

## ANÁLISIS COMPARATIVO ENTRE LOS CRITERIOS DEL DOCENTE Y DE LA INTELIGENCIA ARTIFICIAL EN LA CALIFICACIÓN DE EVALUACIONES PRÁCTICAS DE LA ASIGNATURA COMPUTACIÓN Y CÁLCULO NUMÉRICO

### COMPARATIVE ANALYSIS BETWEEN TEACHER CRITERIA AND ARTIFICIAL INTELLIGENCE IN GRADING PRACTICAL ASSESSMENTS OF COMPUTER SCIENCE AND NUMERICAL CALCULATION SUBJECT

Beatriz I. Pedrotti, Juan F. Weber, María B. Sánchez, Ricardo A. Romero y Carlos A. Fernández

*Cátedra de Computación y Cálculo Numérico – Departamento de Computación – Facultad de Ciencias Exactas, Físicas y Naturales – Universidad Nacional de Córdoba. Av. Vélez Sarsfield 1611, X5016GCA Córdoba, Argentina, [bpedrotti@unc.edu.ar](mailto:bpedrotti@unc.edu.ar).*

**Palabras clave:** Evaluaciones Universitarias Masivas, Cálculo Numérico, Inteligencia Artificial.

**Resumen.** Las oportunidades que presenta la Inteligencia Artificial, en contextos educativos universitarios masivos y como asistente a la labor docente evaluativa, motivan el desarrollo de un proyecto de investigación al interior de la Cátedra de Computación y Cálculo Numérico, de la Facultad de Ingeniería de la Universidad Nacional de Córdoba. Actualmente se desarrolla la segunda fase cuyo fin es responder a la pregunta: ¿Las herramientas de Inteligencia Artificial preseleccionadas son capaces de emular el criterio del cuerpo docente en términos de las evaluaciones prácticas? La prueba se realizó sobre un universo de 51 evaluaciones prácticas, cada una basada en la resolución de tres problemas de ingeniería, con programas en lenguaje Python 3.x y herramientas del cálculo numérico. Las evaluaciones fueron calificadas por dos docentes de la cátedra y por seis *chatbot*. El instrumento de evaluación es la rúbrica, conformada por diez criterios en total y dos niveles de logro por cada uno. El análisis comparativo se realizó principalmente con la herramienta Matriz de Confusión. Los resultados exitosos impulsan al equipo a la tercera fase del proyecto para realizar *fine tuning* sobre un modelo pre entrenado y ejecución local. El cuerpo de conocimiento, experiencias y metodologías desarrolladas se transferirán a cátedras afines y a otras de la unidad académica.

**Keywords:** Massive University Assessments, Numerical Calculation, Artificial Intelligence.

**Abstract.** The opportunities presented by the Artificial Intelligence in massive university educational contexts and as an assistant to teaching assessments motivate the development of a research project within the Computing and Numerical Calculation Cathedra. The second phase of the project try to answer the question: Are the Artificial Intelligence tools capable of emulating faculty judgment in terms of practical assessments? The test was conducted on a set of 51 practical assessments, each based on solving three engineering problems, using Python 3.x and numerical calculation t. They were graded by two professors and six chatbots using a rubric. The comparative analysis was performed using the Confusion Matrix tool. The next phase is related to fine tuning and knowledge transfer.

## 1 INTRODUCCIÓN

### 1.1 Planteamiento del Problema

La asignatura Computación y Cálculo Numérico (CyCN) pertenece al bloque de ciencias básicas de los nuevos planes de estudio de las carreras de ingeniería en la Facultad de Ciencias Exactas, Físicas y Naturales (FCEFN) de la Universidad Nacional de Córdoba (UNC). El proceso de acreditación de los nuevos planes, con enfoque en la enseñanza de competencias y centrados en el alumno, se concreta al inicio del año 2025 (FCEFN, s.f.).

