Facultad de Ciencias Económicas y Empresariales

ICADE

POLÍTICAS INSTITUCIONALES Y SHADOW AI: EVIDENCIA EXPERIMENTAL SOBRE EL EFECTO CAUSAL DE LA REGULACIÓN EN EL USO DE INTELIGENCIA ARTIFICIAL GENERATIVA

Autora: Patricia Martín Martínez

Director: Carlos Martínez de Ibarreta Zorita

MADRID | 06/2026

Resumen

La naturaleza oculta del uso de inteligencia artificial generativa (IAGen) en la educación superior, raras veces declarado por los estudiantes e indetectable para las instituciones, impide aislar en condiciones naturalistas el efecto causal de la política, que solo la manipulación experimental permite identificar. El estudio aplicó este diseño con tres condiciones (permisiva, difusa y restrictiva) asignadas aleatoriamente, sobre una muestra final de 180 estudiantes y graduados, mediante una plataforma que integraba la manipulación y el registro conductual objetivo del uso. El marco teórico integra la Teoría del Comportamiento Planificado, el desacoplamiento moral como mecanismo de la disociación entre conciencia ética y conducta, y el Triángulo del Fraude como moderador de la oportunidad percibida. La política importa, pero no de forma lineal: solo la prohibición explícita suprimió el uso declarado, mientras que la condición difusa fue percibida como permisiva y produjo tanto uso real como la permisiva pero tan poco uso reconocido como la restrictiva, una brecha conductual-declarativa que constituye la manifestación empírica más directa del Shadow AI. La ampliación del modelo con el hábito previo y la racionalización normativa mejoró la predicción; esta última medió la relación entre presión y uso declarado, mientras que la obligación moral no moderó el efecto de la política, trasladando la palanca de intervención del carácter individual al diseño del entorno normativo.

Palabras clave: Shadow AI, inteligencia artificial generativa, políticas institucionales, integridad académica, Teoría del Comportamiento Planificado, desacoplamiento moral

Abstract

The structurally concealed use of generative artificial intelligence (GenAI) in higher education, rarely disclosed by students and undetectable by institutions, makes it impossible under naturalistic conditions to isolate the causal effect of policy, which only experimental manipulation can identify. The study implemented this design with three conditions (permissive, diffuse, and restrictive) randomly assigned across a final sample of 180 students and graduates, through a purpose-built platform that integrated the manipulation with objective behavioral recording of actual use. The theoretical framework integrates the Theory of Planned Behavior, moral disengagement as the mechanism behind the dissociation between ethical awareness and conduct, and the Fraud Triangle as a situational moderator of perceived opportunity. Policy matters, but not in a linear way: only explicit prohibition suppressed declared use, whereas the diffuse condition was perceived as permissive and produced as much actual use as the permissive condition but as little acknowledged use as the restrictive one, a behavioral-declarative gap that constitutes the most direct empirical manifestation of Shadow AI. Expanding the model with prior habit and normative rationalization improved prediction; the latter mediated the relationship between pressure and declared use, while moral obligation did not moderate the policy effect, relocating the lever of intervention from individual character to the design of the normative environment.

Keywords: Shadow AI, generative artificial intelligence, institutional policy, academic integrity, Theory of Planned Behavior, moral disengagement

Índice

Índice de tablas y figuras

Introducción

Transformación digital y el auge de la IA generativa en educación superior

A lo largo de las últimas décadas, el avance tecnológico ha reconfigurado la enseñanza superior hasta alterar las premisas mismas sobre las que se asienta el aprendizaje y la enseñanza, un desplazamiento que ha dado lugar al término de "Educación 4.0", expresión con la que se designa la convergencia de tecnologías emergentes como la inteligencia artificial generativa (IAGen), la automatización y el aprendizaje automático (Portilla et al., 2025).

La universidad ha transitado desde modelos anclados en planes de estudio rígidos hacia sistemas capaces de trazar itinerarios individualizados para cada estudiante, sistemas que ofrecen una retroalimentación continua y adaptada sobre tareas y trabajos y que, con ello, incrementan la transparencia de la evaluación y el compromiso del alumnado (S. Chen et al., 2024; Liang et al., 2024; Messer et al., 2024). Facilitan, asimismo, la construcción de recorridos de aprendizaje que se adaptan en tiempo real al ritmo y a las necesidades de cada alumno, lo que reduce su carga cognitiva (Ma y Zhong, 2025); a ello se añade el concurso de herramientas predictivas que habilitan intervenciones tempranas ante dificultades de aprendizaje o riesgos de abandono académico (P. Chen et al., 2024).

El lanzamiento de ChatGPT por parte de OpenAI, en noviembre de 2022, operó como punto de inflexión en la medida en que democratizó el acceso a la IAGen, al ponerla al alcance de cualquier usuario con conexión a internet. Este gran modelo de lenguaje (Large Language Model o LLM) no tardó en despertar el interés por un ecosistema más amplio de herramientas que comprende, junto a ChatGPT, desarrollos análogos de otras compañías, como Copilot (Microsoft), Gemini (Google) o Claude (Anthropic) (Alqahtani et al., 2023). La aptitud de estas plataformas para producir texto fluido y contextualmente pertinente ha multiplicado sus casos de uso en el terreno educativo, y de manera señalada en la educación superior, donde estudiantes y docentes ensayan sin descanso nuevas aplicaciones para estas tecnologías.

La versatilidad de estas herramientas se manifiesta en su capacidad para generar contenidos de naturaleza diversa (imágenes, textos, vídeos y código) con una creatividad y una flexibilidad que se aproximan a las facultades humanas (Strzelecki y ElArabawy, 2024). En el plano de la creación de contenido, los modelos lingüísticos redactan desde informes técnicos especializados hasta ensayos académicos, y descomponen ideas complejas en disciplinas como la medicina, la física o la ingeniería (Bahroun et al., 2023). Su alcance se extiende, además, a

la resolución de problemas científicos y matemáticos en niveles educativos dispares, desde la enseñanza secundaria hasta los contextos universitarios de ingeniería, donde aportan explicaciones detalladas y asistencia en tareas de notable complejidad, si bien arrastran limitaciones nada triviales en precisión y fiabilidad (Alneyadi y Wardat, 2023; Sánchez-Ruiz et al., 2023; Wang et al., 2024). En el ámbito de la programación, en fin, la IAGen no se limita a traducir descripciones en lenguaje natural a código, sino que acompaña al estudiante en la depuración de errores (*debugging*), esclarece la sintaxis existente y eleva de forma apreciable la exactitud con que se aplican los conceptos de programación (Lyu et al., 2024).

