

Proyecto

Curso: Calidad del Software

Docente: Bryan Miguel Chaves Salas

Puntos: 100 pts (50%).

Fecha de entrega: 29/08/2025.

Hora de entrega máxima: 13:00 pm.

Instrucciones generales:

1. El proyecto es **estrictamente en grupos de 3 personas**.
2. Si algún estudiante es sorprendido en **actos fraudulentos**, ya sea que los provoque o los consienta, su proyecto y el de los demás implicados, será automáticamente **anulado**.
3. **Solo se recibirá por medio del Campus Virtual de la UTN**. Por cualquier otro medio que se envíe no se revisará.
4. Se **debe entregar en la fecha y tiempo establecido**, por lo cual no se recibirá en fechas posteriores y su nota **automáticamente sera de 0**.
5. El **uso de la inteligencia Artificial** se encuentra **prohibido**, si se logra detectar su **nota** automáticamente será **de 0**.

Esquema de Evaluación del Proyecto

La nota total del proyecto se distribuye de acuerdo con el plan de evaluación del curso, de la siguiente manera:

- **Avances de Proyecto (3):** El desarrollo del proyecto se evaluará mediante tres entregas de avance. Cada una de estas entregas tiene un valor del **5%** de la nota final, sumando un total de **15%**.
- **Proyecto Final y Defensa:** La entrega del informe final, junto con una defensa oral del mismo, tendrá un valor total del **35%** de la nota final del curso.

Análisis de Calidad y Fiabilidad para una Funcionalidad de IA

Objetivo

El objetivo es que se enfrenten a uno de los mayores desafíos de la SQA moderna: probar sistemas "inteligentes" y no determinísticos. Deberán investigar, definir y aplicar una metodología de pruebas para evaluar la calidad de una funcionalidad de IA, enfocándose en aspectos que van más allá del simple "pasa/no pasa", como la ética, el sesgo y la fiabilidad. Este proyecto los posiciona en la vanguardia de la profesión.

Ejemplos Sugeridos:

- **Generador de Imágenes (Copilot, Gemini):** Excelente para explorar la subjetividad, la coherencia con el prompt y los sesgos visuales.
- **Chatbot de Lenguaje (ChatGPT, Claude):** Ideal para probar la coherencia de la conversación, la robustez ante preguntas maliciosas (prompt injection) y la utilidad de las respuestas.
- **IA en Productividad (Notion AI):** Permite evaluar un caso de uso más acotado, como la calidad y precisión de los resúmenes de texto o la coherencia de la redacción automática.

Entregables del Proyecto:

Avance 1: Documento de Investigación y Estrategia de Pruebas de IA

Descripción: Dado que no existe un estándar único para probar IA, su primer paso es convertirse en investigadores. Deben definir su propio marco de trabajo, su "constitución" para evaluar la calidad en este nuevo paradigma.

Contenido Requerido y Nivel de Detalle Esperado:

- ✓ **Investigación y Marco de Trabajo:** Presenten un resumen de su investigación sobre metodologías de testing de IA.
- ✓ **Definición de Atributos y Métricas:** Propongan 3-4 atributos de calidad relevantes para la IA (ejemplo: Sesgo Ético, Coherencia, Robustez, Utilidad). Para cada uno, deben definirlo y proponer cómo lo medirán. Por ejemplo: Para medir el Sesgo de Género, se ejecutarán 10 prompts sobre profesiones neutras (ejemplo: 'doctor', 'ingeniero') y se cuantificará el porcentaje de resultados que muestran un género específico.

Avance 2: Banco de Pruebas y Prompts de Desafío

Descripción: En las pruebas de IA, el caso de prueba es **el prompt**. Su trabajo aquí es ser un ingeniero de prompts **malicioso**, diseñando entradas que lleven a la IA a sus límites y revelen sus debilidades.

Contenido Requerido y Nivel de Detalle Esperado:

Banco de Prompts: Entreguen una lista de 30-50 prompts o entradas de prueba.

Categorización Estratégica: No es solo una lista. Deben categorizar cada prompt según el atributo de calidad que busca medir. Ejemplos de categorías: Prompts para inducir sesgos, Prompts con información contradictoria para probar coherencia, Prompts ambiguos para probar robustez.