El diseño curricular de CyCN tiene como objetivo central introducir a la programación estructurada -con lenguaje Python 3.x- y al cálculo numérico a alumnos del segundo año de las carreras de ingeniería. Ellos utilizarán estas herramientas como facilitadoras para el desarrollo de sus actividades en la especialidad, y la asignatura le brinda un primer contacto con las mismas en su trayecto de formación universitaria. En tal sentido los alumnos pertenecen a las siguientes especialidades: Aeroespacial, Agrimensura, Ambiental, Civil, Electromecánica, Industrial, Mecánica y Química. A su vez en las especialidades de Computación, Electrónica y Biomédica se enseñan varias asignaturas relacionadas con la programación sin profundizar en el cálculo numérico.

La organización de las clases presenciales de CyCN fue condicionada por: la nómina prevista de alumnos nuevos (600 anuales) con un aumento de un 30% esperado de recursantes, el plantel docente disponible y los recursos de infraestructura y tecnológicos. Es así que el desarrollo de las mismas se organiza en dos encuentros presenciales semanales de 3 horas cada uno, denominados respectivamente clase teórico-práctica y laboratorio, el primero en aula con pupitres para 150 alumnos, y el segundo en aula con 50 puestos de computadoras.

Durante el primer semestre del año 2025, y según los cursantes reales, se habilitaron 2 comisión teórico-prácticas de 100 alumnos y 1 profesor en cada una, y un total de 4 comisiones de laboratorio con 50 alumnos y 1 profesor en cada una. En el segundo periodo del año, se trabaja con 4 comisiones teórico-prácticas de 100 alumnos y 1 profesor en cada una, y 7 comisiones de laboratorio con 57 alumnos y 1 profesor cada una.

El primer vínculo didáctico entre docente, alumno y contenido se desarrolla en la clase teórico- práctica mediante la propuesta de problemas disparadores del interés y atención del estudiante y ejercicios guiados, posteriormente durante el desarrollo de los laboratorios el profesor, en su rol de tutor, realiza el apoyo y seguimiento personalizado del aprendizaje de cada uno con el fin de atender sus debilidades y potenciar sus fortalezas.

Por lo antes expuesto y según una vasta experiencia desde hace 38 años en el dictado de la vieja asignatura Métodos Numéricos, en la unidad académica y con condiciones de masividad agudizadas desde hace 10 años, el equipo docente se sintió interpelado y motivado para iniciar una búsqueda de soluciones innovadoras con el uso de tecnologías emergentes como la Inteligencia Artificial (IA) y en particular enfocadas en las actividades críticas que desarrolla cada docente en la asignatura. La criticidad se evidencia por la masividad y particularmente durante el proceso de diseño de las evaluaciones, la sustanciación de la actividad evaluativa por parte del alumno y la calificación por parte del docente.

En particular el proceso evaluativo sumativo se realiza a lo largo de toda la cursada, a través de evaluaciones conceptuales parciales, un total de 9, y una evaluación practica final.

Por los motivos expuestos, surge entre algunos miembros del cuerpo docente y alumnos practicantes de pregrado la idea de presentar un proyecto de investigación, actualmente en curso, ante la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Córdoba, radicado en el Laboratorio de Investigación y Desarrollo de Software e Inteligencia Artificial (LIDeSIA) dependiente del Departamento de Computación (DC) de la FCEFN.

El antecedente clave que guía la idea anterior ha sido el trabajo de [Weber y Fernández](#)

(2021), integrantes del equipo de investigación, referido a las experiencias didácticas relacionadas con la implementación de la evaluación automática de las actividades prácticas de la asignatura Cálculo Avanzado mediante el desarrollo de un código específico en lenguaje Python, en la carrera de Ingeniería Civil de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional.

Si bien el objeto de estudio es la Cátedra de CyCN, en esta instancia está focalizado en su proceso de evaluación y en la capacidad que brinda la inteligencia artificial para su soporte. Desde una perspectiva multidimensional, se integran, al tema de investigación, marcos conceptuales desarrollados por referentes de las corrientes didácticas contemporáneas (Anijovich y Cappelletti, 2017; Camilloni, 1998; Litwin, 2013; Maggio, 2013) y del área disciplinar del cálculo computacional (Weber, 2023; Kong et al., 2020).

