



UNIVERSIDAD DE COSTA RICA



CIENCIA MÁS TECNOLOGÍA

Tecnología desarrollada en la UCR

# La inteligencia artificial y NATURAL detrás de las pruebas adaptativas de dominio del inglés

El sistema puede determinar con exactitud el nivel de una persona con solo 15 preguntas

6 MAY 2024

Ciencia y Tecnología



Gracias a la plataforma desarrollada en la UCR, más de 7 000 estudiantes de primer ingreso realizaron el Examen Diagnóstico de Inglés (EDI) en un promedio de 8 minutos y obtuvieron su resultado de manera inmediata. Foto con fines ilustrativos.

Foto: [Karla Richmond](#).

Los exámenes de dos o más horas frente a una computadora para determinar el nivel de inglés de una persona quedaron en el pasado. Y no solo eso, también las conversaciones cara a cara con un evaluador para medir la producción oral en otro idioma.

Ahora, con una prueba de pocos minutos se puede determinar el nivel de inglés en las habilidades de comprensión de lectura, comprensión auditiva y producción oral de acuerdo con los rangos del Marco Común Europeo de Referencia de las Lenguas (MCERL).

Esto, gracias a un trabajo conjunto entre las escuelas de Lenguas Modernas y de Ciencias de la Computación e Informática de la Universidad de Costa Rica, las cuales desarrollaron una plataforma modular escalable capaz de hacer aplicaciones masivas y que brinda los resultados de forma inmediata, junto con una serie de recomendaciones personalizadas que llegan al correo electrónico de cada aplicante, entre muchas otras facilidades.

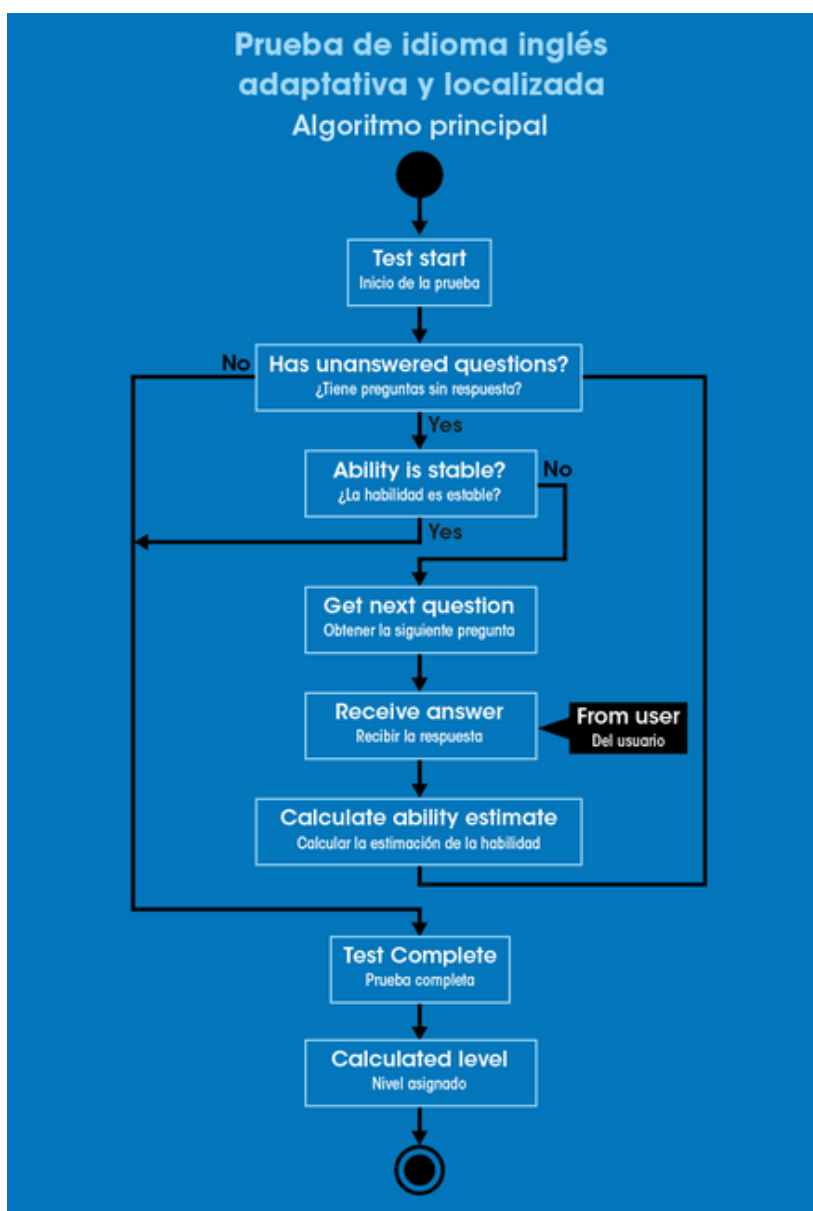
La inteligencia artificial que utiliza este sistema es el fruto de la investigación y conceptualización de expertos evaluadores quienes, a lo largo de los años, han creado y valorado ítems para pruebas sumativas. En el proceso también han participado miles de personas que han contestado esos ítems, así como lingüistas, estadísticos, profesionales del idioma e informáticos, entre otros.