Avance 3: Ejecución y Documentación de Hallazgos Cualitativos

Descripción: Esta es la fase de experimentación. Su rol es interactuar con la IA, registrar sistemáticamente sus respuestas y empezar a identificar patrones de comportamiento anómalos.

Contenido Requerido y Nivel de Detalle Esperado:

- ✓ **Log de Ejecución Detallado:** Un documento que muestre, para al menos un 40% de sus prompts, la entrada exacta, la salida literal de la IA (texto o imagen) y sus anotaciones iniciales. La **evidencia es fundamental**.
- ✓ **Reporte de Hallazgos:** Documenten de 3 a 5 de los hallazgos más significativos. Un hallazgo no es un "bug", sino una observación estructurada, como: Hallazgo #1: Sesgo de Representación. Al solicitar imágenes de 'personas en una oficina', el 90% de los resultados mostraron individuos de etnia caucásica, fallando en representar la diversidad.

Entrega Final y Defensa: Informe de Análisis de Calidad y Fiabilidad de IA

Descripción: Su informe final debe analizar los patrones de comportamiento de la IA y traducir sus hallazgos cualitativos en una evaluación de riesgo. Deberán entregar el informe y preparar una defensa oral de su innovadora metodología y conclusiones.

Contenido Requerido y Nivel de Detalle Esperado: El informe debe ir más allá de la lista de hallazgos. Deben presentar un análisis de resultados que discuta cada atributo de calidad, usando sus métricas personalizadas. La conclusión debe responder: ¿Es esta IA segura para el público? ¿Cuáles son los mayores riesgos (éticos, funcionales, de reputación) que presenta? ¿Qué se debería mejorar antes de un lanzamiento?

Rúbricas

Entrega Avance#1

Criterio	Excelente (20 puntos)	Regular (10 puntos)	Insuficiente (0 puntos)	Total
Calidad y Profundidad de la Investigación	El resumen de la investigación es profundo, bien estructurado y demuestra una consulta de fuentes relevantes. Se evidencia un entendimiento claro de los desafíos actuales en el testing de IA y se utiliza esta investigación para fundamentar el marco de trabajo propuesto.	El resumen de la investigación es superficial o se basa en fuentes muy genéricas. Se mencionan los desafíos del testing de IA, pero sin profundizar o conectar la investigación con la estrategia que se propone en el resto del documento.	No se presenta un resumen de la investigación, o este es una copia directa de una o dos fuentes sin análisis ni síntesis propia.	
Relevancia y Definición de Atributos de Calidad	Se proponen de 3 a 4 atributos de calidad (ejemplo: Sesgo Ético, Coherencia) altamente relevantes y apropiados para la funcionalidad de IA elegida. Cada atributo está definido de manera clara, precisa y conceptualmente sólida.	Se proponen atributos, pero algunos son poco relevantes para la IA seleccionada, o sus definiciones son vagas, ambiguas o incorrectas. La selección parece poco meditada.	No se proponen al menos 3 atributos, los atributos son irrelevantes, o sus definiciones están ausentes o son erróneas.	
Diseño de Métricas y Propuesta de Medición	Para cada atributo, se propone un método de medición claro, práctico y cuantificable, similar al ejemplo provisto en la guía (ejemplo: cuantificar un porcentaje a partir de N prompts). La propuesta demuestra creatividad y un enfoque metodológico para evaluar un sistema complejo.	Se propone una forma de medir los atributos, pero es vaga, difícil de ejecutar o no es realmente cuantificable (ej. "se medirá viendo si las respuestas son coherentes"). Falta el detalle metodológico.	No se propone ningún método para medir los atributos definidos, o la propuesta es inviable o ilógica	
Estructura, Claridad y Profesionalismo del Documento	El documento está excepcionalmente bien estructurado como un informe de investigación y estrategia. Tiene un formato profesional, está libre de errores de ortografía o gramática y la	El documento es comprensible, pero presenta fallos en su estructura, o contiene errores ortográficos/gramaticales que	El documento es desorganizado, difícil de seguir, y/o presenta errores gramaticales y ortográficos de forma generalizada,	

	información es fácil de seguir, cumpliendo su rol como "constitución" del proyecto	le restan profesionalismo.	demonstrando falta de esmero.	
Cumplimiento General de Requisitos e Innovación	Cumple y excede todos los requisitos solicitados (resumen de investigación, 3-4 atributos con definición y métrica). El marco de trabajo propuesto es innovador, coherente y demuestra un alto nivel de compromiso con el desafío del proyecto.	Cumple con la mayoría de los requisitos, pero de forma mínima. Por ejemplo, define los atributos, pero las métricas son débiles, o la investigación es muy básica.	No cumple con los requisitos básicos del entregable. Faltan secciones completas o no se entrega la cantidad mínima de elementos solicitados.	

