



11°
Congreso
TUCUMÁN
2023

20
23



CoNaII SI

XI Congreso Nacional
de Ingeniería en Informática
y Sistemas de Información



XI Congreso Nacional de Ingeniería en Informática y Sistemas de Información.

CoNaISI 2023

Universidad Tecnológica Nacional.
Facultad Regional Tucumán

XI Congreso Nacional de Ingeniería en Informática y Sistemas de Información :
CoNalSI 2023 / Gabriela Ivana Aparicio ... [et al.] ; Compilación de José Augusto
Nasrallah ; Gabriela Ivana Aparicio ; Editado por Fernando Cejas. - 1a ed - Ciudad
Autónoma de Buenos Aires : edUTecNe, 2024.
Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-987-8992-38-9

1. Ingeniería. 2. Sistemas de Información. 3. Tecnología Informática. I. Aparicio, Gabriela
Ivana, comp. II. Nasrallah, José Augusto, comp. III. Cejas, Fernando, ed.
CDD 629.801

Diseño de Tapa e interior: Fernando Cejas



Universidad Tecnológica Nacional – República Argentina
Rector: Ing. Rubén Soro
Vicerrector: Ing. Haroldo Avetta
Secretaría Cultura y Extensión Universitaria: Ing. Federico Olivo Aneiros



Universidad Tecnológica Nacional – Facultad Regional Tucumán
Decano: Mgter Ing Walter Fabián Soria
Vicedecano: Ing. Juan Esteban Campos
Secretario de Vinculación Tecnológica: Ing. Mario Augusto Madariaga
Secretario de relaciones institucionales: Ing. Jorge Molina



edUTecNe – Editorial de la Universidad Tecnológica Nacional
Coordinador General a cargo: Fernando Cejas
Dirección General: Mg. Claudio Véliz
Dirección de Cultura y Comunicación: Ing. Pablo Lassave

Queda hecho el depósito que marca la Ley N° 11.723

© edUTecNe, 2024

Sarmiento 440, Piso 6 (C1041AAJ)

Buenos Aires, República Argentina

Publicado Argentina – Published in Argentina

ISBN 978-987-8992-38-9



Reservados todos los derechos. No se permite la reproducción total o parcial de esta obra, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio (electrónico, mecánico, fotocopia, grabación u otros) sin autorización previa y por escrito de los titulares del copyright. La infracción de dichos derechos puede constituir un delito contra la propiedad intelectual.

Evaluación de GPT-4 como Herramienta Auxiliar en la Elaboración de Criterios para Rúbricas Analíticas

Facundo Nicolas Romeu - Juan Manuel Stecklain
Universidad Abierta Interamericana

FacundoNicolas.romeu@Uai.edu.ar - JuanManuel.Stecklain@alumnos.uai.edu.ar

Resumen

Este estudio explora la aplicabilidad del modelo GPT-4 de OpenAI en la construcción de rúbricas, esenciales en la evaluación académica de competencias. Iniciamos delineando los fundamentos de las rúbricas analíticas y procedemos a describir el entrenamiento personalizado que GPT-4 recibió para tal fin. GPT-4 demostró ser eficaz en la definición de criterios y descriptores con la precisión y flexibilidad requerida al momento de asistir en el desarrollo de rúbricas. Los resultados en el uso de esta herramienta son evaluados, mostrando prometedores beneficios en la práctica educativa. Adicionalmente, se introduce una propuesta basada en React, que se interconecta con la API de ChatGPT, optimizando la interacción con GPT-4 durante la construcción de rúbricas.