El fenómeno de Shadow AI

Conceptualización: de Shadow IT a Shadow AI

El concepto de Shadow IT designa el empleo de dispositivos, sistemas, programas y servicios de tecnología de la información al margen de cualquier aprobación explícita de la organización (Behrens, 2009; Györy et al., 2012; Silic y Back, 2014). El fenómeno suele suceder cuando los sistemas formales no satisfacen una necesidad concreta y los propios usuarios responden con soluciones innovadoras, aunque no autorizadas, concebidas a su medida (Behrens, 2009; Silic et al., 2025). Si bien nació en la década de los 2000 como una optimización de los sistemas de Planificación de Recursos Empresariales, se ha intensificado con el teletrabajo y la consumerización tecnológica (Györy et al., 2012; Trialih, 2023). Su motor principal es la búsqueda de eficiencia operativa frente a la complejidad de los procesos autorizados (Silic y Back, 2014); si bien esa misma búsqueda conlleva riesgos, tanto en el plano de la ciberseguridad, donde cabe la exposición de datos sensibles o la descarga de malware, como en el operativo, derivado de la ausencia de documentación oficial registrada (Behrens, 2009; Silic y Back, 2014).

La proliferación de herramientas impulsadas por inteligencia artificial (IA) ha dado pie a un término emparentado, el de Shadow AI, que abarca el uso de instrumentos de IA por parte de los empleados sin autorización ni supervisión organizacional y que comporta riesgos cualitativamente nuevos (Silic et al., 2025). El peligro deja de ser estático debido a que los instrumentos de Shadow AI amplían sus capacidades respecto de los del Shadow IT, convirtiéndose en adaptativos, generativos y aptos para decidir (Silic et al., 2025); una sola herramienta basta entonces para procesar información, generar contenido y adoptar decisiones (Silic et al., 2025), con lo que la validación humana se vuelve prescindible. De ahí que

disminuyan la visibilidad y la transparencia organizacionales, pues estos procesos se incorporan en los flujos de trabajo y pueden llegar a operar de forma recursiva. Los riesgos propiamente nuevos comprenden las alucinaciones de la IA, la disolución de la responsabilidad por fallos de atribución y el incremento o la perpetuación de desigualdades a causa de los sesgos de los modelos (Silic et al., 2025), a lo que se añaden el envenenamiento de modelos concebido para alterar los sistemas (Puthal et al., 2025) y el impacto cognitivo sobre el pensamiento crítico, la creatividad y la originalidad de los usuarios (Puthal et al., 2025; Silic et al., 2025).

A diferencia de los contextos corporativos, donde los riesgos son ante todo operativos y de seguridad, en la educación superior el peligro de fondo reside en el impacto sobre el proceso de aprendizaje del estudiante y sobre la adquisición de competencias (Galindo-Domínguez et al., 2026). El uso no regulado de Shadow AI en este ámbito arrastra consecuencias como la disminución de la conexión humana, la pereza o la atrofia cognitiva y el debilitamiento del pensamiento crítico y de la creatividad (Al-Zahrani, 2024); incide, en concreto, sobre el desarrollo intelectual, en la medida en que atrofia habilidades fundamentales como la retención o las funciones ejecutivas (Galindo-Domínguez et al., 2026; Zhai et al., 2024). Esta especificidad convierte el Shadow AI en la educación superior en un reto de naturaleza esencialmente pedagógica, que reclama marcos de comprensión diferenciados, orientados a garantizar que la IA opere como extensión de la capacidad humana y no como su sustituto (Zhai et al., 2024).

Manifestación en el contexto universitario

La adopción de instrumentos de IAGen en la educación superior ha alcanzado un grado tal que su presencia en los procesos académicos resulta ya difícilmente reversible. La literatura más reciente cifra las tasas de uso activo entre el estudiantado universitario por encima del 85%, hasta rozar el 98% en el caso de los alumnos de posgrado (Smit et al., 2025); entre quienes cursan estudios de grado, un 65% recurre a estas herramientas al menos una vez por semana para tareas académicas (Askarkyzy y Zhunusbekova, 2024). Tal penetración, masiva y generalizada, contrasta con una respuesta institucional estructuralmente insuficiente, ya que el 74,4% de las instituciones de educación superior carece de política formal sobre IA y se gobierna mediante decisiones *ad hoc* del profesorado (Benayoune et al., 2026), una fragmentación que la literatura confirma de manera sistemática como el estado predominante a escala global (Kangwa et al., 2025). De ahí se sigue la consecuencia más significativa: el 96% del estudiantado reclama claridad normativa a sus universidades mientras que el 57% no

advierte conflicto alguno entre el uso encubierto de IAGen y los principios de integridad académica (Askarkyzy y Zhunusbekova, 2024), señal de que la exigencia de regulación y la propensión al incumplimiento conviven, sin tensión aparente, en un mismo sujeto (Smit et al., 2025).

La coexistencia de este uso de la IA con la ausencia de regulación hace metodológicamente imposible identificar el efecto causal de las políticas mediante estudios observacionales. Esto es porque el comportamiento objeto de estudio permanece estructuralmente oculto, dado que los estudiantes editan, parafrasean o mezclan el contenido generado por IAGen con su propia escritura, reduciendo la sensibilidad de cualquier sistema de detección y generando un sesgo de ocultamiento que invalida la inferencia causal en condiciones naturalistas (Tsigaris y Teixeira da Silva, 2026), mientras que la ambigüedad normativa crea simultáneamente un entorno de riesgo moral (Smit et al., 2025) en el que la variable independiente, el tipo de política institucional, es fragmentada o inexistente (Benayoune et al., 2026; Kangwa et al., 2025), y la variable dependiente, el uso real de IAGen, es activamente encubierta por los propios sujetos. Ante ello, un diseño en el que la política se manipula experimentalmente como variable de tratamiento se observa como la vía óptima para aislar su efecto sobre la intención de uso y los factores conductuales que median esa relación, superando las limitaciones estructurales de la investigación observacional.

