

Nombre: Geovanny Basantes

Fecha: 27 /04/2025

Asignatura: Normativas de Ingeniería de Software

Ensayo: Aplicación de un Modelo de Calidad de Software en la Explicabilidad de Modelos de Lenguaje de Gran Tamaño (LLMs)

1. Introducción

La tecnología avanza a pasos acelerados y, con ella, la necesidad de que los sistemas sean no solo funcionales, sino también comprensibles para los seres humanos que los utilizan. En el mundo del software tradicional, la calidad ha sido objeto de décadas de estudio, refinándose modelos de evaluación como la familia ISO/IEC 25000 para garantizar que los productos cumplan requisitos internos, externos y en uso.

De forma similar, hoy enfrentamos un nuevo desafío: lograr que los modelos de lenguaje de gran tamaño (LLMs), como GPT, BERT o LLaMA, no solo generen respuestas correctas, sino que también **expliquen** de manera transparente cómo y por qué generan esas respuestas.

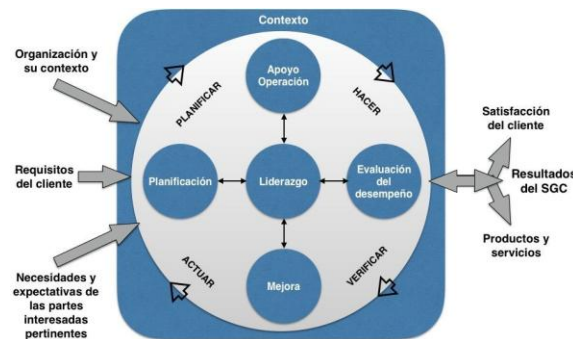


Ilustración 1

Este ensayo explora cómo el enfoque planteado en la tesis "*Modelo de calidad de software aplicado al módulo de talento humano del Sistema Informático Integrado Universitario – UTN*" puede ser adaptado y aplicado al desarrollo de sistemas de **explicabilidad** en modelos de lenguaje, proponiendo un puente entre la ingeniería de software clásica y los desafíos de la inteligencia artificial contemporánea.

2. Desarrollo

La tesis analizada establece un marco sólido para la evaluación de la calidad de software utilizando estándares internacionales. A través de la aplicación de modelos como ISO/IEC 25010, se busca medir aspectos como la funcionalidad, la confiabilidad, la eficiencia y la usabilidad. Esta estrategia no solo permite identificar las fortalezas y debilidades de un sistema, sino también orientar acciones concretas de mejora.

Uno de los principales aportes del trabajo es la división de la calidad en tres dimensiones: Calidad interna que son los atributos que no son visibles para el usuario pero que determinan la solidez del sistema. Calidad externa que el comportamiento observable del software frente a su entorno de operación. Calidad en uso, que se puede entender como la percepción que tiene el usuario sobre el funcionamiento del sistema y su satisfacción al usarlo. Si trasladamos este enfoque al terreno de los LLMs, encontramos una analogía muy potente:

La **calidad interna** de un modelo podría evaluarse observando cómo organiza su atención, qué tan interpretables son sus pesos y cómo están estructuradas sus capas de razonamiento. La **calidad externa** implicaría evaluar si las explicaciones que proporciona un modelo son coherentes, verificables y alineadas con los resultados generados.

La **calidad en uso** se centraría en la experiencia del usuario: ¿las explicaciones son comprensibles para el público objetivo? ¿Aumentan la confianza en el modelo? ¿Permiten al usuario corregir errores o detectar sesgos?

La tesis también enfatiza el uso de instrumentos de evaluación estructurados, como las matrices de calidad, donde cada característica y subcaracterística se evalúa con métricas específicas. En un proyecto de explicabilidad para LLMs, podríamos construir una matriz similar que incluya, por ejemplo:

- **Claridad** de las explicaciones (medida por encuestas de comprensión),
- **Coherencia** entre la respuesta y la explicación,
- **Nivel de detalle** adecuado al contexto del usuario,
- **Rastreo de fuentes** (capacidad de identificar qué partes del input llevaron a una determinada salida).

Otro aspecto vital que rescata el proyecto de la UTN es el reconocimiento de que la calidad no se garantiza en un solo paso. La calidad es un proceso de mejora continua, que requiere medición, análisis, propuesta de mejoras e implementación iterativa.

