



superior politecnica de Chimborazo  
**Facultad de informática y electronica**  
**Escuela de software**



**Nombre:**

Jeferson Vargas

**Codigo:**

6473

**Asignatura:**

Aplicaciones informáticas II

**Docente:**

Dr. Julio Santillan

**Tema:**

**Conceptos de mantenimiento**

**Periodo Académico:**

Marzo - Julio 2025

## 1. Historias de usuario

- Okay, here's everything compiled into a single document, incorporating the new user and technical stories, and the product backlog.

# Historias de Usuario

## Chatbot de Asistencia ESPOH

### HU\_01 Asistencia con "Polito"

- **Descripción:** Como **estudiante o persona interesada**, quiero chatear con un **asistente virtual** llamado "Polito" para obtener respuestas a mis preguntas sobre **documentos de la ESPOL**.
- **Criterios de Validación:**
  - Al iniciar una conversación, "Polito" saluda y pregunta en qué puede ayudar.
  - "Polito" responde a las preguntas utilizando solo la información de los documentos de la base de conocimientos.
  - Si la información no se encuentra en los documentos, "Polito" indica claramente que no tiene información sobre eso.
- **Valor:** Permite a los usuarios obtener información relevante y verificada de manera rápida y eficiente sin la necesidad de buscar manualmente en una gran cantidad de documentos.
- **Prioridad:** Alta
- **Estimación:** 2 días

### HU\_02 Acceso a Documentos Fuente

- **Descripción:** Como **estudiante o persona interesada**, si pregunto por el "link", "enlace" o "descarga" de un documento, quiero que Polito me proporcione el **enlace directo de Google Drive** al archivo si está disponible en la base de datos.
- **Criterios de Validación:**
  - Cuando se solicita un enlace y la herramienta 'Espoch' recupera un documento con `file_id`, "Polito" genera y proporciona el enlace en el formato `https://drive.google.com/uc?export=download&id=[ID_DEL_ARCHIVO_DE_GOOGLE_DRIVE]`.
  - "Polito" menciona brevemente el nombre del documento cuando proporciona el enlace.
  - Si no hay un `file_id` o información relevante para la solicitud de enlace, "Polito" indica que no puede proporcionar un enlace de descarga.
- **Valor:** Facilita el acceso directo a los documentos originales, aumentando la transparencia y la confianza en la información proporcionada.
- **Prioridad:** Alta
- **Estimación:** 1 día

### HU\_04 Claridad en Respuestas

- **Descripción:** Como **estudiante**, quiero que las respuestas de Polito sean **claras y concisas**, sin divagaciones ni información irrelevante.
- **Criterios de Validación:**
  - Las respuestas de Polito son directas al punto y abordan la pregunta del usuario.
  - No hay contenido que no esté directamente relacionado con la consulta o con la información de los documentos.
- **Valor:** Mejora la eficiencia en la obtención de información y reduce la frustración del usuario.
- **Prioridad:** Media
- **Estimación:** 0.5 días

## **HU\_05 Interacción Profesional**

- **Descripción:** Como **usuario**, espero que Polito mantenga un **lenguaje formal y profesional** en todas sus interacciones.
- **Criterios de Validación:**
  - Polito evita el uso de jerga informal, abreviaciones excesivas o un tono coloquial.
  - El tono general de las respuestas es respetuoso y apropiado para un entorno académico.
- **Valor:** Mantiene la credibilidad y la imagen institucional de la ESPOL.
- **Prioridad:** Media
- **Estimación:** 0.5 días

## **Sistema de Gestión Documental**

### **HU\_03 Actualización Automática de Base de Conocimientos**

- **Descripción:** Como **administrador del sistema**, quiero que la **base de conocimientos** se actualice automáticamente cuando se **crea o modifica un documento** en la carpeta de Google Drive designada.
- **Criterios de Validación:**
  - Cuando se detecta un archivo nuevo o modificado en la carpeta monitoreada de Google Drive, el sistema inicia automáticamente el proceso de procesamiento.
  - Los documentos procesados se reflejan en la base de datos vectorial del chatbot.
- **Valor:** Asegura que el chatbot siempre tenga la información más reciente y relevante, reduciendo la necesidad de intervención manual y manteniendo la calidad de las respuestas.
- **Prioridad:** Crítica
- **Estimación:** 3 días