Justificación y objetivos del trabajo de fin de grado

Vacío en la literatura y aportación del estudio

La literatura existente sobre el uso de IA en la educación superior es esencialmente observacional y correlacional, lo que le impide aislar el efecto causal de la política institucional. Ello es agravado por la fragmentación de la variable independiente y el encubrimiento activo de la dependiente. No existen, en consecuencia, estudios que manipulen experimentalmente la política como variable de tratamiento y que, al mismo tiempo, registren la conducta real de uso al margen de lo que el estudiante declara. La aportación de este trabajo busca cubrir ese vacío mediante un diseño experimental que manipula tres entornos normativos, combina una medida declarada y una medida conductual objetiva del uso de IA e integra los determinantes individuales y situacionales en un único modelo de tres niveles, la Teoría del Comportamiento Planificado, el desacoplamiento moral y el Triángulo del Fraude, lo que permite observar no solo si la política modifica el comportamiento, sino a través de qué mecanismos lo hace.

Objetivos generales y específicos

El objetivo general de este trabajo es determinar el efecto causal del tipo de política institucional sobre el uso de inteligencia artificial en tareas académicas y profesionales, así como identificar los factores conductuales que explican ese uso. De él se derivan cinco objetivos específicos: contrastar si la política afecta al uso de IA y si la ambigüedad normativa se asimila a la permisividad; evaluar la capacidad predictiva de los componentes clásicos de la Teoría del Comportamiento Planificado; comprobar si su ampliación con el hábito, la racionalización y los factores del Triángulo del Fraude mejora esa predicción; examinar si la racionalización media la relación entre la presión y el uso; y determinar si la obligación moral modera el efecto de la política.

Marco teórico

Inteligencia artificial generativa y la educación superior

Definición y fundamentos de la inteligencia artificial generativa

La Inteligencia Artificial Generativa (IAGen) se define como un conjunto de algoritmos y modelos de IA capaces de producir nuevo contenido en diferentes formatos como imágenes, texto, vídeos o audio, entre otros, con una capacidad de adaptabilidad y creatividad que se asemejan a la humana (He et al., 2025). Históricamente, la IAGen ha evolucionado desde sistemas basados en reglas que creaban datos nuevos, a modelos fundacionales (como los Grandes Modelos de Lenguaje o LLMs por sus siglas en inglés) que constituyen el precedente de la etapa actual (He et al., 2025). Esta evolución se apoya en la gran escala de parámetros y datos de entrenamiento, que habilita capacidades emergentes capaces de resolver tareas para las que el modelo no fue entrenado (Wei et al., 2022).

El lanzamiento de ChatGPT el 30 de noviembre de 2022 marcó el punto de inflexión en la adopción masiva de la IAGen en el ámbito educativo. Nguyen et al. (2025), en un análisis del discurso educativo en redes sociales, documentaron que en los cuatro meses posteriores al lanzamiento se generaron más de 2,4 millones de publicaciones relacionadas con ChatGPT, de las cuales 91.842 contenían referencias explícitas a contextos educativos, ilustrando la velocidad e intensidad de la disrupción percibida. Este impacto se manifiesta de forma particularmente intensa en la educación superior, cuyo modelo pedagógico, basado en la producción escrita autónoma, el pensamiento analítico y la evaluación de competencias complejas, resulta especialmente sensible a una tecnología capaz de generar contenido académico de apariencia legítima en segundos (Chen y Cheung, 2025).

Oportunidades en la transformación educativa

Las oportunidades que la IAGen abre para la transformación del aprendizaje universitario operan en tres dimensiones que, en conjunto, explican la atracción estructural que ejerce sobre el estudiante y que constituye el factor motivacional del fenómeno Shadow AI: la personalización, el *feedback* inmediato y el *engagement*. La personalización a escala masiva implica adaptar de forma dinámica el contenido, el ritmo y el nivel de dificultad al perfil de cada estudiante, una realidad que los sistemas tradicionales habían perseguido sin poder materializarla más allá de la tutoría individualizada y que la IAGen hace técnicamente accesible

para poblaciones de cualquier tamaño (Pang y Wei, 2025). El *feedback* inmediato elimina la latencia entre el error y la corrección, un intervalo que la literatura identifica como uno de los principales obstáculos para la consolidación del aprendizaje, porque deteriora la capacidad del estudiante para asociar la retroalimentación con el proceso cognitivo que la generó (Pang y Wei, 2025). Y sobre el *engagement*, el metaanálisis de Xia et al. (2025), basado en 33 estudios con población universitaria, constata efectos positivos y significativos en sus cuatro dimensiones, así como un efecto de magnitud alta sobre la motivación, especialmente pronunciados en contextos de aprendizaje individual o en grupos reducidos. La convergencia de estas tres dimensiones configura un perfil de utilidad percibida que opera con relativa independencia del marco normativo, lo que contribuye a explicar por qué los estudiantes recurren a la IAGen incluso donde su uso no está autorizado, si bien esos mismos beneficios son los que otros autores asocian con la dependencia cognitiva y la pereza metacognitiva (Fan et al., 2025; Zhang y Xu, 2025).

Impacto en las habilidades cognitivas y la dependencia

Ambos metaanálisis de referencia (Chen y Cheung, 2025; Ma y Zhong, 2025) organizan el impacto de la IAGen sobre el aprendizaje con taxonomías distintas, tridimensional en Ma y Zhong (2025) y de cinco dominios en Chen y Cheung (2025), y es el contraste entre ambas el que permite identificar dónde la herramienta resulta más eficaz y dónde emergen sus efectos más problemáticos para el desarrollo intelectual autónomo del estudiante. En la dimensión cognitiva, que agrupa la adquisición y el procesamiento de conocimiento, los dos convergen en los efectos más elevados (g = 0,795 en Ma y Zhong (2025); g^{+} = 1,009 en las habilidades lingüísticas de Chen y Cheung (2025)). La paradoja central, sin embargo, se revela en el plano competencial, porque Ma y Zhong (2025) agrupan la resolución de problemas y la autorregulación en un único constructo de efecto elevado (g = 0,711), mientras que cuando Chen y Cheung (2025) aíslan la metacognición como constructo independiente el efecto se vuelve prácticamente nulo (g^{+} = 0,078; p = ,789), de modo que la IAGen actúa como excelente asistente en las tareas básicas y lingüísticas pero apenas incide sobre los procesos cognitivos más profundos e independientes. Las teorías que ambos estudios emplean interpretan esta contradicción: la Teoría de la Carga Cognitiva explica el elevado efecto cognitivo, pero advierte de que una reducción excesiva del esfuerzo genera dependencia al delegar en la máquina el trabajo intelectual (Stadler et al., 2024), mientras que la Teoría del Aprendizaje Autorregulado explica el dato más revelador, ya que unas herramientas diseñadas para generar

respuestas carecen de mecanismos que obliguen al estudiante a monitorizar su progreso y producen una pereza metacognitiva que convierte el éxito académico aparente en un producto de la máquina antes que en un indicador del desarrollo intelectual real del estudiante (Fan et al., 2025).