El sistema mantiene su estabilidad independientemente del número de personas que hace el examen, gracias a su diseño por capas. La primera de ellas alberga todos los datos, desde la información de los aplicantes, los ítems y todas las respuestas brindadas, hasta la secuencia de aciertos y desaciertos de cada individuo y el tiempo que dedicó a cada ejercicio. Esta información se maneja en un servidor exclusivo bajo estrictas medidas de seguridad.

La segunda capa consiste en la lógica de las aplicaciones, la cual se encarga de formular las preguntas en función de las mismas respuestas que da el usuario. Estos ajustes instantáneos hacen que cada prueba sea única y son los que determinan el nivel de dominio del idioma que tiene la persona, rápidamente y con exactitud.

La tercera capa es lo que el usuario ve en la pantalla de su computador y que, en lenguaje informático, se conoce como *frontend*. Esta parte también fue diseñada modularmente para poder colocar y quitar elementos de acuerdo con las necesidades de la evaluación. Por ejemplo, el Examen Diagnóstico de Inglés (EDI) que se aplicó al estudiantado de primer ingreso a la Universidad incluyó un módulo de producción oral desarrollado por la empresa Speechace en conjunto con la Universidad de Costa Rica.

“Este diseño flexible permite que podamos instalarlo en un computador local y llevarlo a una zona rural y fácilmente adaptarlo para correrlo, que desde una red local se conecte y, en un servidor, llevar una copia del sistema y hacer una aplicación local. O bien, de forma remota, desde aquí, aplicar en Panamá o en Colombia, como ya se ha hecho. Entonces, podemos dar un servicio fuera de Costa Rica”, resaltó Edgar Casasola Murillo, docente de la Escuela de Ciencias de la Computación e Informática.



# Cuando la matemática, la estadística, la lingüística y la informática se dan la mano

Una parte fundamental de todo este proceso fue seleccionar el modelo matemático más consistente para clasificar a las personas en un rango del MCERL. Según explica Alejandro Fallas Godínez, docente de la Facultad de Letras y estadístico del Programa de Evaluación en Lenguas Extranjeras (Pelex), el equipo desarrollador se inclinó por el modelo de Rasch porque su simplicidad evita que las personas queden ubicadas en los bordes de los rangos. Es decir, se elimina la posibilidad de alguien con un nivel avanzado del idioma obtenga una puntuación tan cerca del límite inferior que lo haga caer en un nivel intermedio y viceversa.

Otro gran detalle es la elaboración de los ítems, los cuales deben adaptarse al MCERL y a la realidad nacional. Además, deben respetar una serie de lineamientos de evaluación y tener una estructura específica. Por ejemplo, no deben contener dobles negaciones o generar confusión en la persona que toma la prueba.

