

Evaluación Parcial N°3

Encargo: Implementación de Observabilidad

Instrucciones y pauta de evaluación

Estudiante

Sigla	Nombre Asignatura	Tiempo Asignado	% Ponderación
ISY0101	Optativo Ingeniería de Soluciones con IA	2 horas pedagógicas	35%

1. Instrucciones generales

Descripción
<ul style="list-style-type: none">En la Evaluación Parcial N°3 los/las estudiantes desarrollan la tercera parte del proyecto iniciado en la evaluación parcial 1. Implementan métricas de observabilidad y herramientas de trazabilidad sobre agentes de IA ya desarrollados, analicen los registros de ejecución, visualicen el desempeño mediante dashboards y propongan recomendaciones fundamentadas para la optimización del sistema.Medirá los siguientes Indicadores de Logro:<ul style="list-style-type: none">IL3.1 Aplica métricas de observabilidad para medir la precisión, latencia y consistencia de un agente de IA en escenarios con variabilidad de datos.IL3.2 Analiza registros de ejecución del agente, utilizando herramientas de trazabilidad, para identificar puntos de falla o mejora en flujos automatizados.IL3.3 Integra protocolos de seguridad y uso responsable en el diseño de agentes, considerando criterios éticos, normativos y de privacidad, en contextos de producción.IL3.4 Propone mejoras de desempeño o rediseño del agente, basándose en análisis de datos observados, con el fin de aumentar la sostenibilidad y escalabilidad de la solución.

- Las instrucciones de la evaluación se entregan en la semana 13 y el encargo en la semana 17.
- El tiempo asignado para esta evaluación es de **3 semanas y se realiza en parejas**.
- Los/las estudiantes inician su trabajo en el taller de proyectos (TAITE 7) con el/la docente, pero **deben finalizar el encargo en su tiempo de trabajo autónomo**.

Instrucciones específicas de la Evaluación

El encargo debe incluir los siguientes **apartados**:

- A. **Implementación de métricas de observabilidad:** Aplicar al menos tres métricas relevantes (precisión, latencia, consistencia, frecuencia de errores, uso de recursos, etc.) sobre un agente de IA existente. (**IE1, IE2**)
- B. **Análisis de registros y trazabilidad:** Examinar logs de ejecución del agente, identificar puntos críticos de falla o áreas de mejora, y documentar los hallazgos (**IE3**).
- C. **Desarrollo de dashboards de monitoreo:** Construir un dashboard visual (puede ser en Grafana, Kibana, Streamlit, PowerBI, etc.) que muestre el comportamiento del agente según las métricas implementadas. (**IE2, IE4, IE5**)
- D. **Propuesta de recomendaciones:** Elaborar un reporte técnico con recomendaciones prácticas para optimizar el desempeño del agente, justificando cada sugerencia a partir de las métricas y trazabilidad. (**IE6, IE7**)
- E. Redacción técnica, coherente y clara (**IE9**)
- F. Incluir capturas, gráficos y visualizaciones del dashboard (**IE8**)
- G. Citar fuentes y referencias bajo norma APA.

Los **aspectos formales** son:

- El proyecto se entrega mediante un **informe escrito** de un máximo de cinco páginas, elaborado en formato Word o PDF, que integre de manera estructurada todos los aspectos clave del trabajo desarrollado.
- Adicionalmente, se **requerirá la entrega de un repositorio digital (en GitHub o GitLab)** que contenga todo el código fuente desarrollado, junto con documentación sobre su funcionamiento, bocetos de diseño y evidencia de pruebas de software realizadas. Es obligatorio incluir un **archivo README** que explique de manera clara y resumida las instrucciones precisas para ejecutar el sistema y los scripts asociados, permitiendo así que cualquier evaluador **comprenda, aplique y valide** el funcionamiento del proyecto.
- Tanto el informe como el enlace al repositorio deberán **enviarse a través de la plataforma AVA y al correo electrónico del docente**, dentro del plazo estipulado en el cronograma.

Los materiales, herramientas o insumos requeridos son:

- Computador con acceso a internet.
- Herramientas de monitoreo/logs (Grafana, Kibana, Streamlit, PowerBI, etc.).
- Acceso a logs y registros de agentes previos
- Acceso a documentación técnica.
- Documentación técnica y recursos bibliográficos

2. Pauta de Evaluación (Rúbrica, Escala de Valoración, Pauta de Cotejo)

Categoría	% logro	Descripción niveles de logro
Muy buen desempeño	100%	Demuestra un desempeño destacado, evidenciando el logro de todos los aspectos evaluados en el indicador.
Buen desempeño	80%	Demuestra un alto desempeño del indicador, presentando pequeñas omisiones, dificultades y/o errores.
Desempeño aceptable	60%	Demuestra un desempeño competente, evidenciando el logro de los elementos básicos del indicador, pero con omisiones, dificultades o errores.
Desempeño incipiente	30%	Presenta importantes omisiones, dificultades o errores en el desempeño, que no permiten evidenciar los elementos básicos del logro del indicador, por lo que no puede ser considerado competente.
Desempeño no logrado	0%	Presenta ausencia o incorrecto desempeño.