Desafíos éticos, integridad y gobernanza

La IAGen representa un problema cualitativamente diferente al plagio tradicional, conceptualizado como ***AI-giarismo*** (Başer et al., 2026; Chan, 2023), porque fractura simultáneamente las bases de la detección y la autoría académica. A diferencia del plagio clásico, que se fundamenta en la copia de fuentes rastreables, la IAGen genera contenido único y no repetitivo en cada iteración (Chan, 2023), creando una zona gris de coautoría humano-máquina que los detectores existentes son incapaces de resolver ya que sus algoritmos de análisis lingüístico resultan altamente vulnerables a la ingeniería de *prompts* y a la paráfrasis humana o algorítmica, y su eficacia se reduce drásticamente cuando el texto generado es editado por un humano o a la inversa (Ayub et al., 2024; Liu et al., 2024; Tsigaris y Teixeira da Silva, 2026). A esta indetectabilidad estructural se suma que la herramienta elimina las barreras financieras, transaccionales y sociales que a lo largo de la historia han obstaculizado formas graves de deshonestidad, como el *contract cheating* (Chan, 2023), gracias a su coste cero, privacidad e inmediatez. Esto proporciona mecanismos para racionalizar al permitir que el alumno reencuadre el uso como ayuda tecnológica legítima (Başer et al., 2026). La convergencia de ambos factores debilita los mecanismos clásicos de disuasión institucional, trasladando el peso de la contención del fraude en exclusiva a la autorregulación del estudiante, una barrera que con frecuencia cede ante la presión académica (Başer et al., 2026; Chan, 2023).

La respuesta de las universidades ante la IAGen se caracteriza por ser fragmentada, cautelosa y reactiva, con un patrón generalizado de ausencia de marcos normativos cohesionados que opera en su lugar a través de decisiones *ad hoc* del profesorado (Benayoune et al., 2026; Jin et al., 2025). Esta fragmentación se agrava por una carencia crítica de validación empírica a largo plazo, dado que la mayoría de las políticas actuales no están fundamentadas en estudios longitudinales que rastreen su efectividad real, y solo una minoría de instituciones participa activamente en la evaluación continua de su impacto, lo que expone a muchas políticas estáticas al riesgo de obsolescencia acelerada frente al ritmo de avance tecnológico (Kangwa et al., 2025).

El tipo de enfoque adoptado tiene consecuencias conductuales documentadas: las instituciones reactivas registran mayores índices de deshonestidad y confusión ética, mientras que las que integran la literacidad en IAGen en los módulos académicos y enmarcan las tareas asistidas en debates éticos logran menores tasas de fraude y mayor *engagement* (Kangwa et al., 2025). La ambigüedad normativa produce además un riesgo moral en sentido estricto, al incentivar comportamientos oportunistas *ex ante*, cuando el estudiante delega el esfuerzo cognitivo a la máquina anticipando la ausencia de consecuencias, y *ex post*, cuando explota la incapacidad institucional de detectar la infracción (Benayoune et al., 2026; Smit et al., 2025). Las prohibiciones explícitas no resuelven el problema sino que lo desplazan hacia el uso encubierto (Smit et al., 2025), empujando incluso a estudiantes honestos al uso no declarado como respuesta racional ante la percepción de ventaja competitiva de sus pares (Tsigaris y Teixeira da Silva, 2026), lo que revela que la gestión de la IAGen no puede reducirse a un problema de control y detección sino que requiere comprender los determinantes conductuales que median entre la política y el comportamiento del estudiante.

Fundamentos conductuales del Shadow AI: predictores, mecanismos y moderadores del efecto de la política institucional sobre el uso de IA

La Teoría del Comportamiento Planificado como modelo base predictivo

La Teoría del Comportamiento Planificado (TPB por sus siglas en inglés) desarrollada por Ajzen (1991) es un modelo psicológico diseñado para predecir y explicar la toma de decisiones y la conducta humana. La TPB sostiene que la realización de un comportamiento concreto viene dada por la intención de la persona de llevarlo a cabo. Esta intención de actuar se forma a partir de tres componentes independientes: la actitud, las normas subjetivas y el control conductual percibido. La actitud implica la evaluación favorable o desfavorable que un individuo hace sobre la realización de un comportamiento en específico y las creencias sobre sus consecuencias. Las normas subjetivas representan la presión social percibida que ejercen otros grupos de pares o referentes respecto a si el comportamiento se debe, o no, realizar. Por último, el control conductual percibido es la apreciación subjetiva del individuo sobre la facilidad o dificultad de realizar el comportamiento, incluyendo tanto la capacidad del individuo para realizar la acción como el control que tiene sobre ella. Cuanto mayor sea cada uno de los componentes de la TPB, más probable es que el individuo lleve a cabo la acción.

La evidencia empírica muestra de forma consistente que la TPB es un modelo eficaz para predecir tanto la intención de cometer fraude académico como su materialización (Stone et al., 2010). Su versión extendida resulta especialmente potente, como demostró el análisis transcultural de Chudzicka-Czupała et al. (2016), en el que un modelo que incorporaba la obligación moral explicaba entre el 46% y el 75% de la varianza en las intenciones de hacer trampa según el país, frente a las pruebas de la estructura pura de la TPB, que se sitúan en torno al 21% para la intención y al 36% para el comportamiento (Stone et al., 2010), una distancia coherente con los metaanálisis que cifran este impacto en torno al 28% (Whitley, 1998).