En el contexto de LLMs, esto significa que no basta con diseñar un primer sistema de explicabilidad: se necesita establecer un proceso cíclico donde las explicaciones generadas se evalúan regularmente, se reciben retroalimentaciones de usuarios reales y se realizan ajustes continuos en los mecanismos de explicación. Echeverry (2008)

Finalmente, la tesis resalta algo que es esencial en la calidad también es una cuestión de cultura organizacional. Implantar calidad implica concienciar a todos los actores involucrados. En tu proyecto de explicabilidad, esto se traduce en integrar la explicabilidad no como un accesorio, sino como un principio desde el diseño, involucrando tanto a desarrolladores, como a testers, expertos en ética y usuarios finales.

Aplicar un modelo de calidad de software como el propuesto en la tesis de la UTN al problema de la explicabilidad de modelos de lenguaje es más que una adaptación técnica: es una evolución conceptual.

Así como la calidad de un software garantiza que cumpla sus funciones de forma confiable y satisfactoria, la explicabilidad en un LLM garantiza que sus operaciones sean transparentes, comprensibles y, en última instancia, confiables para los usuarios humanos. Adaptar las nociones de calidad interna, externa y en uso nos permite construir modelos de evaluación robustos que no solo midan qué tan "inteligente" es un LLM, sino también qué tan responsable, comprensible y seguro es.

Al incorporar prácticas como el uso de métricas específicas, matrices de evaluación y procesos iterativos de mejora, podemos avanzar hacia una inteligencia artificial más humana, más ética y más integrada en nuestras sociedades.

La tesis nos recuerda que la calidad no es un destino, sino un camino. De la misma manera, la explicabilidad no debe ser vista como una característica opcional en los LLMs, sino como un compromiso continuo con la transparencia y la confianza.

2.1. De la calidad de software a la calidad de explicabilidad

La tesis analizada no se limita a validar el correcto funcionamiento de un sistema. Más bien, busca garantizar que su desempeño sea comprensible, verificable y útil para los

usuarios reales. Esto va mucho más allá de la lógica técnica: introduce la necesidad de humanizar el software, de hacerlo accesible, transparente y confiable.

Del mismo modo, cuando pensamos en la explicabilidad de los LLMs, el objetivo no es únicamente que el modelo funcione "bien", sino que su razonamiento pueda ser entendido.

Una respuesta correcta que no puede explicarse genera desconfianza; una respuesta imperfecta pero honesta y clara puede, en cambio, fortalecer la relación entre humanos y máquinas.

2.2. Calidad Interna, Externa y en Uso en LLMs

Una de las grandes contribuciones del trabajo de tesis es la adopción de un modelo tridimensional de calidad. Este marco puede ser trasladado de manera natural a los LLMs:

Calidad Interna: Aquí evaluaríamos si el modelo ha sido diseñado con arquitecturas que favorecen la trazabilidad de las decisiones. Por ejemplo: ¿Es posible identificar qué capas de la red han influido más en una decisión?, ¿Se utilizan mecanismos como mapas de atención o razonamiento simbólico para favorecer interpretabilidad?

Calidad Externa: Es la forma en que se evalúa desde el enfoque del comportamiento observable. Para explicabilidad, implica responder: ¿Las explicaciones generadas son consistentes entre distintos inputs similares?, ¿Pueden ser validadas externamente por expertos o usuarios comunes?

Calidad en Uso: En la calidad de uso se busca enfocar en lo humano: ¿Los usuarios comprenden fácilmente las explicaciones ofrecidas?, ¿Las explicaciones aumentan la confianza del usuario en el sistema?, ¿Son útiles para tomar decisiones informadas? De este modo, el mismo paradigma que ayuda a evaluar un software administrativo en una universidad puede guiar el desarrollo de una IA ética y responsable.

2.3. Instrumentos de medición: hacia una matriz de explicabilidad

En su tesis, Tulia Vaca Sierra propone una Matriz de Calidad como herramienta clave. Esta matriz no solo organiza las características a evaluar, sino que asocia cada una con métricas objetivas y ponderaciones.

Para tu proyecto de explicabilidad en LLMs, construir una matriz similar sería esencial.

