabilities:**

- Applies transformation rules (e.g., merge sequential tasks, parallelize paths, balance workload)
- Leverages logs and organizational best practices
- Uses LLM to explain before/after changes

**Architecture:**

- Baseline workflow JSON → transformation engine applies optimizations
- Comparison module parses ASTs of both versions
- Prompts generated to summarize improvements and trade-offs
- Frontend shows side-by-side diagrams + natural language summary

**Output:** An improved version of the workflow, complete with AI commentary, visual diff view, and the ability to accept, reject, or edit proposed changes.

## EPIC 4 – Simulation & Optimization

This EPIC enables **interactive simulation** of process modifications and **data-driven optimization**. Users can run what-if scenarios, evaluate risks, and compare process versions based on time, cost, resource efficiency, and compliance.

### 4.1 What-if Simulation

**Purpose:** Enables users to simulate **hypothetical edits** to the workflow — such as removing a step, reassigning responsibilities, adding delays, or introducing automation — and immediately observe downstream effects.

**Key Capabilities:** Each simulation estimates how proposed changes affect overall process performance: including total execution time, actor workload distribution, and potential logic issues or risk exposure. The system presents results in intuitive graphs and summaries.

**Architecture:**

- Converts dynamic swimlane into an executable simulation graph with annotations for time, cost, and resources.
- Uses probabilistic simulations (e.g., Monte Carlo) or rule-based logic propagation to model the outcome of edits.
- Results are cached and visualized in real-time.

**Output:** Users receive dynamic charts and summaries quantifying changes in KPIs, such as “+2 hours, -1 actor, risk increased at gateway X”. Multiple simulation scenarios can be toggled for comparison.

## 4.2 Compare Flows (Before vs. After)

**Purpose:** After a simulation or AI-driven redesign, the system compares two process versions side-by-side:

- Structural differences (nodes, branches, loops).
- KPI metrics (time, task count, complexity, actor distribution).
- AI verdict: is the new version better? Why?

**Key Capabilities:** The comparison highlights structural differences (nodes added/removed, logic branches changed), along with metrics like task duration, actor complexity, and workload balance. AI evaluates whether the redesigned version is more optimal and explains why.

**Architecture:**

- Parses two dynamic swimlane versions into abstract syntax trees (ASTs).
- Compares control flow, actor involvement, and estimated metrics.
- AI uses prompt templates to summarize deltas and justify conclusions.

**Output:** An interactive diff view with color-coded diagram changes, a breakdown of metric deltas in tables, and an AI-generated explanation stating whether — and why — the new process is better.

## 4.3 AI Justification / Risk Analysis

**Purpose:** Provides AI-generated narratives that explain the risks or implications of a given workflow change. Questions like *“What could go wrong if we remove the approval step?”* or *“Is combining steps A and B safe?”* are answered clearly.

**Key Capabilities:** The system applies encoded business rules, regulatory constraints, and past error patterns to simulate possible failure paths. It retrieves and integrates similar historical examples when available.

**Architecture:**

- Encodes domain-specific business rules, regulatory constraints, and known failure cases.
- AI simulates potential failure paths and retrieves relevant case studies or logs.
- Risk is explained in a human-readable, cause-effect narrative.

**Output:** A concise narrative lists potential risks in bullet or paragraph form, along with severity ratings and possible mitigation strategies. Results are optionally linked to simulation outcomes or archived events.

## EPIC 5 – Knowledge & Document Support

This EPIC connects workflow steps with relevant documents and automates the generation of SOPs (Standard Operating Procedures) in multiple languages, ensuring that all process knowledge is consistently accessible and up-to-date.

### 5.1 Link to SOP / Forms / Templates

**Purpose:** Each workflow step can be connected to relevant internal documentation — such as SOP manuals, forms, or shared cloud folders (e.g., Google Drive, SharePoint).

**Key Capabilities:** Users can attach or view files directly from the UI at the task level. Both static URLs and dynamic cloud integrations are supported, with permission-based access for viewing, downloading, or editing files depending on user role.

**Architecture:** Each task node includes a linked\_docs metadata field that connects to external document management APIs.

**Output:** Documents appear as embedded icons or preview panels within the diagram. Users can open files in a new tab or inline viewer.

