

4. Corrección asistida por IA sin vulnerar datos personales

No basta con conocer tipos de instrumentos; también hay que saber qué hace que un instrumento sea bueno. En términos teóricos, un instrumento de evaluación debe ser válido, fiable, claro, pertinente y manejable. La validez se refiere a si realmente evalúa lo que pretende evaluar. La fiabilidad se relaciona con la consistencia de los resultados. La claridad afecta a la comprensión del alumnado y del profesorado. La pertinencia indica si el instrumento encaja con la tarea. La manejabilidad tiene que ver con el tiempo real de uso en el aula.

La IA puede mejorar el lenguaje y proponer versiones más limpias, pero no garantiza por sí sola estos criterios. Por eso es recomendable revisar siempre si el instrumento evita ambigüedades, si no sobrecarga al docente y si permite recoger evidencias suficientes para tomar decisiones fundamentadas.

- [4.1. Corrección automática vs. corrección asistida: diferencias clave](#)
- [4.2. Qué información puede introducirse en una herramienta de IA](#)
- [4.3. Estrategias para trabajar con textos y evidencias sin datos identificativos](#)
- [4.4. Uso de rúbricas, listas de cotejo y criterios de evaluación como base para la corrección con IA](#)
- [4.5. Supervisión docente y validación de resultados](#)

4.1. Corrección automática vs. corrección asistida: diferencias clave

La corrección automática suele implicar que un sistema da una puntuación o juicio con mínima intervención docente. Puede ser útil en ejercicios cerrados, test o comprobaciones muy objetivas, pero resulta limitada para tareas complejas.

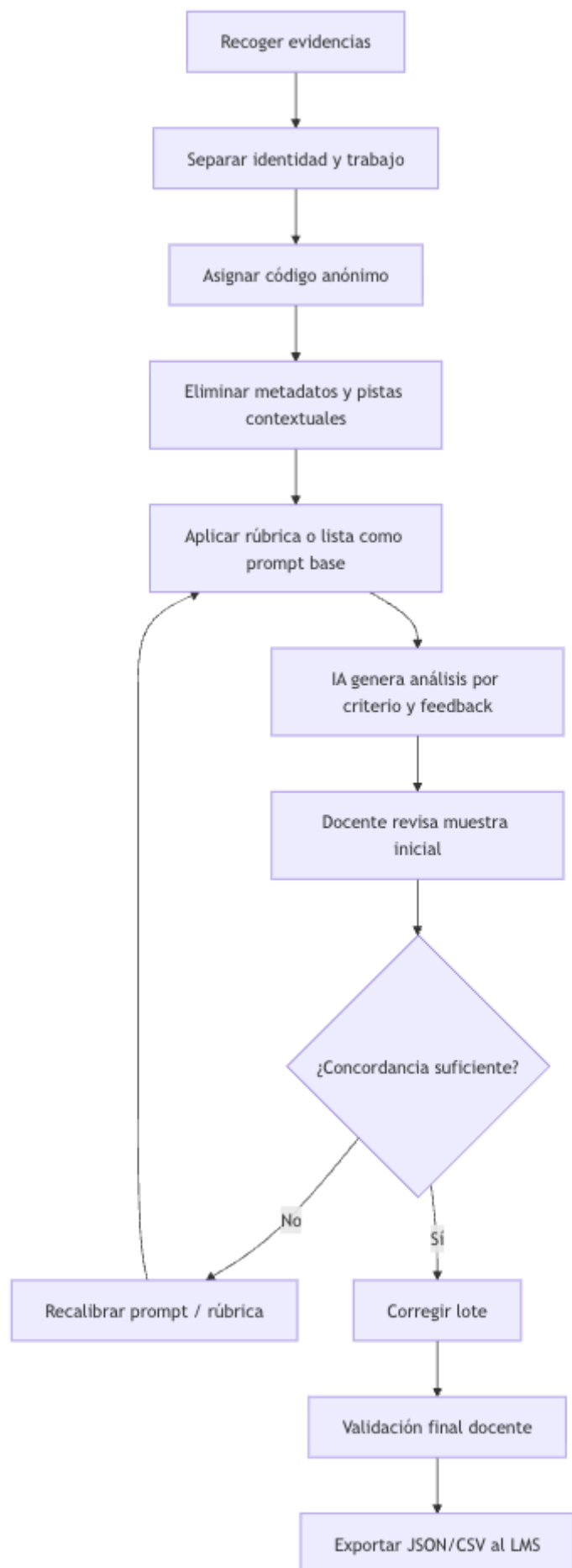
La corrección asistida, en cambio, usa la IA como herramienta de apoyo al análisis (propone borradores, identifica patrones) y a la redacción de observaciones (resume errores, sugiere comentarios), pero el docente conserva el control final revisando y validando todo antes de emitir una calificación. En educación, esta segunda opción es mucho más segura y pedagógicamente más sólida.