Característica de Explicabilidad	Subcaracterística	Métrica	Método de Evaluación
Claridad	Uso de lenguaje comprensible	% de usuarios que entienden la explicación sin ayuda	Encuesta post-interacción
Coherencia	Consistencia lógica entre salida y explicación	Número de inconsistencias detectadas	Revisión de expertos
Precisión	Fidelidad de la explicación al razonamiento real	Diferencia entre trazabilidad real y explicación ofrecida	Análisis técnico
Utilidad	Aplicabilidad práctica de la explicación	% de usuarios que toman mejores decisiones tras explicación	Experimento controlado

Estas métricas no solo permitirían evaluar el desempeño de tu sistema, sino también trazar un plan de mejora continua, exactamente como se hizo en el módulo de talento humano de la UTN.

2.4. La importancia del usuario: explicabilidad centrada en humanos

La tesis enfatiza que la calidad no puede evaluarse de espaldas al usuario.

No basta que un sistema funcione: debe funcionar para alguien, en un contexto real.

De la misma forma, la explicabilidad de LLMs debe centrarse en las necesidades, capacidades y limitaciones humanas:

- Una explicación demasiado técnica, aunque precisa, puede ser inútil para un usuario general.
- Una explicación simplificada puede ser mejor si mejora la comprensión, incluso a costa de omitir ciertos detalles técnicos.
- La confianza del usuario debe ser un objetivo explícito, no un efecto colateral.

Aplicar esta lógica implica diseñar explicaciones adaptadas que cambien su nivel de detalle según el tipo de usuario, sus expertos o el contexto de uso, sobre todo, implica escuchar constantemente a los usuarios, mediante retroalimentación continua, para mejorar la calidad de las explicaciones ofrecidas.

3. Conclusión:

La calidad y en nuestro caso, garantizar la explicabilidad para instalar procesos como de

evaluación, ajuste y mejora de modo que se tiene profundas implicaciones para tu proyecto, sin embargo, basta con diseñar un primer sistema de explicaciones.

Para ello es necesario planificar ciclos de mejora iterativa (por ejemplo, cada seis meses), en los que midas el desempeño explicativo, recojas feedback real y ajustes los mecanismos. Además, deberías fomentar una cultura de la explicabilidad en tu equipo: que todos los diseñadores, desarrolladores, testers y gestores valoren y prioricen la transparencia como un requisito esencial, no como un accesorio.

La experiencia de evaluar la calidad del módulo de talento humano en la UTN nos deja una lección clara como medir, analizar y mejorar de forma sistemática es indispensable para construir sistemas de calidad, el reto de los modelos de lenguaje no es solo alcanzar altos puntajes de precisión que cada respuesta generada no solo sea correcta, sino también pueda ser entendida, verificada y, en última instancia, confiada por sus usuarios al adaptar los principios y métodos de evaluación de calidad de software como la división entre calidad interna, externa y en uso, o la construcción de matrices de evaluación que nos permitan dotar a los LLMs de un marco robusto para crecer no solo en inteligencia, sino en responsabilidad y humanidad, este enfoque recuerda que la confianza es una característica que no es intrínseca si no que se va formando a la par de cumplir las expectativas del usuario y en el mundo de la inteligencia artificial, construir confianza empieza, sin duda, por hacer de la explicabilidad un compromiso real, profundo y permanente.



Referencias:

- Echeverry, A. M. L., Ayala, L. E. V., & Cabrera, C. (2008). Introducción a la calidad de software. *Scientia et technica*, 2(39), 326-331.
- Lovelle, J. M. C. (1999). Calidad del software. *de Grupo Gidis Universidad Nacional de la Pamba, Oviedo*.
- García León, D., & Beltrán Benavides, A. (1995). Un enfoque actual sobre la calidad del software. *Acimed*, 3(3), 40-42.
- Velthuis, M. G. P., & Muñoz, C. C. (2010). *Calidad del producto y proceso software*. Editorial Ra-Ma.
- Sierra, T. N. V. (2017). *Modelo de calidad de software aplicado al módulo de talento humano del sistema informático integrado universitario–UTN* (Doctoral dissertation, UNIVERSIDAD TÉCNICA DEL NORTE).