### **HU\_06 Gestión de Errores en Carga de Documentos**

- **Descripción:** Como **administrador del sistema**, quiero ser notificado si un documento no se puede procesar o cargar correctamente en la base de datos vectorial.
- **Criterios de Validación:**

- El sistema registra los errores de procesamiento de documentos (ej., formato no soportado, error de descarga, etc.).
  - Se genera una alerta o notificación al administrador cuando ocurre un error de carga.
- **Valor:** Permite al administrador identificar y resolver problemas rápidamente para mantener la integridad de la base de conocimientos.
- **Prioridad:** Alta
- **Estimación:** 2 días

#### HU\_07 Auditoría de Carga de Documentos

- **Descripción:** Como **administrador del sistema**, quiero tener un registro de los documentos que han sido procesados, incluyendo quién los creó o modificó y cuándo.
- **Criterios de Validación:**
  - La base de datos vectorial almacena metadatos como `creator`, `created_at` y `last_modified` para cada fragmento de documento.
  - El administrador puede consultar esta información para auditar la carga y actualización de documentos.
- **Valor:** Proporciona trazabilidad y responsabilidad sobre el contenido de la base de conocimientos.
- **Prioridad:** Media
- **Estimación:** 1 día

## 2. Historias técnicas

#### HT\_01 Configuración del Chat Trigger

- **Descripción:** Configurar el nodo **When chat message received** para que actúe como el punto de entrada público del chatbot "Polito" e inicie la conversación con el mensaje inicial.
- **Criterios de Validación:**
  - El chatbot es accesible públicamente.
  - El mensaje inicial "Hola soy Polito. ¿En qué puedo ayudarte hoy?" se muestra al usuario al comenzar la interacción.
- **Valor:** Habilita la interfaz de usuario para el chatbot.
- **Prioridad:** Alta
- **Estimación:** 0.5 días

#### HT\_02 Integración de Memoria de Chat

- **Descripción:** Configurar el nodo **Postgres Chat Memory** para persistir el historial de conversaciones del chatbot utilizando la base de datos PostgreSQL (Agente\_RAG).

- **Criterios de Validación:**
  - Las conversaciones anteriores se recuperan correctamente en interacciones subsiguientes.
  - La base de datos PostgreSQL (Agente\_RAG) almacena los mensajes de manera adecuada.
- **Valor:** Permite al chatbot mantener el contexto de la conversación, mejorando la experiencia del usuario.
- **Prioridad:** Alta
- **Estimación:** 1 día

#### **HT\_03 Conexión con Base de Datos Vectorial (Supabase)**

- **Descripción:** Configurar el nodo Supabase Vector Store para que el agente de IA pueda buscar y recuperar información relevante de la tabla documents en Supabase utilizando la similitud semántica.
- **Criterios de Validación:**
  - El nodo Supabase Vector Store se conecta exitosamente a la base de datos Supabase (Agente\_RAG).
  - La herramienta se denomina "Epoch" y su descripción es "Útil para buscar y recuperar fragmentos de documentos específicos y relevantes de la base de conocimiento de la ESPOL, basándose en la similitud semántica con la consulta del usuario. Proporciona el contenido de los documentos más pertinentes para responder preguntas."
- **Valor:** Habilita la capacidad del chatbot para acceder a la base de conocimientos.
- **Prioridad:** Alta
- **Estimación:** 1 día

#### **HT\_04 Configuración de Embeddings para Recuperación**

- **Descripción:** Configurar el nodo Embeddings Ollama para generar los embeddings de las consultas de los usuarios, permitiendo la búsqueda por similitud en la base de datos vectorial de Supabase utilizando el modelo nomic-embed-text:latest.
- **Criterios de Validación:**
  - El nodo Embeddings Ollama se conecta correctamente al servicio Ollama local.