## 1.2 Preguntas de investigación

El tema de investigación que se aborda, desde una perspectiva general queda inserto en el uso de las TICs para la gestión de masividad en la enseñanza universitaria. Enmarcado en esta temática, el problema de investigación se centra en determinar si la IA tiene la capacidad de ser efectiva como soporte en la enseñanza universitaria, en condiciones de masividad, para las carreras de ingeniería. La búsqueda de una solución, implica investigar, experimentar y evaluar la eficacia de las herramientas disponibles de IA para la mejora del proceso de enseñanza en su instancia evaluativa.

Al interior de la Cátedra de CyCN, como objeto investigativo, y según el problema expuesto, nos planteamos las siguientes preguntas de investigación: ¿cómo puede la IA ser utilizada de manera efectiva para respaldar y mejorar el proceso de evaluación universitaria, en entornos de masividad, en las carreras de ingeniería? ¿Son las herramientas de IA capaces de emular el criterio del cuerpo docente en términos de las evaluaciones? ¿hasta dónde es posible entrenar una IA para que simule dicho criterio? ¿Cuál es el nivel máximo de complejidad que puede enfrentar la IA para una evaluación?

## 1.3 Hipótesis

Según el problema a investigar y las preguntas asociadas, se plantearon las siguientes hipótesis, sujetas a pruebas y análisis, durante el proceso investigativo, para determinar su validez y veracidad.

- La IA puede ser utilizada de manera efectiva como soporte a la enseñanza universitaria, en condiciones de masividad, para las carreras de ingeniería y particularmente en el proceso de evaluación.
- Las herramientas de IA tienen la capacidad de emular el criterio del cuerpo docente en términos de evaluaciones en el ámbito universitario.
- Es posible entrenar una IA de manera de simular el criterio de evaluación de un docente.
- La IA puede enfrentar niveles altos de complejidad en la evaluación de estudiantes en carreras de ingeniería.

## 1.4 Objetivos de la Investigación

Con el fin de dar respuesta a las preguntas antes explicitadas y validar las hipótesis planteadas, el objetivo general de este proyecto de investigación es explorar y experimentar hasta donde es posible entrenar la IA para generar y corregir evaluaciones masivas universitarias, en las ingenierías y como asistente al cuerpo docente.

Los objetivos específicos se refieren a:

- Analizar las distintas herramientas de IA y seleccionar la que mejor se adapte al objetivo de

este proyecto.

- Entrenar y validar la herramienta de IA seleccionada.
- Implementar evaluaciones masivas de corrección automática, generadas por IA, en el aula virtual de la cátedra de Computación y Cálculo Numérico.

### 1.5 Justificación del Trabajo

En consecuencia, el equipo de investigación está desarrollando el proyecto enfocado en posibles herramientas para la implementación de instrumentos de evaluación que se utilizaran en la asignatura, y en directa relación con las competencias a enseñar y las condiciones áulicas expuestas anteriormente. Durante el año 2024 se evaluaron diferentes IA conversacionales de acceso en línea y gratuitas, algunas presentaron suficiente efectividad de tal manera que se utilizan actualmente para generar las preguntas conceptuales de la materia, previa supervisión docente, con formato adecuado para su importación directa al aula virtual.

Se presentan en este trabajo los resultados obtenidos, durante el primer semestre del año 2025, los cuales nos habilitan a continuar durante el resto de año con la implementación de la etapa final cuyo objetivo es mejorar las prestaciones mediante ajuste fino o programación de una IA, de uso local y gratuito, con material específico de la cátedra de CyCN.

## 2 MARCO TEÓRICO

### 2.1 Evaluación Formativa con Enfoque en Competencias

El concepto de competencia, nos remite a dos referentes del ámbito educativo, uno de ellos es [Sergio Tobón \(2007\)](#) quien las define como procesos complejos de desempeño frente a problemas que se enfrentan con idoneidad y ética. La formación en competencias se focaliza, según el autor, en el desarrollo y fortalecimiento de habilidades de pensamiento complejo, formando personas emprendedoras y competentes. Por otro lado [Philippe Perrenoud \(2008\)](#) habla del dilema entre competencia y saber, discutido por la comunidad educativa, y concluye que el enfoque por competencias en la enseñanza, lejos de dar la espalda a los saberes les da una nueva fuerza vinculándolos a las prácticas sociales, a las situaciones complejas, a los problemas y a los proyectos

Una conceptualización más generalizada de competencia la aportan [Cerato y Gallino \(2013\)](#) quienes la consideran como saber hacer en contexto, lo cual involucra conocimientos, habilidades, destrezas y actitudes que se integran a las características personales como capacidades, rasgos, motivaciones, valores y experiencias.