Esta diferencia es importante porque una nota automática puede dar la impresión de neutralidad, aunque en realidad esté basada en reglas poco transparentes o sesgadas. En evaluación educativa, esto es especialmente delicado porque una decisión errónea puede afectar al aprendizaje, a la autoestima y a la trayectoria escolar del alumnado.

Ejemplo comparativo:

- Corrección automática: la IA califica una redacción con un número sin justificación pedagógica clara.
- Corrección asistida: la IA identifica problemas de estructura y propone comentarios; el docente revisa, ajusta y decide la calificación.

La segunda opción es más defendible desde el punto de vista pedagógico y legal.





Aránzazu Cortés + Chat GPT

La recomendación operativa es empezar cada tarea con una **muestra de calibración** y no pasar al lote completo hasta comprobar que la salida es coherente con el criterio docente. Esa lógica coincide con la literatura reciente sobre sistemas híbridos y con la prevención del sesgo de automatización.

Por tanto, la corrección asistida por IA puede ser una aliada muy potente si se usa con prudencia. Permite agilizar tareas repetitivas, detectar patrones de error, redactar feedback y organizar evidencias. Pero en evaluación no todo puede automatizarse, porque la corrección no es solo una operación técnica: implica interpretación pedagógica, sensibilidad contextual y responsabilidad profesional.

Además, cuando se trabaja con datos de alumnado, entran en juego derechos especialmente sensibles. La protección de datos en el ámbito educativo es una cuestión central, y la AEPD insiste en la necesidad de extremar cautelas con menores, plataformas digitales y análisis de información escolar. Por eso, la corrección asistida debe diseñarse desde la minimización de datos y la supervisión docente.

La evaluación de aprendizajes puede afectar al itinerario educativo y, por eso, el marco europeo considera especialmente sensibles los sistemas usados para evaluar resultados de aprendizaje o monitorizar alumnado. Por tanto, la IA debe servir para ahorrar tiempo en tareas mecánicas, no para delegar decisiones sustantivas.

4.2. Qué información puede introducirse en una herramienta de IA

Uno de los puntos clave es decidir qué información puede compartirse con una herramienta de IA. La regla básica es la minimización: introducir solo lo necesario y, si es posible, sin datos identificativos directos. En contexto escolar, esto es especialmente importante cuando hablamos de menores, trabajos académicos, observaciones de conducta o informes de evaluación.

Por tanto, en una herramienta de IA solo debería introducirse información necesaria, mínima y, cuando sea posible, anonimizada o pseudonimizada. No conviene subir nombres, apellidos, fotografías, números de expediente, direcciones, ni documentos con identificación directa salvo que exista base legal, garantías técnicas y autorización institucional clara.

Información recomendable para trabajar con IA:

- Textos o fragmentos anónimos o seudonimizados, es decir, sin nombre del autor.
- Respuestas codificadas o separadas por códigos.
- Fragmentos aislados de producciones o producciones anonimizadas estrictamente necesarios.
- Datos o tablas de resultados agregados de grupo.
- Descriptores de desempeño (criterios, rúbricas, listas y escalas) sin nombre de alumnado, es decir, listas con identificadores neutrales. Por ejemplo, "Alumno 03".