Entrega Avance#2

Criterio	Excelente (20 puntos)	Regular (10 puntos)	Insuficiente (0 puntos)	Total
Calidad y Creatividad de los Prompts	Los prompts están diseñados de manera creativa e inteligente para desafiar a la IA y llevarla a sus límites. Demuestran un pensamiento crítico y "malicioso", como se solicita en la guía, para revelar debilidades, sesgos y fallos.	Los prompts son funcionales pero básicos. Ponen a prueba la funcionalidad principal, pero carecen de la creatividad o el enfoque "malicioso" necesario para descubrir debilidades complejas.	Los prompts son simples preguntas o instrucciones que no constituyen un verdadero desafío para la IA. No demuestran ningún intento de probar los límites del sistema.	
Coherencia de la Categorización Estratégica	Cada prompt está categorizado de forma lógica y coherente según el atributo de calidad que busca medir (definido en el Avance 1). La categorización es estratégica y demuestra un plan de pruebas bien estructurado.	Se realiza una categorización, pero es inconsistente, o algunos prompts no corresponden claramente a la categoría asignada. La estrategia detrás de la categorización es débil o confusa.	Los prompts no están categorizados, o la categorización es ilógica y no tiene relación con los atributos de calidad definidos en la estrategia del proyecto.	
Cobertura de los Atributos	El banco de pruebas en su	La cobertura es	El banco de pruebas no	

de Calidad	conjunto proporciona una cobertura equilibrada y completa de todos los atributos de calidad definidos en la estrategia (ejemplo: Sesgo, Coherencia, Robustez). Hay una cantidad suficiente de prompts para cada categoría.	desequilibrada. El banco de pruebas se enfoca demasiado en uno o dos atributos, dejando otros con una cobertura mínima o nula.	cubre los atributos de calidad definidos en la estrategia del Avance 1.	
Cumplimiento de Requisitos Cuantitativos	Se entrega una lista que cumple o excede el requisito cuantitativo, con 30-50 prompts de alta calidad, demostrando un trabajo exhaustivo	Se entrega una lista de prompts, pero no alcanza la cantidad mínima requerida de 30, o muchos de los prompts son repetitivos o de baja calidad para alcanzar el número.	La cantidad de prompts entregada es significativamente menor a la requerida (menos de la mitad), o no se entrega el banco de pruebas.	
Formato y Claridad del Banco de Pruebas	El documento que contiene el banco de pruebas está perfectamente organizado, es claro y fácil de leer. La estructura (ejemplo. una tabla con columnas para ID, Prompt, y Categoría) es profesional y facilita su uso en la siguiente fase de ejecución.	El documento es comprensible, pero su formato es desordenado o inconsistente, lo que dificulta un poco su lectura y uso.	El banco de pruebas se presenta en un formato desorganizado, confuso o descuidado, lo que demuestra una falta de profesionalismo.	