Palabras claves

Rúbricas, inteligencia artificial, GPT

Rúbricas analíticas

La evaluación orienta y motiva el aprendizaje y la propia enseñanza [1] y no debe ser considerada como un proceso separado del aprendizaje, sino una oportunidad para fomentarlo [2]. En este sentido, las rúbricas analíticas no se limitan a la evaluación, sino que también son estrategias de aprendizaje que permiten descomponer tareas complejas, relacionadas con competencias, en tareas más simples, distribuidas de forma gradual y operativa [3]. Por otro lado, el uso de las rúbricas analíticas se ve fortalecido en los entornos virtuales de enseñanza y aprendizaje, ya que su diseño y utilización empoderan al docente, al estudiante y a la institución, fomentando así una cultura de evaluación [4]. Además, el uso de rúbricas analíticas puede ser un instrumento válido para promover el aprendizaje, especialmente cuando se combinan con diferentes actividades metacognitivas, como la autorregulación, la evaluación por pares o la autoevaluación [5]. La rúbrica analítica presentada en [6] mide el nivel de logro en la resolución de problemas al

relacionar indicadores con niveles de dominio que incluyen un formato de descriptores analíticos que desambigua su uso. El rasgo diferenciador está centrado en la construcción de los *descriptores*; cada criterio de evaluación se puede subdividir en uno o más descriptores y cada uno de ellos, a su vez, se componen de un *contexto* vinculado y dos *criterios analíticos* A+ y B, que podrán ser contrastados con el trabajo evaluado para establecer si lo cumplen, o no.

Componentes principales de una Rúbrica

Criterios de Evaluación: Representan los aspectos esenciales que se van a evaluar en una tarea específica. Estos pueden variar según la complejidad del tema o problema en estudio. Por ejemplo, si se evalúa un ensayo escrito, los criterios podrían incluir la organización del texto, la claridad de la tesis, la adecuación de las fuentes, entre otros.

Descriptores: Son las especificaciones que detallan las características de cada nivel de logro dentro de un criterio. Los descriptores aportan claridad sobre lo que se espera en cada nivel y permiten que la evaluación sea menos subjetiva. Por ejemplo, en el criterio "claridad de la tesis" de un ensayo, un descriptor podría especificar que, para alcanzar el nivel "Excelente", la tesis debe ser clara, concisa y original.

Niveles de Logro: Cada criterio en una rúbrica analítica se descompone en varios niveles que representan la calidad del trabajo del estudiante en relación con ese criterio. Los niveles comunes pueden ser "Excelente", "Bueno", "Aceptable" y "Insuficiente", aunque estos pueden variar según la rúbrica.

Puntuación: Cada nivel de logro puede tener asociada una puntuación que permite calcular una calificación final. La suma de las puntuaciones de todos los criterios brinda una visión integral del rendimiento

del estudiante en la tarea.

Beneficios de las Rúbricas

Objetividad: Al tener criterios y descriptores claramente definidos, se reduce la subjetividad en el proceso de evaluación.

Transparencia: Las rúbricas, cuando se comparten con los estudiantes, les proporcionan una comprensión clara de las expectativas, permitiéndoles orientar mejor su trabajo.

Retroalimentación Constructiva: Los docentes pueden proporcionar feedback específico y constructivo basado en los criterios y descriptores.

Consistencia: Las rúbricas garantizan que todos los estudiantes sean evaluados bajo los mismos estándares, lo que asegura una mayor justicia en el proceso de evaluación. En resumen, una rúbrica es una herramienta esencial que estructura y guía el proceso de evaluación, proporcionando claridad, objetividad y coherencia. La combinación de criterios bien definidos y descriptores detallados permite a los docentes evaluar de manera justa y a los estudiantes comprender y mejorar su desempeño.

Descripción del problema

En el ámbito educativo y profesional, las rúbricas analíticas representan una herramienta valiosa para estructurar y estandarizar evaluaciones, permitiendo una interpretación coherente y objetiva de los resultados obtenidos. Sin embargo, uno de los desafíos más significativos en su desarrollo radica en la definición de criterios y descriptores que sean a la vez claros y pertinentes. Esta tarea, a pesar de su aparente simplicidad, es en realidad compleja. Los criterios deben ser exhaustivos y abordar todos los aspectos relevantes de la actividad o tarea evaluada, sin ser demasiado generales ni demasiado específicos. A su vez, los descriptores, que amplían y detallan los criterios, requieren precisión para reducir la ambigüedad y evitar interpretaciones erróneas o subjetivas.