Información que conviene evitar:

- Datos identificativos directos: Nombre y apellidos, foto o voz identificable.
- Identificadores oficiales: correo o número de expediente.
- Informes completos con datos personales.
- Informes completos con información sensible.
- Producciones con metadatos personales no depurados.
- Imágenes con rostros reconocibles.
- Capturas de pantalla con nombres visibles o contexto fácilmente reidentificable.
- Información sensible sin base legal clara (datos de menores sin necesidad pedagógica clara).



La lógica es sencilla: si la IA no necesita el dato para cumplir su función, no debe recibirlo. Esto reduce riesgos de reidentificación, filtración y uso indebido de información.

Checklist de privacidad y cumplimiento

Comprobación	Pregunta práctica	Base
Finalidad	¿La IA ayuda a analizar o solo "porque sí"?	RGPD: limitación de finalidad y minimización
Minimización	¿He eliminado todo dato no necesario?	RGPD art. 5
Seudonimización	¿La identidad está separada de la evidencia?	RGPD / recital sobreseudonimización
Contrato	¿Existe DPA/encargo y subencargados identificados?	RGPD art. 28
Retención	¿Sé cuánto tiempo se conservan prompts y salidas?	RGPD limitación temporal; políticas del proveedor
Trazabilidad	¿Puedo registrar prompt, fecha, versión y validación?	RGPD registros; AI Act human oversight/logs
Supervisión	¿La nota final la decide una persona?	AI Act y UNESCO
DPD/DPO	¿He consultado al DPD del centro o de la administración?	RGPD art. 37 sector público

4.3. Estrategias para trabajar con textos y evidencias sin datos identificativos

La anonimización es una estrategia básica para poder usar IA con cierto nivel de seguridad. No consiste solo en borrar el nombre, sino en eliminar o sustituir toda información que permita identificar a una persona directa o indirectamente, es decir, reducir la reidentificación por contexto, contenido o metadatos.

En educación, esto implica revisar no solo el encabezado, sino también el contenido, los metadatos y cualquier pista contextual.

En evaluación, una estrategia sencilla es separar identidad y evidencia mediante códigos, y trabajar únicamente con el texto o el producto que se va a analizar.

Flujo seguro recomendado:

1. Recoger la evidencia.
2. Sustituir nombres por códigos.
3. Eliminar datos del documento.
4. Revisar que no haya pistas evidentes de identidad.
5. Introducir solo el fragmento necesario en la IA.
6. Verificar la respuesta generada.
7. Archivar la versión original en entorno institucional seguro.

Ejemplo práctico:

- “
- Documento original: “María Pérez, 4.º ESO B”.
 - Versión para IA: “Alumno 07, 4.º ESO”.
 - Fragmento enviado: solo el texto de la redacción o el apartado relevante.

A veces basta con cambiar un nombre por un código, pero otras veces es necesario eliminar referencias a profesores, lugares, proyectos o circunstancias familiares que podrían hacer

reconocible al alumno. Cuanto más sensible sea la información, más estricta debe ser la revisión.

Herramientas y niveles de madurez

La decisión no debe hacerse por “marca”, sino por **madurez de gobernanza**. La siguiente tabla resume el uso más prudente en centros.

Nivel	Herramientas tipo	Uso recomendable	Condición de datos	Requisitos contractuales y riesgos
Exploración básica	ChatGPT personal, Claude consumer, Gemini Apps	Ideación de actividades, borradores de rúbricas, ejemplos ficticios	Sin datos reales del alumnado	En Gemini Apps, una parte de los chats puede revisarse por humanos y utilizarse para mejorar servicios; ciertos datos revisados pueden conservarse hasta 3 años. En servicios individuales de OpenAI, los datos de versiones para individuos pueden usarse para entrenamiento. En Claude consumer, inputs/outputs pueden usarse para entrenar salvo opt-out.
Entorno institucional gestionado	ChatGPT Business/Edu/API	Creación de instrumentos, revisión de textos seudonimizados, feedback asistido	Datos minimizados y seudonimizados	OpenAI indica que por defecto no entrena con datos de Business, Enterprise, Edu ni API; ofrece DPA, controles de retención y, en API, casos con zero data retention.