- Los embeddings se generan para las consultas de entrada.
- **Valor:** Esencial para la funcionalidad RAG (Retrieval Augmented Generation) del chatbot.
- **Prioridad:** Alta
- **Estimación:** 0.5 días

#### **HT\_05 Implementación del Agente de IA**

- **Descripción:** Configurar el nodo AI Agent con el modelo de lenguaje (deepseek/deepseek-chat-v3-0324:free) y definir el systemMessage para guiar su comportamiento.
- **Criterios de Validación:**
  - El agente utiliza el modelo de lenguaje especificado.
  - El agente solo responde con información de la herramienta "Epoch".
  - El agente indica "Lo siento, no tengo información sobre eso en los documentos que puedo consultar." si la información no está disponible.
  - El agente genera enlaces de descarga para file\_id cuando se solicitan.
- **Valor:** Define la lógica central del chatbot y asegura respuestas precisas y controladas.
- **Prioridad:** Crítica
- **Estimación:** 2 días

#### **HT\_20 Optimización del Prompt del Agente**

- **Descripción:** Refinar el systemMessage del AI Agent para asegurar que las respuestas sean concisas y profesionales, y que siempre se adhieran a la información de la base de datos, evitando divagaciones.
- **Criterios de Validación:**
  - Las revisiones del systemMessage resultan en respuestas más directas y libres de contenido superfluo.
  - Se realizan pruebas de QA para verificar que el agente no "alucina" o utiliza conocimiento externo.
- **Valor:** Mejora la calidad y la utilidad de las interacciones del chatbot.
- **Prioridad:** Alta
- **Estimación:** 1 día

## **Sistema de Gestión Documental (Carga y Actualización de Documentos)**

### **HT\_06 Configuración de Triggers de Google Drive**

- **Descripción:** Configurar los nodos Archivo Creado y Archivo Actualizado (googleDriveTrigger) para monitorear la carpeta de Google Drive (1AmnBHTwTXR1FFykXNZKyyZ0j4Abx-lj0) y activar el flujo cuando se creen o actualicen documentos.
- **Criterios de Validación:**
  - Los eventos de creación y actualización de archivos en la carpeta designada de Google Drive disparan el flujo.
  - Se recupera la información básica del archivo (ID, nombre, tipo MIME).
- **Valor:** Punto de entrada para la ingesta y actualización automática de documentos.
- **Prioridad:** Crítica
- **Estimación:** 1 día

### **HT\_07 Extracción de ID y Tipo de Archivo**

- **Descripción:** Implementar el nodo Obtener ID (y Obtener ID1 para actualizaciones) para extraer el file\_id y mimeType de los archivos detectados por los activadores de Google Drive.
- **Criterios de Validación:**
  - Los campos file\_id y file\_type se extraen correctamente de los datos de entrada.
- **Valor:** Proporciona los identificadores clave para el procesamiento posterior del documento.
- **Prioridad:** Alta
- **Estimación:** 0.5 días

### **HT\_08 Manejo de Tipos de Archivo (Switch)**

- **Descripción:** Configurar el nodo Switch1 para enrutar el flujo basado en el file\_type del documento, dirigiéndolo a los extractores de texto apropiados (PDF, texto, Excel, Word).
- **Criterios de Validación:**
  - El flujo se desvía correctamente al extractor de PDF si el file\_type es application/pdf.

- El flujo se desvía correctamente al extractor de texto si el file\_type es application/vnd.google-apps.document.
- El flujo se desvía correctamente al extractor de Excel si el file\_type es application/vnd.openxmlformats-officedocument.spreadsheetml.sheet.
- El flujo se desvía correctamente al conversor de Google Doc si el file\_type es application/vnd.openxmlformats-officedocument.wordprocessingml.document, application/msword o application/vnd.ms-word.
- **Valor:** Permite el procesamiento diferenciado de distintos formatos de documento.
- **Prioridad:** Alta
- **Estimación:** 1 día

#### **HT\_09 Descarga y Conversión de Archivos**

- **Descripción:** Utilizar el nodo Descargar Archivo (y Descargar Archivo1) para obtener el contenido del documento de Google Drive, con la capacidad de convertir documentos de Google a texto plano. Para documentos de Windows, usar el nodo Convertir a Google Doc para su conversión antes de la extracción.
- **Criterios de Validación:**
  - Los archivos se descargan correctamente de Google Drive.
  - Los documentos de Google (Docs, Sheets, Slides) se convierten a texto plano.
  - Los archivos de Word (DOCX, DOC) se convierten a formato de Google Doc antes de la descarga si es necesario.
- **Valor:** Transforma los documentos a un formato apto para la extracción de texto.
- **Prioridad:** Alta
- **Estimación:** 1.5 días