La inconsistencia o ambigüedad en la definición de estos elementos puede llevar

a evaluaciones sesgadas, lo que puede tener consecuencias negativas en la formación y feedback de los estudiantes o profesionales. Esto puede resultar en interpretaciones erróneas de las habilidades o conocimientos de un individuo y en decisiones educativas o profesionales inapropiadas.

A pesar de la formación y experiencia de los educadores y profesionales, la elaboración manual de criterios y descriptores sigue siendo susceptible a la subjetividad y a las limitaciones humanas. En este contexto, surge la pregunta: ¿Es posible utilizar herramientas avanzadas, como el modelo GPT-4 de OpenAI, para auxiliar en esta tarea y reducir su complejidad?

El problema central que este paper busca abordar, por lo tanto, es explorar la potencialidad de GPT-4 para asistir en la elaboración de criterios y descriptores de rúbricas analíticas, y determinar si puede ofrecer una solución más sistemática, coherente y objetiva que los métodos tradicionales.

Metodología

Antes de presentar esta investigación, desarrollamos una herramienta denominada "UAI Rubrics". Esta aplicación fue creada con el propósito de ofrecer una plataforma dedicada al desarrollo, búsqueda, uso y publicación de rúbricas analíticas. A partir de la utilización de "UAI Rubrics" por parte de los docentes y los feedbacks proporcionados, identificamos que los desafíos más significativos en el diseño de rúbricas analíticas radican en la precisa definición de criterios y descriptores. En base a esta observación, el objetivo de este trabajo es entrenar un modelo GPT-4 incorporando reglas y conocimientos específicos, con la finalidad de facilitar y brindar soporte a los profesores durante el proceso de creación de estas rúbricas. Paralelamente, estamos avanzando en la concepción de un modelo que pueda ser integrado a "UAI Rubrics" y que actúe como asistente en esta tarea.

Capacitación de modelos de lenguaje generativo

Durante el proceso de entrenamiento de un modelo de lenguaje generativo, es esencial prestar especial atención a la información que le proporcionamos. No solo es importante el contenido, sino también el contexto en el que lo situamos y la manera en que establecemos la comunicación con la inteligencia artificial. Ofrecerle un marco adecuado es crucial para obtener respuestas precisas y pertinentes. Según Dathathri et al. (2020), los "Prompts" o indicaciones que se deben considerar al interactuar con el modelo incluyen: tema, estilo, tono, contexto general y detalles del contexto. Estos componentes aseguran que la IA comprenda y responda de manera coherente y relevante al input proporcionado. Al redactar un "prompt", consideramos esencial adherirse a las recomendaciones propuestas por [7]. Estas sugerencias enfatizan la importancia de:

a) Establecer un objetivo claro al interactuar con Chat GPT. b) Mantener el mensaje breve y al punto. c) Emplear un lenguaje natural, evitando ambigüedades [8]. d) Proporcionar el contexto necesario para que Chat GPT entienda la esencia de la conversación y pueda responder adecuadamente. e) Evitar preguntas generales o abiertas, optando en su lugar por palabras y frases concretas que orienten hacia el resultado esperado [9]. f) Ser conscientes del tono y del lenguaje utilizado durante la interacción. g) En situaciones donde se utilice jerga o términos técnicos, es recomendable ofrecer definiciones claras para prevenir malentendidos. h) Antes de enviar el mensaje, es prudente revisarlo detenidamente para asegurar su claridad y precisión.

