

Informe técnico del portal-dashboard

Arquitectura, pipeline de datos y proceso de actualización de la plataforma de Comisiones Temáticas del Plan Perú 2050. Documento de referencia para el CNPP — Colegio de Ingenieros del Perú.

65 Comisiones	19 Validadas	45 En revisión	180 Indicadores
-------------------------	------------------------	--------------------------	---------------------------

1. Resumen

El portal es una aplicación web **estática** (sin servidor de aplicaciones) que publica la información de las comisiones del Plan Perú 2050 y permite consultarla, filtrarla, simular escenarios y descargar los documentos. Los datos se producen mediante un **pipeline asistido por IA** que transforma las redacciones de cada comisión en datos estructurados verificables. El diseño prioriza **rapidez, bajo costo, trazabilidad de fuentes y facilidad de actualización**.

2. Pipeline de datos

Cada documento de comisión recorre cinco etapas hasta llegar al portal:

Pipeline de datos — del documento al portal



Figura 1. Del documento de la comisión al portal publicado.

- **1. Ingesta:** se reciben los Informes Ejecutivos (.docx, .pdf, .pptx) de cada comisión.
- **2. Conversión:** se extrae el texto con herramientas estándar (pandoc, pdftotext, python-pptx).
- **3. Extracción IA multi-agente:** un agente por comisión estructura el contenido al esquema oficial (I. Situación futura ... VIII. Articulación con Programas Presupuestales), con regla **anti-overclaiming**: solo se capturan cifras explícitas en el documento; si un dato no consta, se marca como «sin dato» (nunca se inventa).
- **4. Validación:** los datos quedan en JSON versionado. Cada comisión se marca «validada» (redacción oficial) o «en revisión» (línea base preliminar, claramente etiquetada).
- **5. Publicación:** se generan automáticamente las fichas, los PDFs (con índice) y se despliega al servidor.

3. Arquitectura y despliegue

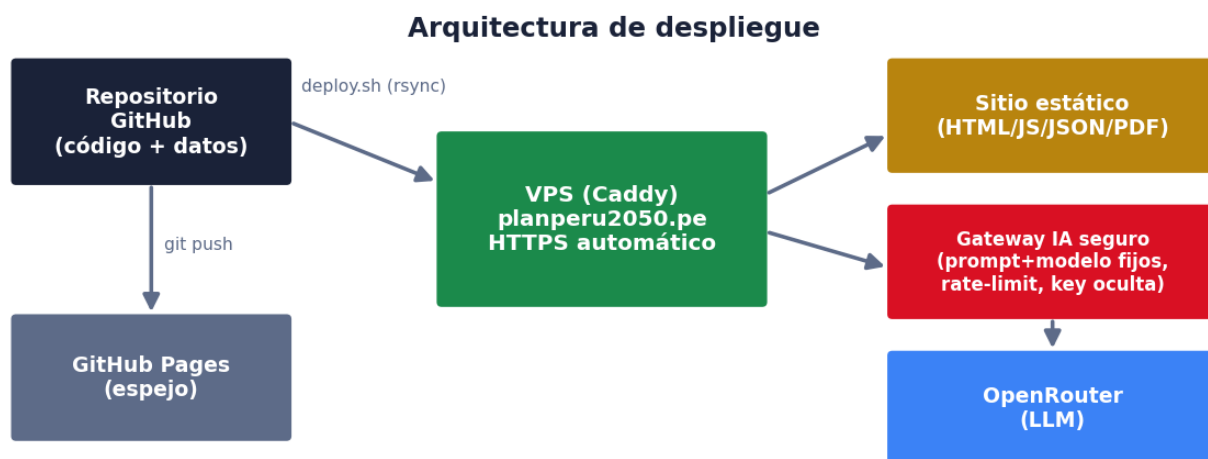


Figura 2. El código y los datos viven en GitHub; el sitio se sirve desde un VPS con Caddy (HTTPS automático). La consulta con IA pasa por un proxy que mantiene la llave del modelo oculta en el servidor.

4. Stack tecnológico

Capa	Tecnología	Rol
Frontend	HTML5 + JavaScript (sin framework)	Interfaz, filtros, fichas, navegación por pestañas
Visualización	Chart.js + Leaflet	Gráficos de avance/brechas y mapa del Perú
Simulación	JS (fórmulas polinómicas)	Escenarios y trayectoria 2026→2050
Datos	JSON versionado en Git	Comisiones, indicadores, puntos del mapa
Extracción	Pipeline IA multi-agente	Documentos → datos estructurados
Documentos	Python (reportlab, python-docx, matplotlib)	PDFs/Word con índice y gráficos
IA / Asistente	OpenRouter vía proxy Caddy	Consulta en lenguaje natural, key oculta
Hosting	VPS + Caddy (HTTPS automático)	planperu2050.pe — servidor del portal
CI / Fuente	GitHub + GitHub Pages (espejo)	Versionado y respaldo
Analítica	Google Analytics (GA4)	Métricas de uso