[Anijovich y Cappelletti \(2017\)](#), consideran a la evaluación como un oportunidad y reflexionan sobre la importancia, en la evaluación formativa, de establecer una instancia de diálogo a través de la retroalimentación de resultados de tal manera de promover la autonomía en el aprendizaje y la concientización del propio proceso por parte del estudiante. Atendiendo a lo expuesto proponen como instrumento de asistencia a la evaluación el uso de rúbricas y matrices de valoración.

Diseñar una rúbrica implica el intento de evaluar de la manera más objetiva posible el trabajo y avance de los estudiantes, por lo que el profesor deberá especificar claramente qué espera de sus alumnos, mencionando los criterios bajo los cuales se evaluará su desempeño en relación a la tarea o actividad encomendada, el número de niveles de la rúbrica y los valores de desempeño asignados a cada nivel ([FCEFN, 2019](#)).

## 2.2 Los Chatbot

La Inteligencia Artificial es el campo de la computación que trabaja con sistemas que pueden emular procesos de razonamiento, simular procesos de aprendizaje a partir de los datos de una manera muy particular, y pueden actuar con distintos grados de autonomía (Russell y Norvig, 2010), pero no lo hacen como el ser humano. Entre las aplicaciones diversas de la IA, se encuentran los sistemas conversacionales o *chatbots*, que son programas o plataformas que interactúan y se comunican con el usuario en forma similar a una conversación humana (Pérez et al., 2020). Por otro lado, los Modelos de Lenguaje Generativo (GLM) son entrenados para crear textos con coherencia, y se utilizan, por ejemplo, para responder a preguntas realizadas por el usuario. Existen sistemas de IA que son simultáneamente GLM y *chatbots*, entre ellos se encuentra ChatGPT (Chat Generative Pre-trained Transformer) desarrollado por OpenAI.

En el ámbito de la educación, el uso de los *chatbots* ha estado en constante desarrollo en los últimos años, aunque su uso empezó a popularizarse en la última década. Se han utilizado como asistentes de aprendizaje y tutores virtuales, proporcionando ayuda y respuesta inmediata al estudiante. De esta manera, el aprendizaje será más interactivo, pudiéndose reforzar conceptos y aclarar dudas en tiempo real (Arias-Chávez et al., 2024).

## 3 METODOLOGÍA

### 3.1 Tipo de Investigación

El proyecto de investigación se encuadra en el cruce entre las TICs, la Tecnología Educativa y la Didáctica, por ello se desarrolla integrando las disciplinas con foco en la IA. El proceso investigativo, atendiendo a su naturaleza compleja, se implementa desde una perspectiva mixta, es decir cualitativa y cuantitativa, con un diseño secuencial de dos etapas y alcance descriptivo.

En particular, en la etapa presentada, del tipo experimental cuantitativa se realiza un análisis comparativo entre los criterios del docente y varios *chatbots* en la calificación de las evaluaciones prácticas de la asignatura CyCN.

### 3.2 Instrumentos de Recolección de Datos

La fuente de datos se obtiene de manera confidencial y anónima, de tres instancias evaluativas prácticas de la misma asignatura dictada por el responsable de CyCN de la UNC, en la Universidad Tecnológica Nacional, bajo el nombre de Cálculo Avanzado y en su mismo rol. Esto fue motivado por la falta de datos históricos en la nueva asignatura CyCN, de reciente creación. La cantidad de alumnos cursantes es de 51 estudiantes, y las instancias evaluativas se corresponden con: evaluación práctica durante la cursada, recuperación de evaluación práctica durante la cursada, evaluación práctica en fecha de examen final.

