



UNIVERSIDAD
COMPLUTENSE
MADRID

Proyecto de Innovación

Convocatoria 2024/2025

Nº de proyecto

447

Evaluación de la amenaza de la inteligencia artificial sobre algunas actividades
de la evaluación continua

Responsable del Proyecto

JOSÉ FERNÁNDEZ MENÉNDEZ

Facultad de CC. Económicas y Empresariales

Departamento

Organización de Empresas

1. Objetivos propuestos en la presentación del proyecto

Una de las consecuencias de la aplicación del Plan Bolonia fue el énfasis puesto en la evaluación continua para los nuevos grados implantados. Este aumento en el peso dado a la evaluación continua se justifica en el hecho de que el aprendizaje es también un proceso continuo y que no parece razonable que la calificación que obtenga el alumno al final de todo este proceso se base exclusivamente en su nivel de desempeño en un punto aislado de ese proceso, el momento de realización del examen final.

Como consecuencia de ello, los planes de estudio de los distintos grados establecieron que porcentajes considerables de la nota final del alumno derivarían de trabajos y actividades realizados por él, individualmente o en grupo, fuera del aula y de forma más o menos continua a lo largo del curso.

A la hora de evaluar estos trabajos realizados fuera del aula, para el profesor es virtualmente imposible determinar quién ha realizado el trabajo. Siempre ha existido una picaresca, probablemente agravada en los últimos años, de TFGs y TFMs, e incluso tesis doctorales, cuya realización se “subcontrata” a academias, profesionales, etc., pero, hasta recientemente, los costes que esta práctica suponen para el alumno no la convierten en una opción razonable para la realización de los trabajos de la evaluación continua.

Esta situación ha cambiado de forma drástica recientemente con la generalización y la facilidad de uso de los *Large Language Models* (LLM) basados en Inteligencia Artificial (AI), como es el caso del archiconocido ChatGPT, y de toda una plétora de modelos similares que no paran de surgir. Estos modelos han supuesto que se “democratice” la subcontratación de los trabajos para realizar fuera del aula. Se vuelve tan sencillo y tan cómodo (y esencialmente sin coste) pedirle los trabajos a ChatGPT, que realmente todos los incentivos van en la dirección de que así ocurra. Los profesores que participan en este proyecto han sido testigos involuntarios de conversaciones entre alumnos en las que se informaban entre ellos de haber pedido a ChatGPT la realización de los ejercicios evaluables propuestos por sus profesores.

Como consecuencia de ello pierde todo sentido la parte de la evaluación que se basa en trabajos que el alumno deba realizar fuera del aula. Con este Proyecto de Innovación docente lo que se pretende es analizar el impacto que los LLM basados en IA pueden tener y están teniendo sobre las calificaciones que obtengan los alumnos como consecuencia de trabajos que se realicen fueran del aula, y plantear posibles alternativas a estos trabajos que permitan mantener la relevancia y el papel que tiene la evaluación continua, pero garantizando que lo que se está evaluando es realmente la actividad del alumno y su adquisición de las competencias requeridas, y no la sofisticación de un sistema informático, obviamente exceptuando aquellas asignaturas cuyo contenido esté relacionado con las tecnologías de la información y el profesor considere que el uso de la IA sea necesario para demostrar la adquisición de competencias sobre ella.

En muchas memorias de los Grados se indica que la evaluación del estudiante tendrá un carácter continuo y formativo, y podrá realizarse tanto a nivel individual como grupal, en coherencia con la naturaleza de las actividades formativas planteadas. Entre los ítems más utilizados para llevar a cabo dicha evaluación continua se encuentran la

participación activa, los trabajos en equipo y/o individuales e informes, así como las presentaciones en clase. En muchas asignaturas estos ítems correspondientes a la evaluación continua pueden tener un peso que oscila entre el 40 y el 60% de la calificación final. Estas elevadas ponderaciones claramente reflejan la importancia de estas actividades en la calificación final del estudiante que, en numerosas ocasiones, pueden mover la balanza hacia un sentido u otro.

En el último año, una TIC que ha evolucionado con extrema rapidez y con un gran impacto en la enseñanza (entre otros sectores) ha sido la Inteligencia Artificial. En ciertas asignaturas vinculadas con sistemas de información y con tecnología, el aprendizaje del uso de la Inteligencia Artificial, puede ser una diferenciación en el mercado laboral con respecto a otras personas. Sin embargo, la aparición de la Inteligencia Artificial ha propiciado la aparición de algunas prácticas fraudulentas, bastante difíciles de detectar por el profesor, en muchos de los ítems de evaluación continua que cuestionan la autoría del estudiante.