Entrega Avance #3

Criterio	Excelente (20 puntos)	Regular (10 puntos)	Insuficiente (0 puntos)	Total
Calidad y Detalle del Log de Ejecución	El log de ejecución es detallado y sistemático. Para cada prompt ejecutado, se incluye la entrada exacta, la salida literal y "anotaciones iniciales" perspicaces. La evidencia presentada es clara y completa, cumpliendo con la premisa de que "la evidencia es fundamental".	El log de ejecución está presente, pero la documentación es inconsistente. Faltan detalles en la entrada, la salida, o las anotaciones iniciales son superficiales en varios de los casos. La evidencia es incompleta.	El log de ejecución es desordenado, está incompleto, carece de la evidencia fundamental (salidas literales) o no se entrega.	
Cumplimiento Cuantitativo (Porcentaje y Hallazgos)	Se cumple y excede el requisito cuantitativo: se documenta la ejecución de "al menos un 40% de sus prompts" y se reportan entre "3 a 5 de los hallazgos más significativos" y bien documentados.	Se cumple parcialmente con los requisitos. Por ejemplo, se ejecuta menos del 40% de los prompts, o se reportan menos de 3 hallazgos.	No se cumple con los requisitos cuantitativos mínimos, ya sea en el porcentaje de ejecución o en la cantidad de hallazgos reportados.	
Calidad y Estructura de los Reportes de Hallazgos	Cada hallazgo se presenta como una "observación estructurada", tal como se ejemplifica en la guía. El reporte es claro, analítico y demuestra la capacidad de identificar patrones de comportamiento anómalos en la IA.	Los hallazgos son válidos, pero se presentan más como "bugs" simples que como observaciones estructuradas. Les falta el análisis o la contextualización que se espera (ej. "la imagen salió mal" en lugar de analizar el patrón de error).	Los "hallazgos" reportados no son significativos, no están estructurados, o no describen un problema de calidad real en la IA.	
Profundidad del Análisis en los Hallazgos Reportados	Los hallazgos documentados son "significativos" y demuestran una capacidad de análisis crítico superior. Identifican	Los hallazgos son válidos, pero tienden a ser obvios o de bajo impacto. Se evidencia una experimentación superficial	Los hallazgos reportados son triviales, irrelevantes para los atributos de calidad definidos, o son una mala interpretación de la	

	problemas complejos como sesgos, falta de coherencia o vulnerabilidades, en lugar de errores superficiales.	sin profundizar en los problemas más complejos de la IA.	respuesta de la IA.	
Profesionalismo y Claridad del Documento	El documento en su totalidad es profesional, está bien organizado, es fácil de leer y está libre de errores de redacción. La presentación de la evidencia y los hallazgos es impecable.	El documento es comprensible, pero su formato es desordenado o inconsistente. Presenta algunos errores de ortografía o gramática que le restan profesionalismo.	El documento es descuidado, confuso, contiene errores generalizados y no cumple con los estándares mínimos de calidad profesional.	

Entrega Final

Criterio	Excelente (20 puntos)	Regular (10 puntos)	Insuficiente (0 puntos)	Total
Análisis de Resultados por Atributo de Calidad	El informe presenta un análisis profundo y detallado para cada atributo de calidad definido en la estrategia. Se utilizan de manera efectiva las "métricas personalizadas" para discutir los patrones de comportamiento y los hallazgos cualitativos.	Se discuten los atributos de calidad, pero el análisis es superficial o no utiliza consistentemente las métricas personalizadas. La conexión entre los hallazgos y los atributos definidos es débil.	No se presenta un análisis de resultados por atributo de calidad, o el análisis no tiene relación con las métricas definidas en la estrategia del proyecto.	
Evaluación de Riesgos (Éticos, Funcionales y de Reputación)	El informe aborda de manera explícita y profunda la evaluación de los "mayores riesgos (éticos, funcionales, de reputación)" que presenta la IA. El análisis de riesgo es maduro, bien argumentado y basado en la evidencia recopilada.	Se mencionan algunos riesgos, pero el análisis es superficial o se enfoca en un solo tipo de riesgo (ej. solo funcional), ignorando los demás. La evaluación de riesgo carece de profundidad.	El informe no realiza ninguna evaluación de riesgos, o la evaluación es trivial y no está conectada con los hallazgos del proyecto.	

Calidad de la Conclusión y Recomendaciones de Mejora	La conclusión responde claramente a la pregunta sobre si la "IA es segura para el público" y propone recomendaciones de mejora específicas, prácticas y accionables sobre "qué se debería mejorar antes de un lanzamiento".	La conclusión es ambigua sobre la seguridad de la IA, o las recomendaciones de mejora son genéricas (ejemplo: "se debe mejorar la IA") y poco útiles.	El informe no presenta una conclusión clara sobre la seguridad de la IA ni ofrece recomendaciones de mejora.	
Integración de Avances y Coherencia Narrativa	El informe integra de manera efectiva todos los avances (investigación, banco de prompts, hallazgos) en un documento único y coherente que "va más allá de la lista de hallazgos" y cuenta la historia completa del análisis.	El informe incluye los elementos de los avances anteriores, pero se siente desconectado o como una simple unión de documentos sin una narrativa unificadora. Las correcciones de feedback previo son superficiales.	El informe no integra los avances anteriores, está incompleto o es incoherente. No logra "traducir sus hallazgos cualitativos en una evaluación de riesgo".	
Profesionalismo y Calidad del Informe de Análisis	El informe está redactado con un nivel profesional excepcional. La estructura es lógica, la "traducción de hallazgos cualitativos en una evaluación de riesgo" es clara, y está completamente libre de errores gramaticales y de formato.	El informe es funcional, pero la redacción carece de la claridad esperada para un informe de análisis complejo, o presenta varios errores de formato o gramática que le restan profesionalismo.	El informe está mal redactado, es desorganizado y/o contiene errores generalizados, lo que lo hace inaceptable como un documento de análisis profesional.	