#### **HT\_10 Extracción de Texto de Documentos**

- **Descripción:** Implementar los nodos Extraer Texto PDF, Extraer Archivo Texto, y Extraer Excel para procesar el contenido de los diferentes tipos de archivos y extraer el texto relevante.
- **Criterios de Validación:**



- El texto se extrae exitosamente de archivos PDF.
- El texto se extrae exitosamente de archivos de texto plano y documentos de Google Doc.
- Los datos se extraen correctamente de archivos Excel.
- **Valor:** Obtiene el contenido textual de los documentos para su posterior indexación.
- **Prioridad:** Alta
- **Estimación:** 1.5 días

#### **HT\_11 Agregación y Resumen de Contenido**

- **Descripción:** Utilizar los nodos Aggregate (y Aggregate1) y Summarize (y Summarize1) para consolidar el contenido extraído de los documentos.
- **Criterios de Validación:**
  - Todo el contenido extraído de un documento se concatena en un único campo.
- **Valor:** Prepara el texto para el proceso de "chunking", asegurando que todo el contenido se considere antes de dividirlo.
- **Prioridad:** Media
- **Estimación:** 0.5 días

#### **HT\_12 División de Texto en Chunks**

- **Descripción:** Configurar los nodos Recursive Character Text Splitter (y Recursive Character Text Splitter1) para dividir el texto largo de los documentos en "chunks" más pequeños (chunkSize: 2000, chunkOverlap: 200).
- **Criterios de Validación:**
  - El texto se divide en fragmentos de tamaño máximo 2000 caracteres con un solapamiento de 200 caracteres.
- **Valor:** Optimiza el tamaño de los fragmentos para la ingesta en la base de datos vectorial y mejora la calidad de la recuperación.
- **Prioridad:** Alta
- **Estimación:** 1 día

#### **HT\_13 Asignación de Metadatos a Chunks**

- **Descripción:** Configurar los nodos Enhanced Default Data Loader (y Enhanced Default Data Loader1) para adjuntar metadatos relevantes

(file\_id, version, creator, created\_at, last\_modified, folder\_path, file\_name, file\_extension) a cada chunk de texto.

- **Criterios de Validación:**
  - Cada chunk de texto incluye los metadatos especificados, obtenidos de la información del archivo original y del proceso de actualización.
- **Valor:** Enriquece los datos en la base de datos vectorial, permitiendo una búsqueda y filtrado más precisos.
- **Prioridad:** Alta
- **Estimación:** 1 día

#### **HT\_14 Generación de Embeddings para Ingesta**

- **Descripción:** Configurar los nodos Embeddings Ollama (y Embeddings Ollama3) para generar los embeddings vectoriales de cada chunk de texto utilizando el modelo nomic-embed-text:latest.
- **Criterios de Validación:**
  - Los embeddings se crean correctamente para cada chunk de texto.
  - El modelo nomic-embed-text:latest se utiliza para la generación de embeddings.
- **Valor:** Convierte el texto en un formato numérico que puede ser buscado por similitud en la base de datos vectorial.
- **Prioridad:** Alta
- **Estimación:** 1 día

#### **HT\_15 Inserción en Base de Datos Vectorial**

- **Descripción:** Configurar el nodo Insertar en BD Vectorizada Supabase (y Insertar en BD Vectorizada Supabase1) para almacenar los chunks de texto y sus embeddings junto con los metadatos en la tabla documents de Supabase.
- **Criterios de Validación:**
  - Los chunks de texto, embeddings y metadatos se insertan exitosamente en la tabla documents de Supabase.
- **Valor:** Almacena los datos de la base de conocimientos para que el chatbot pueda acceder a ellos.
- **Prioridad:** Crítica
- **Estimación:** 1.5 días

### **HT\_16 Gestión de Actualizaciones (Eliminación Previa)**

- **Descripción:** Configurar el nodo **Borrar Filas Doc Antiguo** para eliminar los registros existentes en la base de datos vectorial asociados a un `file_id` antes de insertar una nueva versión del documento.
- **Criterios de Validación:**
  - Las filas con el `file_id` correspondiente se eliminan de la tabla `documents` antes de la nueva inserción.
- **Valor:** Asegura que la base de datos vectorial esté siempre al día y evita duplicados o información obsoleta.
- **Prioridad:** Alta
- **Estimación:** 0.5 días