Estas prácticas fraudulentas también se pueden hacer extensibles a los Trabajos Fin de Grado (TFG) o incluso los Trabajos Fin de Máster (TFM). En ese caso, el problema con el que se encuentra el tutor de TFG o de TFM es poder comprobar la autoría del trabajo realizado por el estudiante. Un problema similar se observó en los años previos a la popularización del uso de herramientas de IA, cuando comenzaron a proliferar páginas web donde personas y empresas ofrecían sus servicios para la superación de estos trabajos. Este problema ya fue en el Proyecto de Innovación 161 de la convocatoria 2018-2019 por algunos de los autores de este Proyecto de Innovación.

Al igual que ocurre en otros aspectos de la existencia, la Inteligencia Artificial está evolucionando con extrema rapidez. Nadie duda de los beneficios que nos proporciona la Inteligencia Artificial en el ámbito docente, pero claramente se corre el riesgo de que fomente el uso de prácticas fraudulentas en aquellos ítems que conforman la evaluación continua (trabajos individuales y/o en equipo, e informes, presentaciones en clase, etc.) en los que resulte muy difícil identificar la autoría final del trabajo. Al igual que los países necesitan desarrollar legislación relacionada con la Inteligencia Artificial, ¿las universidades no deberían empezar a modificar sus planes de estudio para que el profesor evalúe realmente las capacidades y competencias del estudiante y no las prestaciones que ofrecen los diferentes softwares existentes de Inteligencia Artificial?

Por ello, el propósito de este proyecto de innovación es valorar en qué medida un uso inadecuado de la Inteligencia Artificial influye en los actuales ítems que configuran la evaluación continua del Grado en ADE y del Grado en Comercio de la Universidad Complutense y de la Universidad Rey Juan Carlos.

2. Objetivos alcanzados

Los objetivos del proyecto que se han alcanzado son los siguientes:

- Se ha comprobado que, aunque existe una gran uniformidad en las guías docentes dentro de los grados analizados, no se observa de forma generalizada una información con el suficiente nivel de detalle como para mostrar los distintos ítems de evaluación continua ni su peso dentro de la calificación final. Solamente en el Grado de Comercio proporciona de manera detallada, en la casi totalidad de las asignaturas, el peso que tiene cada uno de los ítems de evaluación dentro de la evaluación continua.
- Se ha comprobado que en las Guías docentes revisadas existen ítems de evaluación que contribuyen con porcentajes significativos a la calificación final y que son susceptibles de ser influenciados, en ocasiones de forma muy acusada, por el uso de la IA.
- Resultan en particular especialmente problemáticos los trabajos en equipo o individuales, que en la mayoría de los casos pueden ser realizados por el alumno de forma prácticamente exclusiva usando herramientas de IA que, aunque se puede sospechar su uso, en la práctica son imposibles de detectar de forma fehaciente.
- Otros ítems de evaluación, como las presentaciones en clase y las resoluciones de ejercicios, también se pueden ver afectados por el uso de la IA, pero probablemente no de forma tan acusada como los trabajos individuales o en equipo (para una presentación en clase se requiere un cierto nivel de familiaridad y conocimiento del tema expuesto, independientemente de quién haya sido el autor de la presentación; en ejercicios poco estandarizados o sobre materias poco estandarizadas, la capacidad de la IA para resolver ejercicios correctamente es baja).
- En cualquier caso, sí que resulta conveniente que el profesor adopte medidas para reducir la posibilidad de uso de IA en estos últimos ítems; medidas razonables serían:
 - llevar a cabo un seguimiento continuado del alumno durante todo el proceso de elaboración de las presentaciones y trabajos, sean individuales o en equipo
 - proponer ejercicios de mayor complejidad y menor estandarización para reducir la posibilidad de que sean solucionados con ayuda de IA
 - Simplificar los ejercicios y trasladar su realización al aula bajo la supervisión directa del profesor
- El problema estriba en que, tanto la elaboración de ejercicios más complejos y menos estandarizados, como especialmente el seguimiento continuado de la realización de trabajos, suponen tal aumento de la carga de trabajo del profesor, si se quiere que sean medidas efectivas, que no es razonable pretender su adopción de forma generalizada, sino tan solo de manera puntual
- Realmente, la única forma efectiva de evitar el fraude mediante el uso de IA en la evaluación continua consiste en que esa evaluación continua se haga mediante el procedimiento tradicional de pruebas (del tipo que sean) por medio de papel y lápiz en el aula bajo la supervisión directa del profesor

3. Metodología empleada en el proyecto

La metodología de trabajo ha consistido en la búsqueda de información por parte de los miembros del equipo y la realización de una serie de reuniones periódicas para poner en común esa información, analizarla y valorarla.