Indicador de Evaluación	Categorías de Respuesta					Ponderación Indicador de Evaluación
	Muy buen desempeño 100%	Buen desempeño 80%	Desempeño aceptable 60%	Desempeño incipiente 30%	Desempeño no logrado 0%	
IE1. Implementa métricas de observabilidad para medir la precisión, consistencia y frecuencia de errores de un agente de IA en escenarios con variabilidad de datos.	Implementa métricas precisas y pertinentes que permiten evaluar en profundidad la precisión, consistencia y errores del agente, adaptándose a escenarios variados.	Implementa métricas adecuadas que cubren mayormente los aspectos de precisión, consistencia y errores del agente en contextos variables.	Implementa métricas básicas que permiten observar parcialmente la precisión y errores del agente, aunque no abordan todos los aspectos.	Implementa métricas poco pertinentes o incompletas, sin lograr una evaluación clara del comportamiento del agente.	No implementa métricas de observabilidad o lo hace de manera totalmente incorrecta.	15%
IE2. Aplica métricas de observabilidad para medir la latencia y el uso de recursos de un agente de IA en escenarios con variabilidad de datos.	Aplica métricas precisas y contextualizadas que permiten medir eficazmente la latencia y uso de recursos del agente en distintos escenarios.	Aplica métricas relevantes para analizar latencia y recursos, con resultados útiles en la mayoría de los contextos.	Aplica algunas métricas funcionales, aunque de forma limitada o con escasa contextualización.	Aplica métricas con errores o sin relación clara con la latencia o el uso de recursos.	No aplica métricas de observabilidad o no logra relacionarlas con el uso de recursos o la latencia.	15%
IE3. Examina logs y eventos durante la ejecución, para identificar posibles errores o cuellos de botella.	Examina con profundidad y precisión los logs y eventos, identificando correctamente errores y cuellos de botella en la ejecución del agente.	Examina los logs de forma adecuada y detecta la mayoría de los problemas presentes.	Examina parcialmente los logs, reconociendo algunos errores pero sin profundizar en su origen.	Examina los logs de manera superficial, sin identificar correctamente los problemas.	No examina los logs o lo hace de manera incorrecta o irrelevante.	15%
IE4. Identifica patrones o anomalías en los registros, para detectar áreas críticas de mejora.	Identifica de forma clara y fundamentada patrones y anomalías, proponiendo mejoras estratégicas con base en los hallazgos.	Identifica correctamente patrones y anomalías relevantes, proponiendo mejoras pertinentes.	Identifica algunos patrones o anomalías, pero sin vincularlos completamente a mejoras concretas.	Identifica patrones o anomalías de forma superficial o sin conexión con mejoras significativas.	No identifica patrones ni anomalías en los registros.	10%

IE5. Construye un dashboard visual que muestra el comportamiento del agente, según las métricas implementadas.	Construye un dashboard visual completo, claro e interactivo, que representa con precisión todas las métricas clave del agente.	Construye un dashboard visual adecuado y comprensible, que representa correctamente la mayoría de las métricas.	Construye un dashboard básico, con algunas métricas clave, pero con limitaciones de visualización o diseño.	Construye un dashboard incompleto o poco claro, que representa métricas de forma parcial o incorrecta.	No construye un dashboard visual o lo hace sin utilidad para el análisis.	15%
IE6. Integra protocolos de seguridad y uso responsable en el diseño de agentes, considerando criterios éticos, normativos y de privacidad, en contextos de producción.	Integra protocolos robustos de seguridad y uso responsable, alineados con criterios éticos, normativos y de privacidad en todos los niveles del diseño.	Integra protocolos adecuados en seguridad y responsabilidad, considerando aspectos normativos y éticos en general.	Integra algunos protocolos básicos, aunque de forma incompleta o sin cubrir todos los criterios éticos o de privacidad.	Integra de forma superficial elementos de seguridad o responsabilidad, sin claridad ni alineación normativa.	No integra protocolos de seguridad ni criterios de responsabilidad en el diseño.	10%
IE7. Propone mejoras de desempeño o rediseño del agente, basándose en análisis de datos observados, con el fin de aumentar la sostenibilidad y escalabilidad de la solución.	Propone mejoras o rediseños bien fundamentados, innovadores y alineados con la sostenibilidad y escalabilidad de la solución.	Propone mejoras pertinentes basadas en un análisis adecuado de los datos observados.	Propone algunas mejoras generales, aunque sin suficiente respaldo en el análisis de datos.	Propone mejoras poco relevantes o mal fundamentadas.	No propone mejoras ni rediseño del agente.	10%
IE8. Elabora un informe técnico, que incluye capturas, gráficos y visualizaciones, respaldando las decisiones de diseño con documentación.	Elabora un informe técnico completo, claro y bien estructurado, que integra evidencia visual y documentación sólida.	Elabora un informe técnico bien organizado, con evidencia visual suficiente y respaldo documental pertinente.	Elabora un informe aceptable, con algunos elementos visuales y documentación parcial.	Elabora un informe incompleto o poco claro, con escasa evidencia visual o documentación.	No elabora el informe técnico o lo presenta sin sustento visual o documental.	5%
IE9. Utiliza en el informe un lenguaje técnico, argumentando sus respuestas con respaldo en evidencias y/o ejemplos concretos.	Utiliza lenguaje técnico preciso y adecuado en todo el informe, con	Utiliza lenguaje técnico apropiado y argumentos coherentes, con	Utiliza lenguaje técnico básico y argumenta sus ideas de manera general,	Utiliza lenguaje poco técnico o argumentos débiles, sin evidencias claras.	No utiliza lenguaje técnico ni presenta argumentación basada en evidencias.	5%

	argumentación sólida y ejemplos bien seleccionados.	algunos ejemplos relevantes.	con ejemplos limitados.			
						Total 100%