### **HT\_17 Incremento de Versión del Documento**

- **Descripción:** Implementar el nodo **Incrementar Versión** (o activar y configurar el nodo **Obtener Version con GPT-4o-mini**) para gestionar y actualizar el número de versión de los documentos.
- **Criterios de Validación:**
  - Para cada actualización de documento, se genera un número de versión incrementado (ej. v1 a v2).
  - El número de versión se almacena en los metadatos del chunk.
- **Valor:** Permite rastrear las versiones de los documentos y asegurar que la versión más reciente sea la utilizada.
- **Prioridad:** Media
- **Estimación:** 1 día

### **HT\_18 Eliminación de Archivos Temporales**

- **Descripción:** Configurar los nodos **Eliminar Archivo** (y **Eliminar Archivo1**) para borrar los archivos temporales descargados o convertidos de Google Drive una vez que su contenido ha sido procesado e insertado en la base de datos vectorial.
- **Criterios de Validación:**
  - Los archivos descargados o convertidos temporalmente se eliminan de Google Drive después de su procesamiento.
- **Valor:** Optimiza el almacenamiento y gestiona los recursos de Google Drive.
- **Prioridad:** Media

- **Estimación: 0.5 días**

#### **HT\_19 Procesamiento por Lotes**

- **Descripción:** Utilizar los nodos Loop Over Items (y Loop Over Items1) y Limit para procesar los archivos en lotes y controlar la concurrencia.
- **Criterios de Validación:**
  - Los documentos se procesan en lotes definidos.
  - El flujo maneja múltiples documentos sin saturar los recursos.
- **Valor:** Mejora el rendimiento y la estabilidad del flujo, especialmente con grandes volúmenes de documentos.
- **Prioridad: Media**
- **Estimación: 0.5 días**

#### **HT\_21 Implementación de Manejo de Errores para Carga**

- **Descripción:** Añadir lógica de manejo de errores en los nodos de descarga, extracción y carga a la base de datos vectorial para capturar y registrar fallas en el procesamiento de documentos.
- **Criterios de Validación:**
  - Si un nodo falla (ej. Descargar Archivo, Extraer Texto PDF, Insertar en BD Vectorizada Supabase), el error se registra.
  - El flujo puede manejar excepciones sin detener todo el proceso.
- **Valor:** Aumenta la robustez del sistema y proporciona visibilidad sobre problemas de procesamiento.
- **Prioridad: Alta**
- **Estimación: 2 días**

#### **HT\_22 Notificaciones de Errores de Carga**

- **Descripción:** Configurar un mecanismo de notificación (ej., correo electrónico, Slack) para alertar al administrador del sistema sobre fallos en la carga o procesamiento de documentos.
- **Criterios de Validación:**
  - Una notificación se envía al canal designado cuando se registra un error de carga.
  - La notificación incluye detalles relevantes del error (nombre del archivo, tipo de error).

- **Valor:** Facilita la detección temprana y resolución de problemas de integridad de datos.
  - **Prioridad:** Alta
  - **Estimación:** 1.5 días
- 

### **Product Backlog (25 de Mayo - 1 de Julio)**

**Este Product Backlog está organizado por semanas, asumiendo un enfoque iterativo y flexible. Los puntos son flexibles y pueden ajustarse según el progreso.**

---

#### **Sprint 1: 25 de Mayo - 31 de Mayo**

- **HU\_01 Asistencia con "Polito" (Prioridad: Alta, Estimación: 2 días)**
    - **HT\_01 Configuración del Chat Trigger (Prioridad: Alta, Estimación: 0.5 días)**
    - **HT\_02 Integración de Memoria de Chat (Prioridad: Alta, Estimación: 1 día)**
    - **HT\_03 Conexión con Base de Datos Vectorial (Supabase) (Prioridad: Alta, Estimación: 1 día)**
  - **HU\_03 Actualización Automática de Base de Conocimientos - Parte 1 (Creación) (Prioridad: Crítica, Estimación: 3 días)**
    - **HT\_06 Configuración de Triggers de Google Drive (Creación de Archivo) (Prioridad: Crítica, Estimación: 0.5 días)**
    - **HT\_07 Extracción de ID y Tipo de Archivo (Prioridad: Alta, Estimación: 0.5 días)**
    - **HT\_08 Manejo de Tipos de Archivo (Switch) (Prioridad: Alta, Estimación: 1 día)**
    - **HT\_19 Procesamiento por Lotes (Prioridad: Media, Estimación: 0.5 días)**
- 