Cada evaluación consta de tres problemas a resolver con códigos en lenguaje Python 3.x. Se utiliza como instrumento de evaluación una rúbrica. Los problemas 1 y 3 se evalúan con una rúbrica basada en 3 criterios y 2 niveles de desempeño cada uno (Tabla 1), y el segundo problema utiliza una rúbrica de 4 criterios y dos niveles de desempeño cada uno (Tabla 2). El desempeño de cada problema es incremental: solo se evalúa un criterio si el anterior fue aprobado. En la Tabla 3 se muestran los niveles de desempeño totales factibles para cada problema.

Criterio de evaluación/Nivel de desempeño	0	10
¿Tiene una idea general de la solución?	Reprobado	Aprobado
¿El código se puede ejecutar correctamente?	Reprobado	Aprobado
¿Los resultados son correctos?	Reprobado	Aprobado

Tabla 1: Rúbrica correspondiente a los problemas 1 y 3.

Criterio de evaluación/Nivel de desempeño	0	10
¿Tiene una idea general de la solución?	Reprobado	Aprobado
¿El código está libre de errores sintácticos que impidan su ejecución?	Reprobado	Aprobado
¿El código está libre de errores lógicos que impidan su ejecución?	Reprobado	Aprobado
¿Los resultados son correctos?	Reprobado	Aprobado

Tabla 2: Rúbrica correspondiente al problema 2.

Nivel de desempeño total					
Problema 1	0	10	20	30	NA
Problema 2	0	10	20	30	40
Problema 3	0	10	20	30	NA

Tabla 3: Niveles de desempeño de cada problema

Mediante una matriz de múltiples entradas se registraron los datos reales de las evaluaciones del docente, con el siguiente detalle: la instancia evaluativa, en alumno anónimo identificado por números correlativos, el número de problema, el criterio de la rúbrica de evaluación, el valor del desempeño obtenido en cada criterio, el desempeño total de cada problema y el total del alumno anónimo en la instancia evaluativa.

El proceso de corrección de las evaluaciones por los *chatbots* seleccionados en etapas anteriores del proyecto, según criterios de gratuidad, disponibilidad en línea y efectividad se realizó mediante un mismo *prompt* descriptivo del rol a cumplir, la tarea encomendada, las rúbricas de evaluación y los códigos entregados por el alumno, todo compilado en un archivo de formato pdf, de tal manera que el mismo archivo es el que alimento a cada *chatbot*.

Los *chatbots* que se utilizaron son: Chat GPT (modelo 3.5- 4<sup>o</sup>), Claude (Sonnet 4), Deepseek, Qwen (2.5 max) y Gemini (2.5 flash). Se incorporó a la lista Perplexity, catalogado como un motor de búsqueda compleja que integra varios modelos conversacionales de IA como GPT-4 Omni y Claude 3.5, con el objetivo de determinar si su integración evidencia un desempeño similar o no a los *chatbots* seleccionados.

### 3.3 Población y Muestra

Es importante distinguir entre la cantidad de alumnos cursantes (51) y el tamaño de población y de la muestra analizadas. La población, para este caso, es la cantidad de problemas resueltos en las tres instancias evaluativas, es decir que cada elemento de la población es un problema resuelto por algún alumno y de acuerdo a la información recolectada el tamaño de la población es de 156 elementos.

Ya que eventualmente algún alumno no entregó algún problema, esto indica que el tamaño de la muestra (127 elementos) difiere del tamaño de la población. Pero desde el punto de vista académico la falta de una entrega implica un desempeño nulo por tal motivo se trabajó con análisis estadístico de la población.

### 3.4 Herramienta de Análisis de Datos

El proceso exploratorio para hallar una herramienta estadística adecuada, con el fin de analizar los datos recolectados, nos remite, según varios autores (James et al., 2023; Kuhn y Johnson, 2013), a desistir de las métricas estadísticas que miden solo aciertos o desaciertos en el contexto particular educativo y de predicciones con IA.

