**A continuación, se especifican los cambios de la propuesta del pipeline versión 1 (V1) con la propuesta del pipeline versión 2 (V2)**

**1. Etapa Data Input**

* **V1:** No explicita las fuentes internas y externas ni incluye elementos de seguridad.
* **V2:** Agrega detalles como *registros médicos*, *historias clínicas*, y iconos de seguridad y política de acceso.
  + **¿Por qué?** Mejora la trazabilidad, origen y cumplimiento en gobernanza de datos.
* **V1:** Menor número de herramientas tecnológicas asociadas.
* **V2:** Añade logos de *PostgreSQL*, *DVC*, *BeautifulSoup*, y hace más explícita la herramienta de versionado.
  + **¿Por qué?** V2 busca reflejar con mayor precisión el stack tecnológico y su aporte en reproducibilidad.
* **V1:** No presenta la fase de versionado de datos para trazabilidad y reproducibilidad.
* **V2:** Incluye una fase explícita de “Versionado de datos para trazabilidad y reproducibilidad” después de la recolección de datos.
  + **¿Por qué?** Este cambio aporta trazabilidad y permite la reproducibilidad del pipeline mediante herramientas como DVC, fundamentales en proyectos con datos sensibles como los clínicos.

**2. Etapa Data Explorer & Feature Selection**

* **V1:** No presenta el tracking de artefactos y transformaciones.
* **V2:** Incluye una fase explícita de “**Tracking de artefactos y transformaciones**”.
  + **¿Por qué?** Aporta trazabilidad a las modificaciones realizadas durante el preprocesamiento.
* **V1:** Menor cantidad de herramientas visualizadas.
* **V2:** Añade *Weights & Biases* como herramienta de seguimiento de transformaciones.
  + **¿Por qué?** Brinda mayor control en experimentos y cambios sobre los datos.

**3. Etapa Modeling & Evaluate – Optimize Models**

* **V1:** Esta etapa aparece más resumida.
* **V2:** Se amplía y divide claramente en dos subetapas: *Optimize Models* y *Modeling & Evaluate*.
  + **¿Por qué?** Facilita la comprensión de actividades como tuning, optimización y evaluación.
* **V2:** Añade herramientas especializadas como *Optuna* para optimización bayesiana.
  + **¿Por qué?** Muestra un enfoque más robusto y modular para la búsqueda de hiperparámetros.

**4. Etapa Compilate**

* **V1:** Muestra empaquetado y registro del modelo.
* **V2:** Añade una caja de **CI/CD con GitHub Actions** detallando pruebas automáticas, comentarios en PR y push
  + **¿Por qué?** Refleja las buenas prácticas de DevOps aplicadas a modelos ML.

**5. Etapa Serve**

* **V1:** Contiene despliegue como API y recolección de nuevos datos.
* **V2:** Añade **Infraestructura como Código (IaC)** con Terraform.
  + **¿Por qué?** Mejora la automatización, escalabilidad y replicabilidad de ambientes.

**6. Etapa Monitoring**

* **V2:** Incluye información más detallada, como **alertas automáticas y monitoreo de latencia**.
  + **¿Por qué?** Resalta el uso de herramientas como *Prometheus* y *Grafana* no solo para métricas sino también para performance del sistema.

**7. Etapa Generate Value**

* **V1:** Incluye esta etapa con LLMs pero sin justificación funcional clara.
* **V2:** Mantiene la estructura pero fue complementado con una descripción en entregables del proyecto, justificando la inclusión de LLM para interpretación.
  + **¿Por qué?** Aborda la crítica de que los LLMs agregan complejidad sin aparente beneficio técnico, dando contexto y funcionalidad interpretativa al componente.

A continuación, se listan las tecnologías que aparecen en el pipeline **V2** y que **no están presentes en el pipeline V1**, junto con una breve explicación del propósito o razón de su inclusión:

* **DVC**  
  Razón: Permite versionar datos y modelos, aportando trazabilidad y reproducibilidad al flujo de trabajo MLOps.
* **PostgreSQL**  
  Razón: Se agrega como una base de datos relacional robusta para almacenamiento estructurado, especialmente útil en contextos clínicos.
* **BeautifulSoup**  
  Razón: Herramienta de scraping usada para recolectar información estructurada o no estructurada desde sitios web externos.
* **Weights & Biases (W&B)**  
  Razón: Plataforma para el tracking de experimentos, visualización de métricas y gestión del ciclo de vida de modelos de forma colaborativa.
* **Tracking de artefactos y transformaciones** (MLflow + W&B)  
  Razón: Registra y permite auditar cambios aplicados a los datos y transformaciones realizadas durante la etapa de preprocesamiento y feature engineering.
* **Optuna**  
  Razón: Framework de optimización bayesiana eficiente para el tuning automático de hiperparámetros con bajo consumo de recursos.
* **GitHub Actions**  
  Razón: Motor de automatización para CI/CD que permite ejecutar pruebas, validar PRs y desplegar imágenes Docker de forma automática.
* **Terraform**  
  Razón: Herramienta de infraestructura como código (IaC) que permite aprovisionar entornos reproducibles, escalables y controlados para despliegue de modelos.

**A continuación, y para esta versión, se listan los supuestos detallados en el documento diseño-pipeline-V2 y no determinados en la versión de la primera entrega**

1. Acceso desde entornos locales con recursos limitados
2. Despliegue modular y portable
3. Disponibilidad de infraestructura cloud (cuando se requiera)
4. Conectividad y seguridad
5. Disponibilidad de herramientas de trazabilidad
6. Acceso a los datos históricos y actualizados
7. Supervisión médica garantizada en ciclo de reentrenamiento
8. Dependencia de la infraestructura TI hospitalaria
9. Privacidad y gobernanza de datos
10. Disponibilidad de soporte técnico especializado bajo demanda