La TPB se ha adaptado al ecosistema de la IA en la educación superior, pasando de un modelo de decisiones individuales a marcos integrados que la fusionan con el Modelo de Aceptación de la Tecnología, la Teoría Unificada de Aceptación y Uso de Tecnología y la Teoría del Aprendizaje Autorregulado (Jazim et al., 2025; Kangwa et al., 2025). En esa adaptación, la actitud se explica por la evaluación que el estudiante hace de las ventajas frente a los riesgos, y la evidencia muestra que, en relación con la IAGen, el ahorro de tiempo y la mejora de la calidad final superan a los riesgos percibidos (Ivanov et al., 2024); la norma subjetiva se desplaza de la presión social a la presión institucional, incorporando el peso de la propia universidad junto al de los pares (Kangwa et al., 2025); y el control conductual percibido queda muy influido por la autoeficacia tecnológica y el conocimiento del estudiante (Ha et al., 2025; Tbaishat et al., 2025).

La TPB se presenta como un marco clásico capaz de explicar parte del uso de IAGen no autorizado, específicamente la motivación para usarla; sin embargo, la revisión anterior muestra que la ampliación de conceptos y constructos enriquece de forma clara la predicción y explicación de Shadow AI. En concreto, el análisis transcultural de Chudzicka-Czupała et al. (2016) relativo a la copia en estudiantes demostró que al añadir la variable de obligación moral a la TPB se producía una mejora drástica en su capacidad predictiva y explicativa, aumentando significativamente la varianza explicada y provocando un efecto de supresión en otras variables (que variaban según la cultura). Aunque el estudio transcultural utilizó la obligación moral como freno, las investigaciones más recientes sobre fraude explican que el desacoplamiento moral es precisamente la base psicológica de la racionalización (Bandura, 1999; Başer et al., 2026).

El desacoplamiento moral: por qué la actitud no se traduce en intención ética

Los estudiantes universitarios, situados habitualmente en la etapa de adultez emergente, presentan condiciones que dificultan el razonamiento moral óptimo y los hacen especialmente vulnerables a la deshonestidad académica: la maduración moral incompleta, la autonomía sin experiencia vital consolidada y el desfase cognitivo-emocional propio de esa etapa generan un perfil en el que la conciencia ética existe pero opera de forma inestable, creando las condiciones para que mecanismos de desacoplamiento moral encuentren escasa resistencia interna (Bélanger et al., 2012).

El desacoplamiento moral, fundamentado en la Teoría Cognitiva Social de Bandura, es el proceso de reestructuración cognitiva mediante el cual los individuos se desvinculan de sus estándares morales internos y actúan de forma poco ética sin experimentar culpa o autocensura (Bandura, 1999). En el contexto del *AI-giarismo*, este mecanismo se activa a través de narrativas de neutralización específicas: la distorsión de consecuencias ("no perjudico a nadie"), la difusión de responsabilidad ("todos lo hacen"), la justificación por presión académica y, de forma novedosa, alegar desconocimiento sobre el engaño de la IA (*unawareness of AI deception*), utilizando la ilusión de credibilidad y el riesgo de alucinación de los modelos como excusa para evadir la responsabilidad de verificar la información generada (Başer et al., 2026). La ambigüedad normativa institucional trasciende la mera ausencia de freno para convertirse en facilitador activo de este desacoplamiento al generar un riesgo moral *ex ante*, donde el estudiante anticipa que la institución carece de pautas para sancionar su dependencia de la herramienta, y *ex post*, donde explota la incapacidad institucional de detectar la infracción para obtener ventaja inmediata (Smit et al., 2025), de modo que la política difusa no solo no desincentiva el *Shadow AI* sino que lo incentiva estructuralmente, lo que revela que el análisis no puede detenerse en los factores individuales de la TPB sino que debe incorporar las condiciones situacionales que modulan la oportunidad percibida.

El Triángulo del Fraude: la oportunidad como moderador situacional

El Triángulo del Fraude, propuesto originalmente en el ámbito empresarial y adaptado al contexto académico, establece que el comportamiento fraudulento requiere la convergencia de tres condiciones: la presión, entendida como los factores subjetivos que motivan al individuo, entre ellos la sobrecarga de tareas, la competencia entre pares y la necesidad de obtener altas

calificaciones; la oportunidad, factor objetivo y situacional derivado de debilidades en los controles y la supervisión que permiten al estudiante percibir que puede cometer el fraude sin ser descubierto; y la racionalización, proceso cognitivo mediante el cual el individuo justifica su comportamiento deshonesto para neutralizar el malestar psicológico (Becker et al., 2006; Lewellyn y Rodriguez, 2015). La evidencia empírica confirma de forma consistente la aplicabilidad del modelo al fraude académico, demostrando que los tres factores afectan directa y positivamente la materialización de la deshonestidad (Smith et al., 2023). En este sentido, destaca el hallazgo de Heriyati y Ekasari (2020), quienes demostraron que el razonamiento moral no actúa como moderador significativo del modelo; esto implica que incluso los estudiantes con alta conciencia ética cometen deshonestidad cuando la oportunidad está presente, desplazando el foco de las intervenciones institucionales desde la apelación al carácter individual hacia el diseño del entorno y, en particular, hacia la reducción de la oportunidad percibida como la palanca más efectiva y controlable.

En el contexto del *AI-giarismo*, el vértice de oportunidad del triángulo se redefine estructuralmente: la oportunidad ya no reside en evadir la supervisión de un vigilante sino en explotar la indetectabilidad algorítmica de la herramienta, creando una asimetría en la que el estudiante sabe que la institución carece de capacidad técnica para probar irrefutablemente la infracción (Başer et al., 2026; Tsigaris y Teixeira da Silva, 2026). El Diamante del Fraude extiende el modelo añadiendo la capacidad como cuarto factor, argumentando que el fraude no se materializa si el individuo no posee las habilidades necesarias para ejecutarlo (Smith et al., 2023); en el contexto del *Shadow AI*, esta capacidad se concreta en la competencia digital, dado que los estudiantes con mayor dominio técnico pueden emplear ingeniería de *prompts* iterativa para humanizar el texto y burlar tanto los detectores algorítmicos como la evaluación docente (Ayub et al., 2024; Tsigaris y Teixeira da Silva, 2026), convirtiendo la competencia tecnológica en el mecanismo que transforma la oportunidad latente en fraude sofisticado e invisible. La política institucional es la variable que más directamente modifica la oportunidad percibida: las políticas ambiguas o inexistentes generan riesgo moral al eliminar límites explícitos; las políticas restrictivas no reducen la oportunidad sino que desplazan el comportamiento hacia la clandestinidad (Smit et al., 2025), empujando incluso a estudiantes honestos al uso no declarado como respuesta racional de supervivencia para mantenerse competitivos (Tsigaris y Teixeira da Silva, 2026); y solo las políticas anticipatorias y transparentes reducen genuinamente la oportunidad al reformular el diseño de las evaluaciones y cerrar las brechas

sistémicas que hacen rentable la trampa (Kangwa et al., 2025), lo que convierte el tipo de política en la variable de tratamiento central del diseño experimental de este estudio.