### 5.2 Auto SOP Generator

**Purpose:** Automatically creates a structured Standard Operating Procedure document derived from the visual workflow.

**Key Capabilities:**

- AI parses diagram into a sequence of:
  - **Heading:** Task name
  - **Action:** Purpose of what to do
  - **Actor:** Who performs the task
  - **Rule:** Conditions, dependencies, or SLA
- Optional export as DOCX, PDF, or Markdown.

**Architecture:** Workflow JSON is fed into a prompt template tailored for SOP generation. The LLM produces the SOP content, which is then formatted into exportable documents using a formatter module.

**Output:** A ready-to-use SOP document aligned with the workflow, suitable for internal distribution, audits, or compliance use.

### 5.3 Multi-language Support

**Purpose:** Support SOP generation in **multiple languages**, with a focus on:

- Vietnamese (formal tone for government/internal use)
- English (reporting or documentation with partners)

**Key Capabilities:** A language switcher allows users to toggle between SOP languages. Both AI-generated and static text (headings, labels) are translated. Tone can be controlled to match use cases — such as formal, simplified, or business-style.

**Architecture:** A multilingual LLM (e.g., GPT-4 Turbo or NLLB-200) performs translation and tone adjustment. Both language versions can be stored in parallel with version synchronization.

**Output:** SOPs are available in the selected language, with the option to download or view bilingual versions side-by-side for easier communication or compliance reporting.

## EPIC 6 – Feedback & Improvement Loop

This EPIC introduces mechanisms to continuously refine workflows through expert-driven feedback, crowdsourced insight on **pain points**, and controlled updates.

### 6.1 Receive User Feedback

**Purpose:** Allows experienced users (e.g., bankers or operators) to submit feedback on steps that may be outdated, incorrect, missing context, or causing delays.

**Key Capabilities:** An inline feedback form is embedded in each annual section of the workflow, supporting both anonymous and named submissions. Users can tag their feedback with predefined labels such as “*Confusing*”, “*Too slow*”, or “*Needs update*” for quicker triage and categorization.

**Architecture:** Each feedback entry is associated with the relevant task ID and stores metadata including user identity (if available) and timestamp. Feedback is aggregated in a centralized admin dashboard to streamline triage and prioritization.

**Output:** Admins receive a compiled list of feedback items per task, along with a heatmap visualization that highlights workflow steps receiving frequent concern or critique.

### 6.2 Crowd-vote Pain Point

**Purpose:** Let employees vote on steps they frequently encounter problems with - creating a community-driven heatmap of **pain points**.

**Key Capabilities:** Each node in the workflow diagram includes an upvote button labeled “*This step often causes issues*”. Votes are tallied and visualized directly on the diagram through overlays (e.g., color gradients or emojis). Users can filter voting data by department, role, or specific time periods to identify contextual bottlenecks.

**Architecture:**

- Per-node voting registry (unique per user per version)
- Visualization overlay on diagram (e.g., color gradient or emoji)

**Output:**

- Real-time crowd rating on workflow **pain points**
- Can be exported to reports for process improvement meetings

### 6.3 Edit Workflow from Community Feedback

**Purpose:** Admins or workflow owners can review and apply changes based on community feedback or vote trends on **pain points**.

**Key Capabilities:** A dedicated feedback dashboard lets authorized users sort and filter submissions by most-voted, recently-submitted, or department-specific trends. Suggested edits may include adding a new step, adjusting estimated durations, or reassigning actors. All changes are subject to approval workflows, ensuring only designated roles (e.g., managers, QA leads) can finalize and publish modifications.

**Architecture:**

- Feedback linked to proposed patch (structured diff)
- Integrated with version control (see EPIC 1.3)

**Output:**

- Updated diagram version
- Change log showing feedback → edit → approval → deployment

### EPIC 7 – Access Control & Role-Based View

This EPIC ensures that users only see, edit, and interact with workflow elements relevant to their roles and responsibilities - promoting compliance, simplicity, and data protection.

#### 7.1 Role-Based Diagram Filtering

**Purpose:** Users can view only the workflow steps relevant to their role (e.g., employee, manager, compliance officer).

**Key Capabilities:** The system filters workflow diagrams dynamically based on actor-role mappings and visibility settings (e.g., *public*, *internal*, or *private* steps). Authorized roles such as managers can optionally toggle between full-view and role-filtered view for broader context when needed.