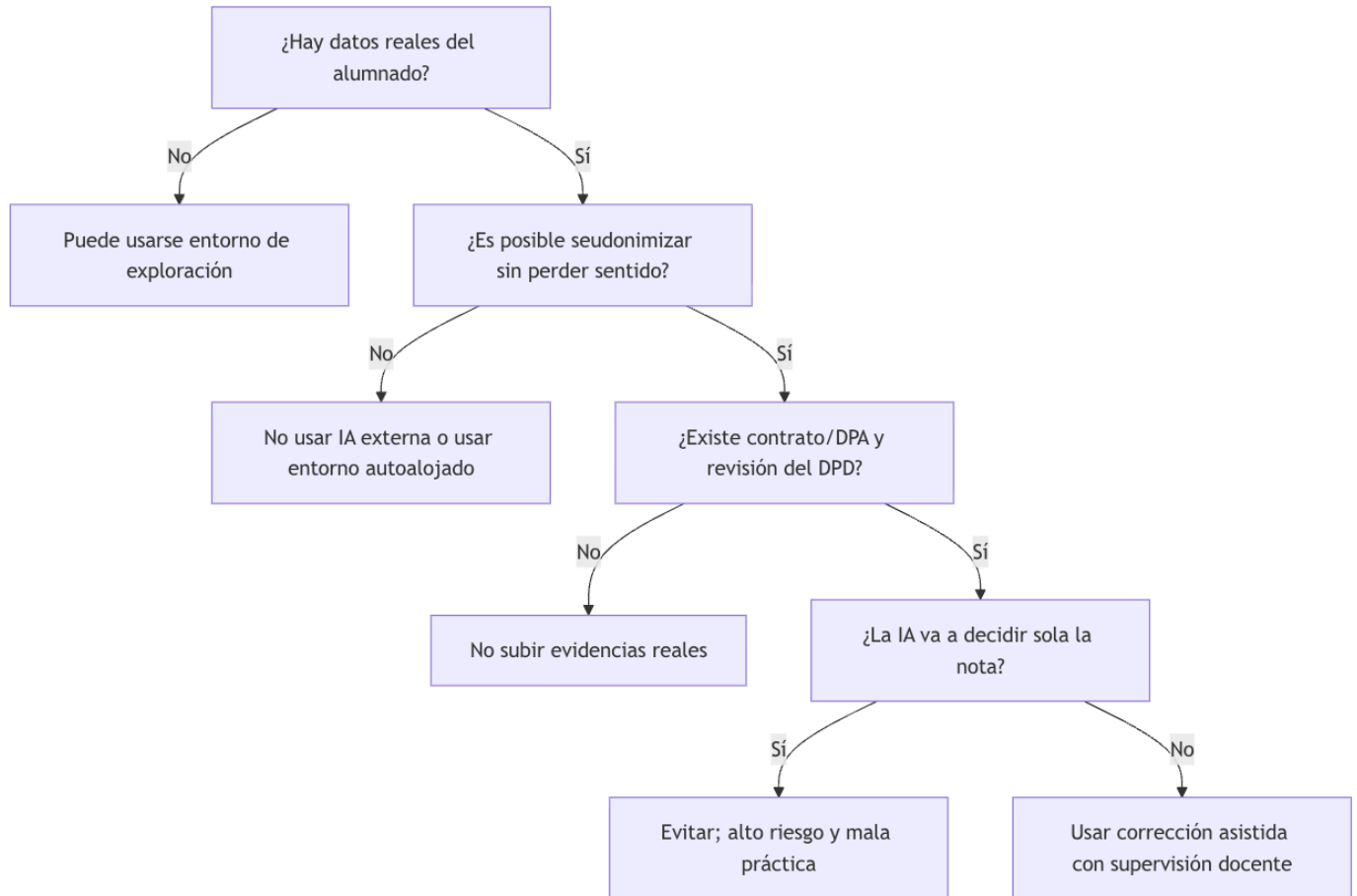
Suite corporativa integrada	Microsoft 365 Copilot	Comentado de documentos, resúmenes y borradores en ecosistema M365	Solo con permisos bien gobernados	Microsoft indica que prompts, respuestas y datos de Graph no se usan para entrenar LLM fundacionales; funciona dentro del perímetro M365. Riesgo principal: permisos mal configurados. Además, algunos modelos Anthropic en experiencias Copilot quedan fuera del EU Data Boundary.
Plataforma cloud avanzada	Vertex AI / Google Cloud	Pipelines con JSON, RAG institucional, despliegues controlados	Datos institucionales bajo diseño técnico y jurídico	Google afirma que no usará datos para entrenar o afinar modelos sin permiso previo; aun así, puede haber logging para abuso y ciertos servicios de grounding almacenan prompts y salidas 30 días.
Servicios comerciales B2B	Anthropic Commercial / API	Corrección asistida o feedback si el centro contrata servicio empresarial	Datos minimizados y contrato	Anthropic establece en términos comerciales que el cliente conserva inputs y outputs y que Anthropic no puede entrenar modelos con Customer Content; además incorpora DPA. En consumo, la política es distinta.