Defensa

Criterio	Excelente (20 puntos)	Regular (10 puntos)	Insuficiente (0 puntos)	Total
Dominio de la Estrategia y Planificación del Proyecto	El estudiante articula con total claridad y confianza la justificación de la selección de aplicaciones, la criticidad de las funcionalidades y la relevancia de los atributos de calidad elegidos. Demuestra un dominio conceptual completo de la estrategia de su proyecto.	El estudiante explica su estrategia, pero titubea o sus respuestas son superficiales. Le cuesta justificar sus decisiones iniciales o no logra conectar los atributos de calidad con su caso de estudio de manera convincente.	El estudiante no puede explicar la estrategia de su proyecto, lee directamente de las diapositivas sin demostrar comprensión, o sus respuestas sobre la planificación son incorrectas o contradictorias.	
Defensa de la Ejecución Técnica y Hallazgos	El estudiante defiende con solvencia el diseño de sus casos de prueba y la lógica detrás de sus reportes de defectos. Explica de manera experta cómo asignó la prioridad y severidad a cada hallazgo, demostrando un conocimiento práctico profundo.	El estudiante describe lo que hizo, pero tiene dificultades para explicar el "porqué" de sus decisiones técnicas. Por ejemplo, no puede justificar claramente por qué un defecto tiene una severidad alta o cómo aplicó una técnica de diseño de pruebas.	El estudiante no puede explicar su proceso de pruebas, no recuerda sus propios hallazgos o no puede justificar la clasificación de los defectos reportados.	
Argumentación del Análisis y la Conclusión Estratégica	El estudiante argumenta de forma persuasiva su análisis de resultados, utilizando los datos y gráficos para respaldar cada afirmación. La defensa de la recomendación final es	El estudiante presenta su conclusión, pero la argumentación es débil o no logra conectar los datos del análisis con la recomendación final de manera efectiva. La defensa se siente más	El estudiante no es capaz de explicar su análisis o su conclusión. La recomendación final parece arbitraria o el estudiante no puede defenderla coherentemente frente a	

	sólida, convincente y demuestra un pensamiento estratégico de alto nivel, tal como lo ejemplifica la guía.	como una opinión que como una conclusión basada en evidencia.	preguntas.	
Capacidad de Respuesta y Manejo de Preguntas	El estudiante responde a las preguntas con seguridad, precisión y profundidad. Sus respuestas demuestran un dominio del tema que va más allá de la presentación, conectando su proyecto con otros conceptos del curso de Calidad del Software.	El estudiante responde a las preguntas de manera básica pero correcta. Sin embargo, puede titubear, dar respuestas incompletas o tener dificultades para responder a preguntas que requieren un pensamiento más allá de lo memorizado para la presentación.	El estudiante no puede responder a las preguntas, sus respuestas son evasivas, incorrectas, o demuestra no haber entendido la pregunta, evidenciando una falta de dominio integral sobre su propio proyecto.	
Profesionalismo en la Comunicación y Presentación Personal	El estudiante utiliza un vocabulario técnico preciso y profesional a lo largo de toda la presentación. Su vestimenta es formal y apropiada para una defensa de proyecto final, proyectando una imagen de seriedad y respeto académico.	El estudiante utiliza un vocabulario a veces informal o impreciso. La vestimenta es casual pero aceptable (no es formal). La presentación personal es adecuada pero no destaca por su profesionalismo.	El estudiante utiliza un vocabulario coloquial, vulgar o incorrecto. La vestimenta es inapropiada para un acto académico formal (ejemplo: gorra, pantaloneta, sandalias), demostrando una falta de respeto por la actividad.	