**Architecture:** User role information is retrieved from integrated identity management systems (SSO/IAM), and the rendering engine conditionally hides irrelevant lanes or nodes at runtime.

**Output:** A simplified, role-specific diagram improves focus, reduces visual clutter, and enhances usability across large, multi-actor processes.

## 7.2 Edit Permissions by Role

**Purpose:** Only authorized roles (e.g., Process Owners, QA Leads) are allowed to edit workflows.

**Key Capabilities:**

- Role levels include:
  - **Viewer** (read-only access)
  - **Editor** (can suggest or apply changes)
  - **Approver** (can publish and lock versions)
- UI dynamically enables/disables edit options based on permissions

**Architecture:** An **RBAC (Role-Based Access Control)** framework governs all permissions, with every user action logged via audit hooks to ensure traceability.

**Output:** A tightly controlled editing environment enhances security, accountability, and version integrity.

## 7.3 Audit Trail

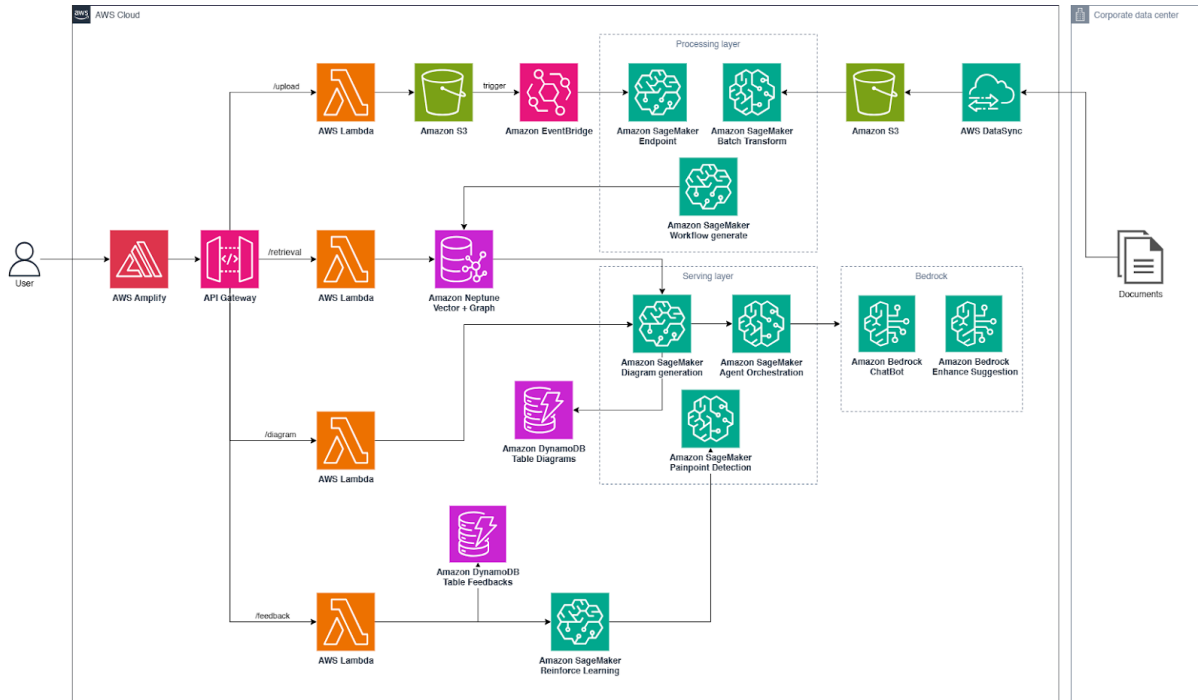
**Purpose:** Maintain a complete and searchable history of all workflow modifications — detailing what was changed, when, by whom, and why.

**Key Capabilities:** Every action — including create, update, delete, or comment — is captured with full metadata. Users can optionally provide reasons for structural edits. The audit log supports filters by user, date, or action type to streamline investigations or reviews.

**Architecture:** The system writes to an append-only changelog tied to version identifiers, with export options for compliance or integration with external audit/reporting systems.

**Output:** An audit dashboard provides administrators and compliance officers with a transparent, filterable activity feed, enabling detailed change tracking and governance assurance.