Los autores recomiendan el uso de una Matriz de Confusión (MC) y a modo de ejemplo en un análisis binario o donde se presentan únicamente dos clases -en este caso 0/10 o Aprobado/Reprobado-, la herramienta permite discernir los siguientes parámetros estadísticos: Verdaderos Positivos (VP), Verdaderos Negativos (VN), Falsos Positivos (FP) y Falsos Negativos (FN). A través de ellos se calculan los indicadores presentados en las siguientes ecuaciones (1, 2, 3, 4) los cuales sirven para medir que tan bien clasifica un *chatbot* o modelo comparado con un profesor.

$$\text{Precisión} = \frac{VP}{VP+FP} \quad (1)$$

$$\text{Exhaustividad} = \frac{VP}{VP+FN} \quad (2)$$

$$\text{Puntuación F1} = 2 \times \frac{\text{Precisión} \times \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}} \quad (3)$$

$$\text{Especificidad} = \frac{VN}{VN+FP} \quad (4)$$

Una interpretación conceptual de las ecuaciones permite establecer que: la precisión indica cuan confiable es el modelo en las predicciones positivas o aprobados, la exhaustividad mide la proporción de casos positivos reales que el modelo detecta correctamente, la puntuación da una idea del equilibrio entre falsos positivos y falsos negativos, y la especificidad establece la proporción de casos negativos reales que el modelo reconoce correctamente como negativos. Adicionalmente, se agrega el cálculo de un coeficiente de concordancia que evalúa el grado de acuerdo entre dos evaluadores, para este caso docente y modelo (Kappa de Cohen) corrigiendo el efecto del azar.

## 4 RESULTADOS Y DISCUSIÓN

### 4.1 Presentación de los Resultados

En este caso en particular, se calcularon los indicadores de la MC y Coeficiente de Kappa de Cohen mediante clases múltiples. Los Problemas 1 y 3 son de 4 clases según los desempeños factibles: 0, 10, 20, 30. El Problema 3 es de 5 clases y sus desempeños posibles son 0,10,20,30,40. Luego de un arduo trabajo realizado en planillas vinculadas para el cálculo de todos los indicadores para cada modelo de IA, mediante un enfoque “uno contra el resto”, de tal manera de analizar cada clase individual como si se tratara de un problema de clasificación binaria, se obtuvieron los indicadores macro de cada *chatbot* versus el criterio docente (Tabla 4).

<i>Chatbot</i>	Precisión	Exhaustividad	Puntuación F1	Especificidad	Kappa de Cohen
CHAT GPT	0,313	0,265	0,163	0,826	0,2372
CLAUDE	0,369	0,339	0,171	0,849	0,3300
PERPLEXITY	0,441	0,255	0,158	0,834	0,2731
DEEP SEEK	0,392	0,213	0,118	0,808	0,1544
QWEN2.5-MAX	0,232	0,112	0,081	0,781	0,0576
GEMINI 2.5 FLASH	0,433	0,177	0,129	0,824	0,2219

Tabla 4: Indicadores macro de cada *chatbot* versus el criterio docente

## 4.2 Limitaciones de la Investigación

Durante el proceso se encontraron algunas dificultades externas. Dentro de las mismas se menciona la limitación de cantidad de *prompts*, cada determinado tiempo, que se le podían solicitar de las IA's por tratarse de versiones gratuitas, como consecuencia el análisis resulta interrumpido periódicamente. Por otro lado, particularmente la IA Gemini 2.5 Flash ha evidenciado, en algunas ocasiones, dificultad para recibir y leer los archivos de formato pdf, por errores en el respectivo sistema.

Otro escenario al cual se enfrentó la investigación consiste en las constantes actualizaciones en las IA. Como consecuencia, resulta difícil determinar cuál resulta más idónea para desarrollar las pruebas. Ante ello es que se decide abordar un número significativo de IA's para trabajar. El criterio que se tuvo para seleccionarlas no fue único, ya que se analizó disponibilidad de uso en el territorio nacional, dinamismo en las interacciones preguntas-respuestas y gratuidad. A pesar de lo enunciado anteriormente, se finalizó con la elaboración y recolección de los datos necesarios para el análisis.

## 4.3 Interpretación de los Resultados

La [Tabla 4](#), a través de los indicadores calculados para cada IA y su interpretación permite inferir que, a la fecha de finalización de la medición y cálculo:

- El mejor modelo global es Claude, evidencia balance entre precisión y exhaustividad, posee la mejor especificidad y el mejor Kappa, aunque no tiene la precisión más alta.
- Los modelos más precisos son Perplexity y Gemini Flash, indicando que son buenos cuando lo que se desea es minimizar los falsos positivos.
- El modelo más débil es Qwen con métricas consistentemente más bajas que el resto.
- Un orden de desempeño global enumerados del mejor al peor es: Claude – Perplexity y Gemini Flash – Chat GPT – Deep Seek – Qwen.