Las fases en las que se ha llevado a cabo son las siguientes:

Etapa 1. Se han recopilado todas las guías docentes del Grado en ADE y Grado en Comercio de la Universidad Complutense de Madrid y Grado en ADE de la Universidad Rey Juan Carlos.

Etapa 2: Se han analizado en detalle, a partir de la información disponible en las Guías docentes recopiladas, las actividades que conforman la evaluación continua en cada una de las asignaturas de los grados seleccionados, así como los ítems de evaluación y sus ponderaciones en la calificación final.

Etapa 3: Se ha tabulado toda la información recopilada para elaborar una base de datos que facilite su análisis.

Etapa 4: Se ha realizado un análisis de los datos recopilados, esencialmente para valorar los porcentajes con que cada ítem de evaluación contribuye a la calificación de la asignatura.

Etapa 5: Se han mantenido múltiples reuniones entre los miembros del equipo para discutir y valorar los resultados obtenidos y se han realizado diversos *brainstormings* para generar y proponer ideas que permitan paliar los problemas que plantea el uso de la IA en los trabajos evaluables no presenciales.

Etapa 6: Se han redactado las principales conclusiones alcanzadas y se han planteado propuestas por parte de todos los miembros del equipo.

Etapa 7: Se han difundido los resultados y las conclusiones alcanzadas en reuniones informales con otros profesores. La falta de financiación ha impedido que, por el momento, no se haya podido alcanzar el objetivo de presentar los resultados en congresos docentes.

4. Recursos humanos

Se trata de un proyecto de Innovación interdepartamental (Departamento de Organización de Empresas y Departamento de Economía Financiera, Actuarial y Estadística de la UCM) en el que han participado los siguientes profesores:

Del Departamento de Organización de Empresas:

- José Fernández Menéndez (responsable del proyecto)
- Emilio Álvarez Suescun
- Diéssica de Oliveira Dias
- Beatriz Minguela Rata

Del Departamento de Economía Financiera, Actuarial y Estadística:

- María Zuleyka Díaz Martínez

El equipo de proyecto está compuesto por integrantes con perfiles personales y profesionales que convergen en una misma línea de adecuación programática en la Economía de la Empresa. Aunque divergen en sus orígenes formativos y en sus propios desarrollos académicos y profesionales, esta diversidad enriquece el desarrollo del proyecto.

Todos los miembros del equipo tienen el grado de Doctor. Tres de ellos son Profesores Titulares de Universidad, otro es Profesor Contratado Doctor y un quinto miembro del equipo es Profesor Ayudante Doctor. Todos ellos tienen una amplia experiencia docente, en particular en el área de la Dirección de Operaciones, Dirección Estratégica, Dirección de la Innovación, y las Matemáticas Empresariales y han participado en un gran número de proyectos de innovación docente así como en la implementación de nuevas asignaturas y nuevas metodologías docentes.

La mayoría de los miembros de equipo de proyecto (el 80% de los mismos) tienen una experiencia docente media superior a los 20 años en la Universidad Complutense de Madrid. Durante todo este tiempo han impartido asignaturas troncales, obligatorias y optativas (de las antiguas licenciaturas) así como básicas, obligatorias y optativas (en los nuevos grados y másteres), obteniendo resultados altamente positivos en las evaluaciones docentes a través del programa Docentia, incluso alguno de ellos ha obtenido el Diploma de Excelencia Docente (entregado por el Rector).

5. Desarrollo de las actividades

El desarrollo del proyecto comenzó con la búsqueda sistemática en las webs de la UCM y la URJC de las Guías Docentes de las asignaturas de los grados en ADE y Comercio, su descarga y su revisión.

A partir de las guías docentes, se cumplimentó una base de datos en la que se recopiló de manera detallada la información relativa a la universidad, el grado, la asignatura, el curso en el que se impartía, el cuatrimestre (primer cuatrimestre o segundo), el número de ECTS correspondiente a la asignatura, el carácter de la misma (básica, obligatoria, optativa), así como los diferentes ítems utilizados para la evaluación del estudiante (con su desglose correspondiente).