Autoalojada / local	Gemma 4 + Ollama u otra infraestructura propia	Mayor soberanía, pruebas con datos internos, prototipos avanzados	Puede admitir datos reales si la arquitectura lo justifica y el centro controla todo el entorno	Gemma 4 se publica bajo licencia Apache 2.0, con soporte para JSON estructurado y despliegue on-prem; Google la presenta como apta para control total de datos e infraestructura. Ollama permite ejecución local sencilla e integración vía CLI/API. Riesgos: seguridad del servidor, evaluación del modelo, coste de mantenimiento.
----------------------------	--	---	---	--

Sobre herramientas edtech específicas como **MagicSchool**, sí existe evidencia reciente de uso docente en generación de rúbricas, con buena percepción como borrador inicial y necesidad clara de edición humana.

¿Cómo decidir qué herramienta usar?



Aránzazu Cortés + Chat GPT

4.4. Uso de rúbricas, listas de cotejo y criterios de evaluación como base para la corrección con IA

La corrección asistida funciona mejor cuando parte de instrumentos muy claros. Las rúbricas y listas de cotejo son la base ideal para la corrección asistida porque convierten la valoración en evaluación de criterios visibles y explícitos y permiten que la IA trabaje sobre categorías definidas por el docente y ayude a comparar una evidencia con esos criterios proponiendo observaciones y redactando el feedback, pero siempre dentro de la estructura definida por el docente. Esto mejora la consistencia del feedback y evita comentarios genéricos del tipo “deberías mejorar” sin especificar qué ni cómo.

Ejemplos de uso práctico de flujo:

- “
- El docente define una rúbrica para un texto argumentativo.
 - La IA analiza una versión anonimizada del texto.
 - La IA devuelve observaciones según criterios: tesis, estructura, argumentos, cohesión y corrección lingüística.
 - El profesor valida y decide la calificación final.

- “
- El docente define una rúbrica de exposición oral.
 - La IA recibe la transcripción anonimizada.
 - La IA devuelve observaciones vinculadas a criterios.
 - El docente valida, corrige y ajusta la calificación.

Ejemplo de feedback asistido:



La tesis está presente y se identifica con claridad. El texto, sin embargo, necesita mejorar la conexión entre argumentos mediante conectores como 'por otra parte' o 'además'. La conclusión recupera la idea principal, pero no cierra del todo el razonamiento.

Este tipo de respuesta es útil porque convierte la corrección en una orientación pedagógica, no solo en un veredicto.

Esto reduce tiempo y mejora la coherencia, especialmente cuando hay muchos trabajos. Además, favorece una retroalimentación más homogénea entre grupos o profesores.

Plantilla de prompt para corrección asistida con texto anonimizado

“ Analiza este texto anonimizado usando exclusivamente la rúbrica adjunta. No pongas nota final. Indica, por criterio, el nivel estimado, una evidencia textual y una sugerencia breve de mejora. Si faltan datos para valorar un criterio, indícalo explícitamente.

“ Actúa como asistente de evaluación, no como evaluador final.

Vas a analizar un texto anonimizado de alumnado usando exclusivamente esta rúbrica:

[pegar rúbrica]

Tu tarea:

- asignar un nivel provisional por criterio,
- justificar cada propuesta con una evidencia literal del texto,
- señalar incertidumbres o ambigüedades,
- redactar feedback breve, accionable y comprensible para el alumno,
- NO dar calificación final global.