Las políticas institucionales como variable de tratamiento

Las políticas institucionales sobre IAGen no constituyen marcos convergentes sino respuestas heterogéneas determinadas por prioridades estratégicas, contextos geopolíticos y valores culturales (Jin et al., 2025; Kaya-Kasikci et al., 2025), una divergencia que se agrava por la brecha estructural entre la velocidad de adopción tecnológica y la capacidad regulatoria institucional, ya que mientras las capacidades de la IAGen evolucionan en ciclos de semanas los ciclos normativos se miden en semestres, de modo que las universidades, ante esa incertidumbre, copian políticas de líderes regionales de forma *ad hoc* sin alineación estratégica real, un fenómeno que Benayoune et al. (2026) explican mediante el isomorfismo institucional y que produce un mosaico de normativas superpuestas e inconsistentes cuya carga interpretativa recae sobre profesores y estudiantes, socavando la confianza y la equidad percibida.

La política institucional no actúa como variable de contexto sino que reconfigura simultáneamente los tres componentes de la TPB: enmarca la legitimidad del uso de IAGen modulando la actitud del estudiante hacia el uso ético frente al encubierto, desplaza el peso de la presión social hacia un estándar normativo institucional que opera como isomorfismo coercitivo sobre la norma subjetiva (Benayoune et al., 2026), y modifica directamente el control conductual percibido, de modo que la ambigüedad genera percepción de alta facilidad de uso sin restricciones mientras que las políticas explícitas actúan como barreras de control (Kangwa et al., 2025). Esta acción simultánea sobre la TPB se extiende a los otros dos niveles teóricos del modelo: sobre el Triángulo del Fraude, las políticas laxas o inexistentes son codificadas por los estudiantes como ventana de oportunidad mientras que las políticas claras y predecibles a nivel de módulo desactivan ese vértice al establecer expectativas invariables (Smit et al., 2025); sobre el desacoplamiento moral, las directrices explícitas bloquean las narrativas de neutralización al obligar al estudiante a confrontar su propia disonancia cognitiva, mientras que el vacío normativo las facilita estructuralmente (Başer et al., 2026; Smit et al., 2025). La política se configura, de este modo, como la variable de tratamiento metodológicamente idónea para un diseño experimental, porque a diferencia de los rasgos psicológicos individuales, es un factor externo, maleable y controlable que permite manipular distintos entornos normativos para observar variaciones causales directas sobre la intención de

uso y las justificaciones éticas, aislando efectos que los estudios correlacionales, contaminados por la cultura institucional general, no pueden separar.

Hipótesis y modelo conceptual

Trasladado este modelo de tres niveles a términos empíricos, la política institucional constituye la variable independiente, manipulada experimentalmente en tres condiciones, mientras que el uso de IA se recoge como variable dependiente de dos formas: la que el propio estudiante declara y la que la plataforma registra de forma objetiva, lo que permite contrastar el efecto de la política tanto sobre lo que el estudiante reconoce como sobre lo que efectivamente hace. Los factores conductuales descritos en los apartados anteriores se incorporan a este diseño desempeñando tres funciones: la de predictores individuales cuya capacidad explicativa se pone a prueba frente a la del modelo clásico de la Teoría del Comportamiento Planificado, la de mecanismo capaz de articular la relación entre la presión situacional y la conducta, y, con carácter exploratorio, su eventual papel moderador del efecto de la política.

A partir de este planteamiento se derivan cinco hipótesis. La primera (H1) se apoya en que la política reconfigura la legitimidad percibida del uso y la oportunidad de realizarlo sin consecuencias, y en que la literatura sobre riesgo moral sostiene que la ausencia de una prohibición explícita tiende a interpretarse como autorización antes que como cautela (Benayoune et al., 2026; Smit et al., 2025), de modo que cabe anticipar que la política ejerza un efecto significativo sobre el uso de IA, con un uso menor en la condición restrictiva que en la permisiva y con una condición difusa que, al no fijar un límite explícito, se prevé más próxima a la permisiva que a una posición intermedia.

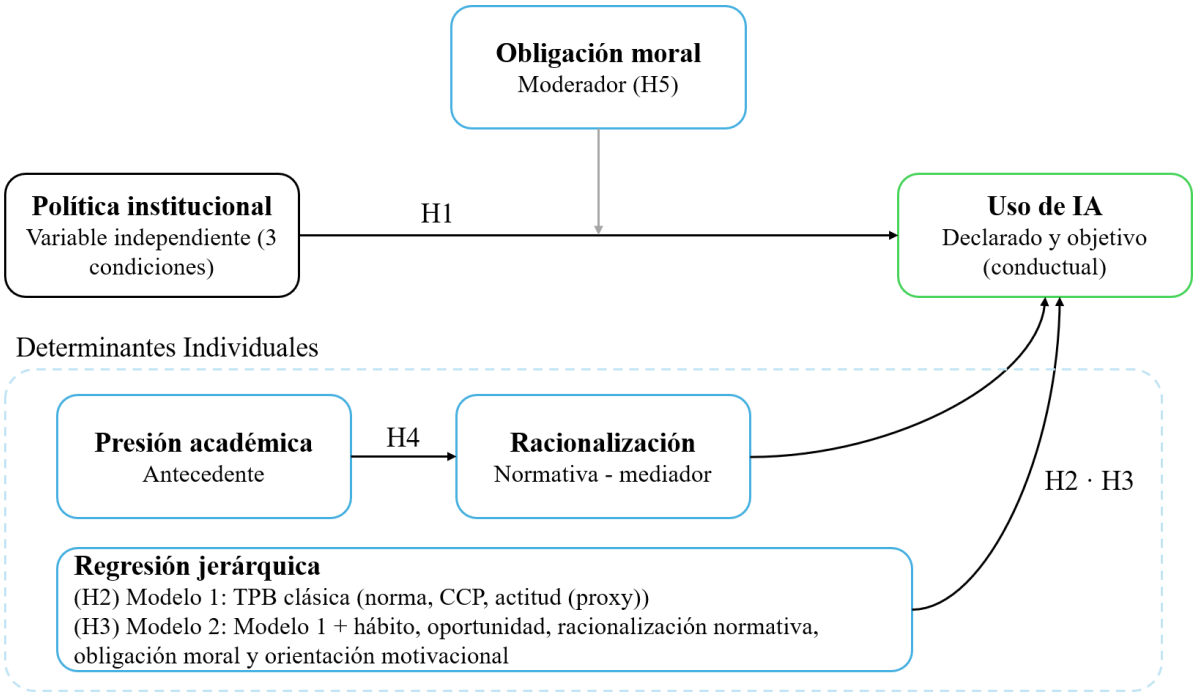
Las dos hipótesis siguientes forman una sola cadena argumental. Dado que la conciencia ética del estudiante tiende a operar de forma inestable y no siempre se traduce en conducta, se espera que el modelo clásico de la Teoría del Comportamiento Planificado, sostenido sobre la actitud, la norma subjetiva y el control conductual percibido, presente por sí solo una capacidad predictiva limitada sobre el uso de IA no autorizado (H2), y es esa misma limitación prevista la que conduce a esperar que la incorporación del comportamiento pasado, la racionalización y los factores del Triángulo del Fraude mejore la predicción respecto al modelo restringido a la TPB clásica (H3), en la línea de la ganancia que Chudzicka-Czupała et al. (2016) obtuvieron al añadir la obligación moral a la ecuación.