#### **Sprint 2: 1 de Junio - 7 de Junio**

- **HU\_03 Actualización Automática de Base de Conocimientos - Parte 2 (Extracción) (Prioridad: Crítica, Estimación: 3 días)**
  - **HT\_09 Descarga y Conversión de Archivos (Prioridad: Alta, Estimación: 1.5 días)**

- HT\_10 Extracción de Texto de Documentos (Prioridad: Alta, Estimación: 1.5 días)
  - HT\_11 Agregación y Resumen de Contenido (Prioridad: Media, Estimación: 0.5 días)
  - HT\_04 Configuración de Embeddings para Recuperación (Prioridad: Alta, Estimación: 0.5 días)
  - HT\_05 Implementación del Agente de IA - Parte 1 (Configuración Básica) (Prioridad: Crítica, Estimación: 1 día)
- 

#### **Sprint 3: 8 de Junio - 14 de Junio**

- HU\_02 Acceso a Documentos Fuente (Prioridad: Alta, Estimación: 1 día)
    - HT\_05 Implementación del Agente de IA - Parte 2 (Lógica de enlaces y no información) (Prioridad: Crítica, Estimación: 1 día)
  - HU\_03 Actualización Automática de Base de Conocimientos - Parte 3 (Indexación) (Prioridad: Crítica, Estimación: 3 días)
    - HT\_12 División de Texto en Chunks (Prioridad: Alta, Estimación: 1 día)
    - HT\_13 Asignación de Metadatos a Chunks (Prioridad: Alta, Estimación: 1 día)
    - HT\_14 Generación de Embeddings para Ingesta (Prioridad: Alta, Estimación: 1 día)
    - HT\_15 Inserción en Base de Datos Vectorial (Prioridad: Crítica, Estimación: 1.5 días)
- 

#### **Sprint 4: 15 de Junio - 21 de Junio**

- HU\_03 Actualización Automática de Base de Conocimientos - Parte 4 (Actualización de archivos) (Prioridad: Crítica, Estimación: 3 días)
  - HT\_06 Configuración de Triggers de Google Drive (Actualización de Archivo) (Prioridad: Crítica, Estimación: 0.5 días)
  - HT\_16 Gestión de Actualizaciones (Eliminación Previa) (Prioridad: Alta, Estimación: 0.5 días)
  - HT\_17 Incremento de Versión del Documento (Prioridad: Media, Estimación: 1 día)

- **HT\_18 Eliminación de Archivos Temporales (Prioridad: Media, Estimación: 0.5 días)**
  - **HU\_04 Claridad en Respuestas (Prioridad: Media, Estimación: 0.5 días)**
    - **HT\_20 Optimización del Prompt del Agente (Prioridad: Alta, Estimación: 1 día)**
- 

#### **Sprint 5: 22 de Junio - 28 de Junio**

- **HU\_05 Interacción Profesional (Prioridad: Media, Estimación: 0.5 días)**
    - **HT\_20 Optimización del Prompt del Agente (Refinamiento) (Prioridad: Alta, Estimación: 0.5 días)**
  - **HU\_06 Gestión de Errores en Carga de Documentos (Prioridad: Alta, Estimación: 2 días)**
    - **HT\_21 Implementación de Manejo de Errores para Carga (Prioridad: Alta, Estimación: 2 días)**
  - **HU\_07 Auditoría de Carga de Documentos (Prioridad: Media, Estimación: 1 día)**
- 

#### **Sprint 6: 29 de Junio - 1 de Julio**

- **HT\_22 Notificaciones de Errores de Carga (Prioridad: Alta, Estimación: 1.5 días)**
- **Refinamiento y Pruebas Generales (Prioridad: Alta, Estimación: ~2 días)**
  - **Pruebas de integración del chatbot con la base de conocimientos.**
  - **Pruebas de estrés del sistema de carga de documentos.**
  - **Corrección de errores menores.**