Proceso de Capacitación de GPT-4

Durante el proceso de entrenamiento del modelo de lenguaje generativo para asistir en la elaboración de rúbricas analíticas, y tras múltiples intentos, coincidimos con las

observaciones presentadas por [7]. Enfatizamos la precisión en la definición de las tareas. Por ello, decidimos entrenar dos modelos distintos: uno enfocado en la definición de criterios y otro en la definición de descriptores. Esta decisión buscó asegurar que las responsabilidades de cada modelo fueran más precisas, claras y menos ambiguas. Además, establecimos con claridad las reglas y el alcance de cada modelo, definiendo tanto sus capacidades como sus limitaciones. Proporcionamos al modelo amplia información sobre cómo estructurar los diferentes componentes de la rúbrica y ofrecimos numerosos ejemplos que ilustran la correcta definición de criterios y descriptores.

Tras ofrecer la formación técnica y proporcionar el contexto necesario para abordar las problemáticas emergentes, llevamos a cabo un proceso de asignación de roles. En este proceso, se le especificó a GPT-4 cuáles serían sus responsabilidades y cómo debería responder ante diferentes preguntas o estímulos

Proceso de entrenamiento

Tras la capacitación, se llevaron a cabo sesiones de prueba donde la IA asumió el papel de asistente. En estas sesiones, se presentaron diversos escenarios en los cuales se debían desarrollar rúbricas con temáticas variadas. Posteriormente, comparamos las respuestas generadas por el modelo con los resultados esperados, basándonos en las reglas propuestas por [6]. Informamos al modelo sobre los resultados obtenidos, tanto en los casos en que sus respuestas fueron adecuadas como en aquellos donde no lo fueron. Tras reiterar este proceso evaluativo y obtener consistentemente los resultados esperados, podemos afirmar que el modelo alcanzó un nivel de estabilidad en su conocimiento.

Resultados obtenidos

Tras el proceso de capacitación, se evaluaron los resultados alcanzados por ChatGPT en la

tarea de creación de criterios y descriptores. A continuación, presentamos un análisis de algunos resultados obtenidos:

Se solicitó al modelo entrenado que proporcionara un ejemplo de criterios y descriptores para la rúbrica de un diagrama de casos de uso UML. La solicitud fue la siguiente: “Necesito un ejemplo de criterios y descriptores para la rúbrica de diagrama de casos de uso UML”

Totalidad de criterios obtenidos:

- Criterio 1: Identificación de Actores.
- Criterio 2: Representación de Casos de Uso.
- Criterio 3: Relaciones entre actores y casos de uso.
- Criterio 4: Inclusión y Extensión.

Si vamos a la descripción de criterios según [10] Los criterios de evaluación tienen como finalidad establecer los elementos sobre los que se va a basar la evaluación del aprendizaje y, por consiguiente, ofrecen al estudiante los aspectos esenciales de la tarea que será objeto de valoración por parte del docente. Analizando los resultados obtenidos podemos entender que la respuesta otorgada por el modelo es correcta y podría ser de utilidad para un modelo de rúbrica orientado para diagramas de casos de uso UML. A continuación, se analizarán los resultados obtenidos por el modelo al detallar el contexto y descriptores.

Ejemplo 1

Se le solicitó a la IA que defina un contexto y descriptores para el Criterio 3

Criterio 3: Relaciones entre actores y casos de uso.

- Contexto 3.1: Uso adecuado de líneas de asociación.
 - Descriptor A+: Las líneas de asociación entre actores y casos de uso están correctamente dibujadas.
 - Descriptor B: Algunas líneas de asociación están

mal ubicadas o son incorrectas.

- Contexto 3.2: Representación clara de multiplicidad
 - Descriptor A+: Todas las relaciones indican claramente la multiplicidad cuando es necesario.
 - Descriptor B: Algunas relaciones carecen de indicadores de multiplicidad o son incorrectas.