Las dos últimas hipótesis descienden al plano situacional y moral. Puesto que el Triángulo del Fraude sostiene que la presión necesita ser neutralizada cognitivamente antes de materializarse en conducta, se prevé que la racionalización, y en particular su componente normativo, medie la relación entre la presión académica y el uso de IA, de manera que la presión ejerza sobre el

uso un efecto indirecto y no directo (H4). Y dado que Heriyati y Ekasari (2020) constataron que el razonamiento moral no modera la materialización del fraude, se espera, con carácter exploratorio, que el efecto de la política sobre el uso se mantenga al margen del nivel de obligación moral del estudiante (H5), un resultado que, de confirmarse, situaría la palanca de la intervención no en el carácter del individuo sino en el diseño del entorno normativo que la institución decide construir. El modelo conceptual y las relaciones contrastadas se recogen en la Figura 1.

Figura 1

Modelo conceptual de los determinantes del uso de IA y de las hipótesis del estudio



Nota. La flecha gruesa que une la política institucional con el uso de IA representa el único efecto manipulado experimentalmente y, por tanto, de naturaleza causal (H1); las relaciones restantes (H2 a H4) son de carácter predictivo y correlacional. La línea que parte de la obligación moral indica la relación de moderación contrastada (H5). El recuadro inferior agrupa los determinantes individuales organizados en los tres niveles del modelo (Teoría del Comportamiento Planificado, desacoplamiento moral y Triángulo del Fraude) e incluye la regresión jerárquica que compara el modelo clásico de la TPB (H2) con su ampliación (H3), así como la vía de mediación de la presión a través de la racionalización (H4). Los colores distinguen la variable independiente (gris), la variable dependiente (verde) y los determinantes individuales junto con el moderador (morado). CCP = control conductual percibido.

Metodología

Diseño experimental

El estudio adopta un diseño experimental entre sujetos de un solo factor con tres niveles, en el que la política institucional sobre el uso de inteligencia artificial constituye la variable independiente manipulada y los participantes fueron asignados aleatoriamente a una de tres condiciones (permisiva, difusa y restrictiva). La variable dependiente, el uso de IA, se operacionalizó en un doble plano que combina una medida de autoinforme (el porcentaje de contenido que el participante declara haber generado o parafraseado con IA) y una medida conductual objetiva (el volumen de texto introducido mediante un asistente de IA integrado en la plataforma), de manera que el efecto del tratamiento pudiera contrastarse tanto sobre la conducta que el estudiante reconoce como sobre la que la plataforma registra al margen de su declaración. Junto a estas variables, un conjunto de constructos psicológicos y situacionales medidos por cuestionario se incorporó en calidad de predictores y, en el caso de la cadena que une presión y racionalización, de mecanismo mediador, explorándose además el eventual papel moderador de la obligación moral sobre el efecto de la política. El estudio se administró íntegramente a través de una aplicación web desarrollada ad hoc que presentaba la manipulación, registraba la conducta e integraba el asistente de IA, lo que permitió combinar el control propio de un experimento con la captura no intrusiva de la conducta real de uso (su arquitectura técnica se detalla en el Anexo A, subapartado A3).

Participantes

Se recogieron 190 respuestas a lo largo de un mes mediante difusión en redes sociales, correo electrónico y otros canales de distribución. Tras aplicar los criterios de exclusión que se detallan en el apartado de resultados, la muestra final quedó compuesta por 180 participantes, de los cuales el 59,4% eran mujeres y el 40,6% hombres, con una edad media de 29,2 años (*DE* = 13,2; rango 18-66) y una nota media autodeclarada de 7,70 sobre 10. En cuanto al nivel de estudios, el 56,1% cursaba un Grado, el 31,7% un Máster, el 7,8% Formación Profesional Superior y el 4,4% restante se repartía entre Doctorado (2,2%) y Bachillerato (2,2%), una composición deliberadamente amplia que responde a que el instrumento se diseñó para resultar accesible tanto a estudiantes en activo como a recién graduados incorporados a entornos profesionales, de modo que los ítems se formularon en términos aplicables tanto al trabajo académico como al laboral con el objetivo de aumentar el tamaño muestral, aunque ello obligue

a cierta prudencia al extrapolar los resultados a la población estrictamente universitaria. La asignación aleatoria produjo grupos de tamaño equilibrado, con 63 participantes en la condición permisiva, 57 en la difusa y 60 en la restrictiva (véanse las Tablas B1 a B3 del Anexo B).

Tarea experimental

La tarea consistía en redactar un texto de 60–120 palabras con un asistente de IA disponible en las tres condiciones; cada inserción de contenido generado quedaba registrada de forma objetiva como número de caracteres, con independencia de lo que el participante declarara después. La arquitectura técnica completa figura en el Anexo A, subapartado A3.