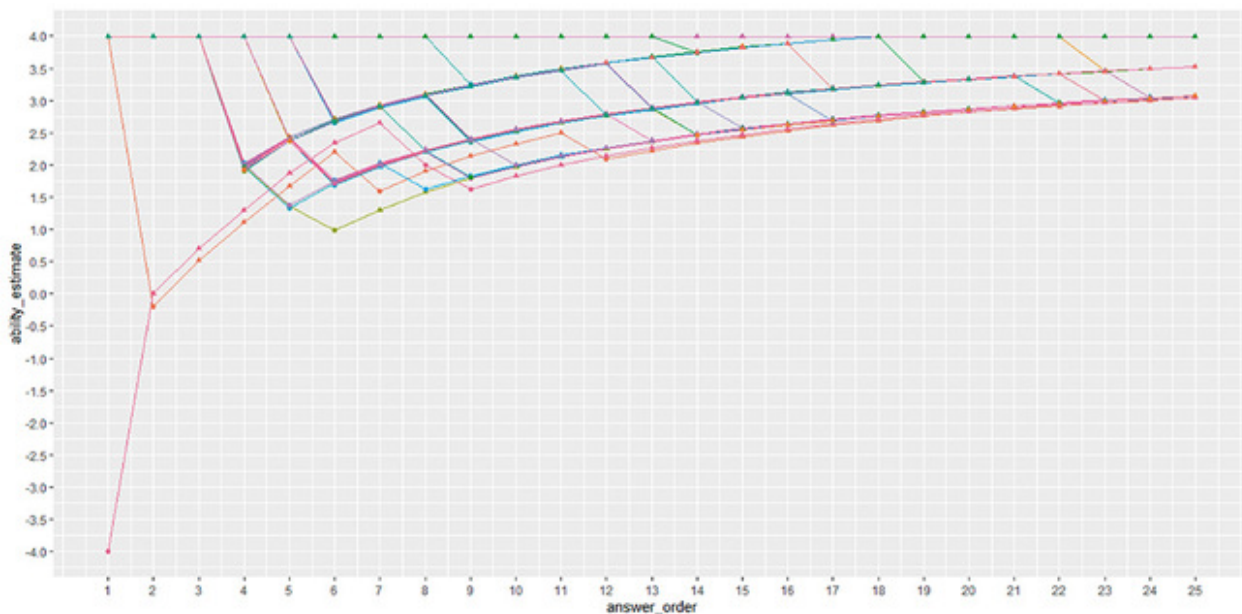
Posteriormente, los ítems deben someterse a pruebas con una población lo más parecida posible a las personas que harán el examen para corroborar la consistencia de cada uno de ellos. Esta población de prueba debe incluir personas con distintos niveles de dominio del idioma para tener certeza de que el ítem discrimina quién puede contestarlo y quién no, a partir del conocimiento que tenga de la persona por las respuestas que ha brindado anteriormente. Una vez verificada la calidad de un ítem, se pasa al banco de ítems y se implementa en el modelo para que funcione en el sistema.

Es así como la inteligencia artificial y los sistemas automatizados que utilizan las pruebas adaptativas de dominio del inglés desarrolladas en la UCR se sustentan en el trabajo permanente de muchas personas, porque, como señala Fallas, no se trata de pedirle a una máquina que invente una prueba.

“Yo le puedo pedir al ChatGPT que genere nuevos ítems a partir de un modelo que yo le doy, pero eso no garantiza que los nuevos ítems tengan la misma calidad. Es necesario respetar todo el proceso de revisión con especialistas y hacer pruebas con personas reales con el propósito de ver el desempeño de cada ítem en un ambiente real”, acotó el investigador.

A partir de esta plataforma, el equipo investigador está desarrollando pruebas específicas y experimentando con otras técnicas de inteligencia artificial para evaluar otras habilidades, como la dicción y la redacción en inglés. Paralelamente, está trabajando en la prueba adaptativa de dominio del francés, la cual está en su fase final.