## Architecture of Solution



**Image 2:** The architecture of our solution

Our solution utilizes a robust combination of AWS services across **processing**, **serving**, and **feedback learning layers** to provide a seamless, scalable, and intelligent document processing pipeline. This architecture is designed to integrate document upload, AI-powered processing, vector-based retrieval, diagram generation, and reinforcement learning for feedback improvement. Here's how we effectively use and combine AWS services to achieve this:

/diagram, and /feedback) that serve as entry points for various user actions. These endpoints orchestrate service invocation using **AWS Lambda** functions.

### 1. Frontend Interface and API Integration

**AWS Amplify** hosts the frontend application, providing users with a secure and responsive interface to interact with the system. Amplify integrates directly with **API Gateway**, allowing RESTful communication to backend services.

**Amazon API Gateway** exposes multiple endpoints (/upload, /retrieval, ensuring a scalable and serverless backend.

## 2. Document Upload and Processing (Processing Layer)

When a user uploads a document (/upload), an **AWS Lambda** function handles the request and stores the file in **Amazon S3**.

An **Amazon EventBridge** rule monitors this S3 bucket and triggers the processing pipeline.

The uploaded document is first passed through:

- **Amazon SageMaker Endpoint** for real-time predictions or quick metadata extraction.
- **Amazon SageMaker Batch Transform** for large-scale or batch processing of document content.

The output is stored back into **Amazon S3**, from where **AWS DataSync** transfers it to an on-premises corporate data center if needed.

This tightly coupled flow ensures automation, scalability, and near real-time availability of processed documents.

## 3. Semantic Retrieval and Vector Search

- Upon a retrieval request (/retrieval), the corresponding Lambda function queries **Amazon Neptune**, which is used to manage both **graph-based relationships** and **vector similarity search** between documents and concepts.
- This integration allows for more **intelligent retrieval** based on context, similarity, and entity relationships-an essential part of our knowledge graph implementation.

## 4. Intelligent Diagram Generation and AI Orchestration (Serving Layer)

For requests to generate diagrams (/diagram), a Lambda function fetches diagram data from **Amazon DynamoDB**, which stores metadata and previously generated diagrams.

**Amazon SageMaker Diagram Generation** processes this data to create visual representations, powered by trained ML models.

These diagram insights are fed into **Amazon SageMaker Agent Orchestration**, which coordinates multiple SageMaker tasks including:

- **Painpoint Detection** (e.g., identifying potential data issues or missing insights),
- Interfacing with **Amazon Bedrock** components such as:
  - **Bedrock lightRAG-ChatBot** – Provides conversational explanations with Retrieval Augmented Generation for accurate result.



- **Bedrock Enhance Suggestion** – Offers improvement suggestions based on document structure and content.

This seamless combination of vector search, generation models, and LLMs ensures rich and context-aware outputs for end users.

## 5. Feedback Loop and Reinforcement Learning

Users (e.g., Banker, Diagram expert) can submit feedback on suggestions or diagram quality via the /feedback endpoint.

This data is written to **Amazon DynamoDB (Table Feedbacks)** and then processed by **Amazon SageMaker Reinforcement Learning**, which retrains or fine-tunes models to improve future responses. This loop makes the system **self-improving**, adapting to user preferences and enhancing model performance over time.

## 6. End-to-End Seamless Integration

Our architecture is carefully crafted to ensure **tight integration across components**:

- **Event-driven orchestration** (EventBridge + Lambda) ensures real-time responsiveness.
- **Data-centric services** (S3, DynamoDB, Neptune) serve as persistent and efficient stores.
- **AI/ML integration** (SageMaker + Bedrock) brings intelligence to processing, generation, and feedback.
- **Cloud-to-on-prem** connectivity via **DataSync** supports hybrid workflows with corporate data centers.

Each AWS service plays a specific role, but they're connected in a way that enables **automatic transitions, minimal latency, and AI-augmented intelligence** across the entire document lifecycle.

## Conclusion

This solution is not just a set of AWS services—it's a well-orchestrated, intelligent system where each component enhances the Key Capabilities of the other. By using serverless functions, managed AI services, vector databases, and feedback loops, we provide a responsive, scalable, and continuously improving platform for enterprise-grade document analysis and enhancement.