Como resultado de esta revisión se identificaron cinco grandes grupos en los que encuadrar los distintos ítems de evaluación:

- Item eval_01: Prueba oral y/o escrita
- Item eval_02: Trabajos equipo y/o individuales
- Item eval_03: Presentación en clase
- Item eval_04: Resolución problemas y ejercicios
- Item eval_05: Participación activa

En cada guía docente se identificaron, en la medida en que estaban detallados, los porcentajes en que cada uno de estos ítems contribuye a la calificación final de la asignatura.

Dentro de estos ítems existen dos que no se ven afectados de forma negativa por el uso de la IA, las pruebas orales o escritas y la participación en el aula, ya que en ellos se necesita la presencia del alumno en el aula.

Las pruebas orales o escritas suponen en promedio un 63% de la calificación final con una desviación típica de 13 puntos de porcentaje. Esa contribución varía entre un mínimo de un 30% y un máximo de un 100%, máximo que se observa en un 2% de las guías.

La participación en el aula supone un 8% de la calificación final en promedio, con una mediana de un 10% y una desviación típica de 3 puntos de porcentaje. Oscila entre un mínimo de un 0% y un máximo de un 20%, alcanzándose cada uno de estos extremos tan sólo en una única guía.

Más problemáticos resultan, por el potencial efecto que pueda tener sobre ellos el uso de la IA, los otros 3 ítems de evaluación, los trabajos en equipo o individuales, las presentaciones en clase y la resolución de problemas o ejercicios.

Los trabajos en equipo o individuales suponen en promedio el 25,7% de la calificación final de las asignaturas contempladas, con una mediana del 20% y una desviación típica de 14,6 puntos de porcentaje. El porcentaje en el que contribuyen estos ítems a la calificación final oscila entre un 2% y un muy elevado 80%. Existe un 8% de guías en

las que la contribución de los trabajos individuales o en equipo a la calificación final es igual o superior al 50%. Este es el ítem más susceptible a la adulteración de resultados mediante el uso de la IA, con lo que dado el peso elevado que tiene en la calificación se deberían extremar las precauciones por parte del profesor a la hora de incluirlo entre los mecanismos de evaluación de los alumnos.

Las presentaciones en clase suponen en promedio un 20% de la calificación final (la mediana también está en el 20% con una desviación típica de 9,6 puntos de porcentaje). Su participación en la calificación final oscila entre un mínimo de un 5% (en una guía), y máximo de un 50% (también en una guía). Aunque el contenido de las presentaciones puede ser elaborado por el alumno recurriendo a la IA, el hecho de que por su propia naturaleza las presentaciones deban ser defendidas públicamente por el alumno ante el profesor proporciona a éste un mecanismo (por ejemplo realizando preguntas concretas sobre el contenido o sobre aspectos metodológicos del trabajo) para valorar la adquisición de competencias por parte del alumno y para mitigar, parcialmente, el efecto que pueda tener la IA en la elaboración de las presentaciones.

La resolución de problemas o ejercicios supone en promedio un 22% de la calificación final (mediana del 20%) con una desviación típica de 11 puntos de porcentaje y oscila entre un mínimo de un 10% y un máximo de un 60% (en un 11% de las Guías supone un 40% o más de la calificación final). Estos son porcentajes relativamente elevados, con lo cual conviene valorar cuidadosamente por parte del profesor el potencial efecto de la IA sobre estos ejercicios. Probablemente en aquéllos con un carácter más sencillo, rutinario y estandarizado, este impacto vaya a ser muy elevado, con lo cual convendría modificar la naturaleza de estos ejercicios o mover su realización al aula.

Estos resultados deben interpretarse con cierta precaución. Si bien nos hubiese gustado realizar un análisis más exhaustivo, nos hemos encontrado que, al revisar las Guías docentes de los 3 grados analizados, un buen número de ellas no detallan en absoluto los tipos de ítems que se utilizan para la evaluación, sino que se indica de manera genérica que la evaluación continua supondrá un determinado porcentaje de la calificación final, sin mayor detalle.

Si bien a nivel de VERIFICA la flexibilidad que esta falta de concreción supone puede ser positiva para facilitar que el profesor organice la evaluación de la forma que estime más adecuada (aunque nos cuestionamos si en el VERIFICA verdaderamente aparece dicha flexibilidad puesto que no hemos podido acceder al mismo para efectuar dicha comprobación), parece que, a nivel de Guía docente, esta discrecionalidad e indefinición (recordemos que en la mayor parte de los casos oscila entre un 40 y un 50% de la calificación final) en los ítems que componen la evaluación continua (y sus ponderaciones) parecen un tanto excesivas. No obstante, esta reflexión obtenida tras el análisis detallado de las guías docentes cae fuera del objeto del presente proyecto.