## 5 CONCLUSIONES

Los resultados obtenidos nos permiten establecer conclusiones parciales y contestar una de las preguntas de investigación: según el estado del arte de los *chatbots* gratuitos y de ejecución remota, a la fecha de confección de este documento, la IA no tiene suficiente capacidad para emular el criterio del cuerpo docente en términos de evaluaciones prácticas de CyCN.

La evolución del proyecto nos remite a la próxima etapa de programación o entrenamiento de un modelo de IA gratuito en forma local y mediante contenidos, problemas y resoluciones particulares de la cátedra, además de la base de datos de las evaluaciones reales sustanciadas durante la cursada inaugural del año 2025. La finalidad es validar o no la siguiente hipótesis: Es posible entrenar una IA de manera de simular el criterio de evaluación de un docente.

Con la experiencia que ha adquirido a la fecha el equipo de trabajo, además de las



consideraciones éticas sobre el uso de estas herramientas en ambientes educativos, y los resultados parciales obtenidos nos permite establecer que este tipo de evaluaciones tienen un rol complementario y solo como asistente para el propio docente en su actividad.

## REFERENCIAS

- Anijovich, R. y Cappelletti, G., *La evaluación como oportunidad*. Paidós, 2017.
- Arias-Chávez, D., Ramos-Quispe, T., y Cangalaya Sevillano, L., Análisis y tendencias en el uso de chatbots y agentes conversacionales en el campo de la educación: una revisión bibliométrica. *Innovaciones Educativas*, 26(41), 242–260, 2024.
- Camilloni, A., *La Evaluación de los Aprendizajes en el Debate Didáctico Contemporáneo*. Paidós, 1998.
- Cerato, A. y Gallino, M., Competencias genéricas en carreras de ingeniería. *Revista Ciencia y Tecnología*, 13, 83-94, 2013.
- FCEfYN (s.f.) *Programa de la materia Computación y Cálculo Numérico*, <https://fcefyn.unc.edu.ar/facultad/secretarias/academica/departamentos/computacion/departamento-de-computacion/10-04072-computacion-y-calculo-numerico/> , accedido 1 de septiembre de 2025.
- FCEfYN, *Manual Descriptivo de Instrumentos de Evaluación bajo el Modelo de Competencias*, Curso la Evaluación de los Aprendizajes en las Ciencias y las Tecnologías, 2019.
- Kong, Q., Siau, T., y Bayen, A., *Python Programming and Numerical Methods: A Guide for Engineers and Scientists*. Elsevier Science, 2020.
- James, Gareth et al., *An Introduction to Statistical Learning: With Applications in Python*. Springer, 2023.
- Kuhn, M. y Johnson, K., *Applied Predictive Modeling*. Springer, 2013.
- Litwin, E., *El oficio de enseñar. Condiciones y contextos*. (7ma. ed.). Paidós, 2013.
- Maggio, M., *Enriquecer la enseñanza. Los ambientes de alta disposición tecnológica como oportunidad*. Paidós, 2013.
- Pérez, J., Daradoumis, T., y Puig, J., Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6), 1549-1565, 2020.
- Perrenoud, P., Construir las competencias, ¿es darle la espalda a los saberes? *Red U. Revista de Docencia Universitaria*, número monográfico II “Formación centrada en competencias (II)”, 2008.
- Russell, S., y Norvig, P., *Artificial Intelligence: A Modern Approach*. (3rd ed.). Pearson Education, 2010.
- Tobón, S., El enfoque complejo de las competencias y el diseño curricular por ciclos propedéuticos. *Revista Acción Pedagógica*, 16(1): 14-28, 2007.
- Weber, J., y Fernández, C., Sobre la Automatización de las Evaluaciones Prácticas de la Asignatura Cálculo Avanzado. *Mecánica Computacional*, 38(18), 733-742, 2021.
- Weber, J. F., *Computación y Cálculo Numérico*. Universitas, 2023.