Si analizamos las respuestas obtenidas y vamos a la definición de descriptores según [6] los descriptores permiten detallar las características de cada nivel de logro según el grado en que los criterios de evaluación hayan sido alcanzados. Cada uno de ellos, a su vez, estará compuesto por un contexto vinculado y dos criterios analíticos A+ y B, que podrán ser verdaderos o falsos. Podemos observar que la respuesta otorgada cumple con las condiciones presentadas en la definición de descriptores, además podemos observar que la definición de descriptores vino acompañada del contexto vinculado necesario para entender los descriptores.

Ejemplo 2

Se le solicitó a la IA que defina un contexto y descriptores para el Criterio 4

Criterio 4: Inclusión y Extensión

- Contexto 4.1: Correcta representación de relaciones de inclusión.
 - Descriptor A+: Las relaciones de Inclusión están correctamente identificadas con flechas y etiquetas adecuadas.
 - Descriptor B: Algunas relaciones de inclusión son incorrectas o carecen de etiquetas.
- Contexto 4.2: Correcta representación de relaciones de extensión

En referencia a la definición dada anteriormente vemos que este resultado también está acompañado de su contexto vinculado y tiene descriptores capaces de ser verdaderos o falsos.

Conclusión y futuros trabajos.

Como conclusión, entendemos que el uso de modelos de lenguaje generativos como ChatGPT nos permite facilitar la definición de criterios de aprendizaje y descriptores para la creación de rúbricas analíticas. Actualmente, en base a lo expuesto en este trabajo, se está desarrollando un modelo capaz de integrarse en aplicaciones, lo cual nos permitirá asistir a los usuarios en la creación de rúbricas analíticas. Este modelo va a tener la ventaja de poder ser entrenado y mejorar su funcionamiento mediante los procesos de entrenamiento detallados en este trabajo.

Referencias

- [1] M. De Miguel Díaz, *Modalidades de enseñanza centradas en el desarrollo de competencias: Orientaciones para promover el cambio metodológico en el Espacio Europeo de Educación Superior*, 1st ed. 2005.
- [2] V. Torres-Sanz, J. Tramullas, J. A. Sanguesa, P. Garrido, and F. J. Martínez, "Rúbricas como estrategia de evaluación en entornos TICS [Rubrics as an evaluation strategy in ICT environments]," 2017.
- [3] J. Alsina Masmitjà, "Rúbricas para la evaluación de competencias," *Barcelona: Octaedro. Recuperado el*, vol. 19, no. 09, p. 2018, 2013.
- [4] N. Battaglia, C. Neil, and M. De Vincenzi, "Software Engineering Competence-Based Learning in Collaborative Virtual Environments," *V IEEE Congreso Mundial de Educación en Ingeniería (EDUNINE2021)*, 2021.
- [5] J. Valverde Berrocoso, "El uso de e-rúbricas para la evaluación de competencias en estudiantes universitarios: Estudio sobre fiabilidad del instrumento," *REDU: Revista de Docencia Universitaria*, 2014.
- [6] C. Neil, N. Battaglia, and M. De Vincenzi, "Marco metodológico para el diseño de rúbricas analíticas," *Edulec Revista Electrónica*, vol. 80, no. 1, pp. 198–215, 2022, doi: 10.21556/edulec.2022.80.2425.
- [7] M. A. Morales-Chan, "Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza," 2023.
- [8] P. Vossen, T. Caselli, G. Lapesa, and H. Saggion, "A survey of current datasets for natural language processing and understanding," 2018.
- [9] L. Zhou and K. Small, "Multi-domain dialogue state tracking as dynamic knowledge graph enhanced question answering," *arXiv preprint arXiv:1911.06192*, 2019.
- [10] Nicolás Battaglia, Carlos Neil, and Marcelo De Vincenzi, "Desarrollo y evaluación de competencias en la ingeniería de software en un entorno virtual de aprendizaje colaborativo," in *XXIII Workshop de Investigadores en Ciencias de la Computación*, 2021, pp. 875–880.