5. Estado actual de la cobertura

Cobertura: 65 comisiones

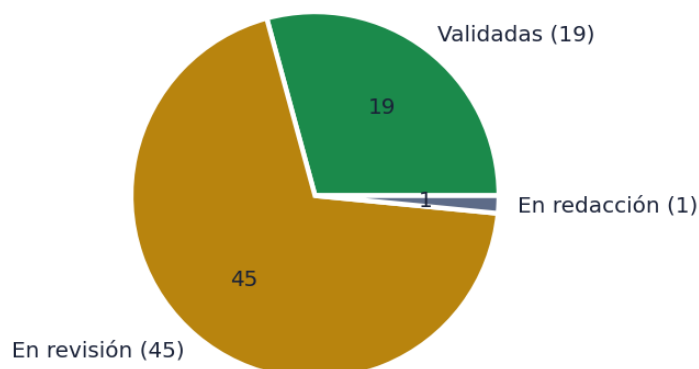


Figura 3. 19 comisiones validadas y 45 en revisión, de 65 comisiones temáticas.

Las comisiones «validadas» provienen de redacciones oficiales; las «en revisión» son una línea base preliminar inferida que el equipo del CIP valida progresivamente. Cada ficha indica su estado de forma visible.

6. Proceso de actualización

Incorporar o actualizar una comisión es un proceso reproducible de pocos pasos:

1. Recibir el Informe Ejecutivo de la comisión y colocarlo en la carpeta de fuentes.
2. Convertir a texto y ejecutar la extracción IA (un agente por documento) al esquema oficial.
3. Revisar el JSON resultante; mover la comisión de «en revisión» a «validada» cuando corresponda.
4. Regenerar automáticamente las fichas y los PDFs (scripts de Python).
5. Publicar: «git push» (versiona + espejo) y «deploy.sh» (actualiza el servidor). El HTTPS se renueva solo.

Tiempo típico por lote de comisiones: minutos. No se requiere intervención manual en el servidor.

7. Funcionalidades del portal

- Directorio de comisiones con buscador y filtros por eje y por estado.
- Ficha por comisión con la estructura del Informe Ejecutivo (I–VIII), incluida la articulación con el Acuerdo Nacional, el PEDN 2050 y los Programas Presupuestales.
- Panorama nacional con gráficos comparativos y mapa territorial del Perú.
- Simulación de escenarios con fórmulas polinómicas (trayectoria 2026→2050).
- Asistente de consulta con IA (respuestas breves, lenguaje natural).
- Descarga de documentos en PDF (fichas, plan de 100 días, síntesis por ejes) con índice.
- Sección de preguntas frecuentes (FAQ) y dosificación de la vista por hitos.

8. Seguridad del asistente de IA

El asistente conversacional fue sometido a un **análisis adversarial** (pruebas de abuso) y posteriormente blindado. Diseño actual:

- **Gateway propio del lado del servidor:** el navegador solo envía la pregunta; el servidor controla el prompt del sistema, el modelo y el contexto. El usuario no puede alterar el rol del asistente.
- **Alcance restringido:** responde únicamente sobre el Plan Perú 2050; cualquier pregunta de otro tema (o intento de cambiar sus reglas) recibe un mensaje fijo de rechazo.

- **Modelo y respuesta acotados:** el modelo y el límite de tokens los fija el servidor (no el cliente), evitando el uso de modelos costosos con la cuenta del proyecto.
- **Límite de frecuencia (rate-limit) por IP:** impide el abuso por volumen de consultas.
- **Llave de IA protegida:** la credencial del modelo reside solo en el servidor (variable de entorno del servicio), nunca en el navegador ni en la configuración pública del proxy.

Resultado de la verificación posterior: los intentos de abuso (uso fuera de tema, modelos arbitrarios, inyección de instrucciones y consultas masivas) quedan bloqueados, manteniéndose el servicio funcional para consultas legítimas.

9. Principios de diseño

- **Anti-overclaiming:** no se inventan cifras; las fuentes son trazables y el estado de validación es visible.
- **Reproducibilidad:** todo el proceso (extracción, documentos, despliegue) está automatizado y versionado.
- **Bajo costo y robustez:** sitio estático + un servidor; sin base de datos compleja ni dependencias frágiles.
- **Privacidad:** los datos personales de los integrantes de las comisiones no se publican.
- **Seguridad por diseño:** los controles del asistente viven en el servidor, no en el cliente.

Desarrollo, analítica de datos e ingeniería del portal: **Carlos Cárdenas Fernández**. Portal en producción: planperu2050.pe