Alejandro Mairena Jiménez, estudiante de maestría de la Escuela de Ciencias de la Computación e Informática, tuvo a su cargo la depuración del software y la alimentación de todo el sistema. Él destaca las ventajas administrativas de la plataforma, la cual permite que los docentes monitoreen el progreso de sus estudiantes, revisar las respuestas y descargarlas. Además, la herramienta puede hacer reportes, generar documentos en PDF y enviar los resultados por correo electrónico.



La imagen muestra a todas las personas que fueron ubicadas en el rango C1 (avanzado) en comprensión de lectura. Cada línea de color representa a una persona. El eje X muestra el orden de las respuestas y el eje Y la medición de la habilidad. Aunque unos iniciaron la prueba un poco abajo y otros muy arriba, todos terminaron convergiendo en el mismo nivel.

*“Estamos apostando responsablemente a una inteligencia artificial supervisada, lo cual conlleva un proceso extra. Dejar que solo la máquina evalúe a los estudiantes, para nosotros no es responsable. [...] Cada vez que hacemos un examen, un porcentaje de esa población es escuchada por nuestros evaluadores para seguir mejorando los algoritmos.”*

Allen Quesada Pacheco, director de la Escuela de  
Lenguas Modernas

## Un evaluador de la oralidad más amigable

En cuanto al módulo de producción oral con un avatar, Allen Quesada Pacheco, director de la Escuela de Lenguas Modernas, detalló la colaboración estrecha que ha habido entre la empresa de inteligencia artificial Speechace y la UCR para mejorar los algoritmos y las tareas que enfrentan los usuarios. Gracias a esa alianza se creó una interfase más amigable



para niños y jóvenes, quienes pueden escoger el avatar con el que interactuarán, entre otras características.

Esta parte de la prueba combina varios métodos de algoritmos como el Reconocimiento Automático del Habla (RAH), el Procesamiento del Lenguaje Natural (NLP), el aprendizaje automático (machine learning) y el aprendizaje profundo (deep learning). Pero es la calibración realizada por el equipo experto de la UCR la que permite eliminar sesgos por acentos no nativos del inglés. Esto es de fundamental importancia para quienes han aprendido el inglés como segunda lengua.

“Lo que se da es un monólogo. El usuario se enfrenta a un avatar y mantiene un intercambio con él a través de una presentación. Con esto se miden las competencias de la producción, la pronunciación, la fluidez, la gramática y el vocabulario específico para esa pregunta. Inclusive, se está perfeccionando un algoritmo que medirá la pertinencia de lo que habla la persona”, puntualizó el director de la Escuela de Lenguas Modernas.

Este evaluador virtual tiene la ventaja de que mide con exactitud cada uno de los componentes de la producción oral de miles de personas sin necesidad de estar tomando apuntes en una libreta, como lo hacen los evaluadores humanos, lo cual puede estresar al aplicante. Además, brinda los resultados y una serie de consejos para mejorar la producción oral de manera inmediata, sin cansarse, algo imposible para un evaluador de carne y hueso.

Para Quesada, esta tecnología es revolucionaria porque se pasa de una evaluación del aprendizaje a una evaluación para el aprendizaje, porque el sistema no solo ubica a la persona en un rango de dominio del idioma, sino que también le hace una devolución de los aspectos que puede mejorar.

Gracias a su diseño flexible, esta plataforma puede utilizarse para evaluar otro tipo de conocimientos, como la matemática.



**[Fernando Montero Bolaños](#)**  
Periodista, Oficina de Comunicación Institucional  
[fernando.monterobolanos@ucr.ac.cr](mailto:fernando.monterobolanos@ucr.ac.cr)

**Etiquetas:** [ingles](#), [evaluacion](#), [ia](#), [inteligencia artificial](#), [pruebas adaptativas](#), [c+t](#).