Devuelve una tabla con estas columnas:

criterio | nivel_provisional | evidencia_textual | duda_detectada |
sugerencia_de_mejora

Si falta evidencia suficiente, escribe “requiere revisión docente”.



Este último prompt es especialmente potente porque obliga a la IA a trabajar como **asistente de evidencia**, no como juez final.

Plantilla JSON para corrección asistida

```
{
  "id_anonimo": "ALU_07",
  "tarea": "texto_argumentativo",
  "puntuaciones": {
    "C1": 3,
    "C2": 2,
    "C3": 3,
    "C4": 2
  },
  "evidencias": [
    {
      "criterio": "C1",
      "fragmento": "El uso del móvil debería limitarse en clase..."
    }
  ],
  "feedback_breve": "La tesis es clara; conviene reforzar la conexión entre argumentos.",
  "requiere_revision_docente": true
}
```

La utilidad de estas salidas estructuradas es doble: reducen ambigüedad y facilitan exportaciones a hojas de cálculo, scripts o LMS.

Prompt innovador para salida LMS

“Devuelve el análisis en JSON con campos `id_anonimo`, `criterio`, `nivel`, `evidencia`, `comentario`, `requiere_revision_docente`. No añadas texto fuera del JSON.

Prompt para feedback adaptado

Reescribe este feedback con lenguaje comprensible para , tono motivador y dos pasos concretos de mejora.

4.5. Supervisión docente y validación de resultados

La supervisión del docente es **irrenunciable e indispensable**. La IA puede detectar patrones, pero no comprende por completo el contexto del aula, las características del grupo, la evolución del alumno ni los objetivos específicos de una situación didáctica. Además, la IA puede equivocarse, sesgarse o interpretar mal una evidencia. Por eso, cualquier resultado generado por IA debe ser entendido como borrador, no como decisión final.

Además, en educación hay un problema de fondo: la IA puede parecer segura incluso cuando se equivoca. Por eso es importante que el profesorado mantenga una actitud crítica y no delegue en la herramienta la interpretación de evidencias complejas. La IA debe ahorrar tiempo en tareas mecánicas, no desplazar el juicio profesional.

En evaluación, un comentario aparentemente convincente no equivale a una valoración válida. La decisión final debe mantenerse siempre en manos del profesorado, que conoce el contexto, el nivel real del grupo y las circunstancias del alumnado.

Regla práctica útil:

- La IA propone o sugiere.
- El docente dispone: interpreta y decide.

Prompt para validación docente

“ Revisa si la siguiente propuesta de corrección es coherente con la rúbrica. Señala posibles sobrevaloraciones, infravaloraciones o inferencias no justificadas por la evidencia.

Ejemplo completo de aplicación o corrección asistida



Imagina una actividad de 2.º de ESO: redactar un texto argumentativo sobre el uso del móvil en clase. El proceso podría ser este:

- El alumnado entrega el texto con un código.
- El docente elimina datos personales.
- La IA revisa ortografía, estructura y presencia de tesis, argumentos y conclusión.
- La IA genera un borrador de feedback: “La tesis es clara, pero faltan conectores argumentativos y ejemplos”.
- El docente revisa el resultado, comprueba si la observación es correcta y añade matices.
- La calificación final la decide el profesor con ayuda de la rúbrica.

“ Imaginemos una actividad de 1.º de Bachillerato: comentario crítico sobre un texto periodístico. El proceso podría ser este:

- El alumnado entrega el texto con un código.
- El docente elimina nombres y datos personales.
- La IA revisa estructura, argumentación, uso de conectores y coherencia.
- La IA propone un borrador de feedback.
- El docente revisa si la observación es correcta, añade matices y ajusta la valoración.

Posible salida de la IA:

“El texto presenta una postura clara, pero necesita ejemplos más concretos para sostener el argumento. La coherencia global es adecuada, aunque algunos párrafos podrían unirse mejor mediante conectores.”

Ese comentario puede ser útil, pero solo si el profesor comprueba que refleja fielmente el texto. Si el alumno ya había hecho esa conexión y la IA no la detecta, o si la IA sobreestima un error menor, la intervención docente corrige el sesgo.