PROGRAMA

PERSISTE

Programa Escolar de Reutilización, Sostenibilidad e Innovación Tecnológica Educativa

"Lo viejo funciona"

Guía técnica: IA local, Ollama y AnythingLLM

Instalación, uso y verificación del asistente local

Colección PERSISTE · Documento 3 de 3

Muestra para evaluación · Versión 0.1 · Noviembre de 2026

Formación docente en investigación-acción · Aprendizaje por proyectos · Economía circular aplicada a la gestión tecnológica escolar

Muestra para evaluación

Este documento ilustra el formato, el tono y el nivel de detalle que tendría el material definitivo. Es una redacción genérica y no reemplaza la documentación de implementación, que se elabora con cada escuela en la etapa de diagnóstico. Los nombres de menús y las versiones pueden variar: antes de implementar, los pasos se verifican contra la documentación vigente de cada herramienta.

1. Qué es y para qué sirve

Es un asistente que funciona en una computadora de la escuela y permite consultar en lenguaje natural la documentación propia, con indicación de la fuente. Se necesita internet solo para instalar los programas y descargar el modelo; después funciona sin conexión.

2. Componentes

Componente	Función	Licencia
Ollama	Motor que ejecuta el modelo de lenguaje en la computadora	Código abierto
Modelo qwen2.5:3b (unos 2 GB)	Genera las respuestas	Licencia Qwen, de investigación y uso no comercial; no es software libre
AnythingLLM Desktop	Interfaz de consulta con recuperación de información sobre documentos propios	Código abierto
Archivo Técnico Institucional	Fichas y procedimientos de la escuela	La que defina la escuela

Sobre la licencia del modelo

El modelo usado en las pruebas del programa no cumple el criterio de software libre. Está prevista una nueva evaluación de candidatos con licencias libres (Apache 2.0 o MIT), para que la recomendación sea coherente con la prioridad que el programa otorga al software libre.

3. Requisitos

- Computadora reacondicionada con 8 GB de RAM, procesador de al menos 4 núcleos y SSD de 128 GB.
- Sistema operativo Ubuntu MATE o Windows.
- Conexión a internet solo durante la instalación y la descarga del modelo.

4. Pasos generales

1. Instalar Ollama con el instalador de su sistema operativo.
2. Descargar el modelo desde Ollama.
3. Instalar AnythingLLM Desktop.
4. En AnythingLLM, elegir Ollama como proveedor del modelo de lenguaje y seleccionar el modelo descargado.
5. Crear un espacio de trabajo, por ejemplo “Archivo Técnico”.
6. Cargar los documentos y pedir que se procesen.

7. Probar con una consulta cuya respuesta se conozca y comprobar que cita la fuente.

5. El Prompt Oficial PERSISTE y buenas prácticas

El prompt le pide al asistente que redacte una ficha técnica abordando los seis criterios de adquisición tecnológica responsable. Buenas prácticas:

- Incluir el contexto del programa en cada consulta.
- Formular una consulta por vez.
- Pedir siempre la fuente.
- No aceptar ninguna respuesta sin verificarla.

6. Errores frecuentes

Error	Ejemplo observado	Qué hacer
Precio inventado o erróneo	Un precio de $100 para una impresora de unos $400	Cotejar con una fuente oficial o un proveedor
Nombre de paquete o repuesto inexistente	Visto en otros modelos de prueba	Buscarlo en el repositorio o catálogo oficial
Dirección web inventada	Visto en otros modelos de prueba	Abrirla antes de citarla
Campos sin completar	Visto en otros modelos de prueba	Completar a mano o reformular la consulta

7. Mantenimiento

- Actualizar las herramientas y revisar los modelos disponibles al inicio de cada ciclo lectivo.
- Repetir el benchmark con el prompt estándar.
- Respaldar el Archivo Técnico y el repositorio.
- Registrar tiempos de respuesta y errores para retroalimentar al programa.

8. Escalamiento

Cuando se cumplan al menos dos de las siguientes condiciones, se pasa al Nivel 2:

- Los tiempos de respuesta superan consistentemente los 3 minutos.
- Hay 20 o más consultas semanales.
- Se necesita consulta simultánea desde varios equipos.
- La documentación inicial está completa y el foco pasa a la consulta.

El Nivel 2 implica 16 GB de RAM y el modo multiusuario del mismo software.

Elaborado por: Equipo técnico del Programa PERSISTE

Fecha: Noviembre de 2026

Versión: 0.1

Estado: Muestra para evaluación; no constituye documentación de implementación