Instrumentos y medidas

La restrictividad percibida de la política se midió con un ítem de comprobación de la manipulación en una escala de 1 a 7. El resto de los constructos se evaluó mediante una batería de catorce ítems en escala Likert de 1 a 5, salvo el de comportamiento pasado, formulado como escala de frecuencia. La batería cubría los componentes clásicos de la Teoría del Comportamiento Planificado (norma subjetiva descriptiva e injuntiva, y control conductual percibido en sus facetas de capacidad y de evasión de la detección), anclados en Ajzen (1991) y en su adaptación al fraude académico (Beck y Ajzen, 1991; Stone et al., 2010); los tres vértices del Triángulo del Fraude (oportunidad percibida, presión académica o laboral y racionalización, esta en sus variantes utilitaria y de normalización social), apoyados en Becker et al. (2006), Smith et al. (2023) y Başer et al. (2026); la obligación moral en sus facetas de internalización, culpa y principios, fundamentada en Bandura (1999), Bélanger et al. (2012) y Chudzicka-Czupała et al. (2016); y, por último, la ambigüedad normativa percibida, la orientación motivacional al rendimiento y la frecuencia previa de uso de IA como indicador de comportamiento pasado (Harding et al., 2007). La actitud no se midió con un ítem propio y se aproximó, con cautela, mediante la racionalización utilitaria, dado su solapamiento con la evaluación favorable del comportamiento, decisión que se asume como limitación y se retoma en la discusión. El instrumento se diseñó deliberadamente breve para resultar compatible con una tarea conductual y maximizar las respuestas completas, lo que llevó a operacionalizar la mayoría de los constructos mediante ítems únicos; esta decisión prioriza la viabilidad pero impide estimar su consistencia interna, de modo que solo la obligación moral, medida con tres

ítems, permite hacerlo (α = ,872). La formulación de los catorce ítems y su correspondencia con cada constructo y su fuente teórica figuran en la Tabla A1 del Anexo A.

Condiciones experimentales

Antes de iniciar la tarea, cada participante leía un texto breve que presentaba la normativa sobre el uso de IA y que constituía la manipulación experimental. Los tres textos mantenían constantes la longitud y la estructura para que la única variación entre condiciones fuera el grado de permisividad: la permisiva autorizaba explícitamente el uso, la restrictiva lo prohibía de forma expresa y la difusa describía un marco ambiguo que no establecía ni autorización ni prohibición claras. El texto exacto de las tres condiciones se recoge en la Tabla A2 del Anexo A.

Procedimiento

El procedimiento siguió una secuencia fija de pantallas. Tras una pantalla de bienvenida en la que se recababa el consentimiento informado mediante una casilla de aceptación obligatoria, el participante completaba una breve sección de datos demográficos y accedía a las instrucciones junto con el texto de la política asignada de forma aleatoria, que permanecía visible un mínimo de tres segundos. A continuación, realizaba la tarea de redacción con el asistente disponible y declaraba qué porcentaje de su texto había sido generado o parafraseado por IA. Después respondía a las preguntas de control sobre la restrictividad percibida de la política y sobre si había utilizado el botón de IA o herramientas externas, completaba una segunda sección demográfica (universidad, rama de conocimiento y nota media) y, por último, los dos bloques de escalas Likert sobre su entorno y sus valores. La sesión se cerraba con una pantalla de agradecimiento que ofrecía la posibilidad de dejar un correo para recibir información sobre el estudio.

Plan de análisis

El plan de análisis siguió la lógica causal del diseño y se ejecutó íntegramente en Jamovi. En primer lugar, se verificó, mediante un análisis de la varianza de un factor sobre la restrictividad percibida, que la manipulación hubiera sido percibida de forma diferencial entre condiciones, requisito para atribuir al tratamiento las diferencias conductuales posteriores. A continuación, se contrastó el efecto principal de la política sobre el uso de IA, primero declarado y después objetivo, recurriendo a la corrección de Welch y a las comparaciones post hoc de Games-

Howell ante la violación del supuesto de homocedasticidad, mientras que la consistencia entre la conducta registrada y el autoinforme se examinó con correlaciones de Spearman, más robustas ante distribuciones sesgadas. La capacidad predictiva de los factores individuales se evaluó mediante una regresión jerárquica que comparó un primer modelo restringido a los componentes clásicos de la Teoría del Comportamiento Planificado con un segundo que incorporó el comportamiento pasado, la racionalización y los factores del Triángulo del Fraude, de modo que el incremento de varianza explicada permitiera valorar si el modelo ampliado mejoraba la predicción del clásico. La relación indirecta entre presión y conducta se exploró mediante un análisis de mediación con intervalos de confianza obtenidos por *bootstrap*, y el eventual papel moderador sobre el efecto de la política se analizó con modelos lineales generales con el término de interacción correspondiente, tanto para la obligación moral como, de forma exploratoria, para el uso habitual de IA, interpretando los resultados con la cautela que impone la potencia estadística del estudio.

Consideraciones éticas

El estudio se realizó de forma anónima y voluntaria. Antes de iniciar la tarea, los participantes aceptaron un consentimiento informado en el que se explicaba la finalidad académica del estudio, el tipo de datos recogidos y la posibilidad de abandonar la participación en cualquier momento. Los correos electrónicos, cuando se proporcionaron, se almacenaron separadamente de las respuestas y solo se utilizaron para comunicaciones relativas al estudio. La interacción con el asistente de IA se limitó al fragmento seleccionado por el participante o a la generación de una frase breve dentro de la tarea.

Resultados

Depuración de los datos y validación de los registros

Se excluyeron del análisis 10 casos en los que la suma del porcentaje declarado como generado y como parafraseado por IA superaba el 100%, un valor matemáticamente imposible, con lo que la muestra de análisis quedó en los 180 participantes ya descritos; los casos excluidos se distribuían de forma homogénea entre condiciones, de modo que su eliminación no introduce sesgos sistemáticos. Como comprobación adicional de la calidad de los registros se contrastaron las trazas conductuales del sistema (registro automático) con las declaraciones de los participantes (autoinforme). El uso del botón de IA coincidió entre ambas fuentes en todos los casos registrados en la plataforma, mientras que nueve participantes declararon haber recurrido a la IA sin dejar traza alguna en el entorno controlado, lo que delimita desde el inicio el alcance de la medida objetiva: captura el uso dentro de la plataforma, pero no el realizado mediante herramientas externas.

Comprobación de la manipulación

Antes de atribuir a la política cualquier diferencia conductual posterior, se verificó que las condiciones fueron percibidas como distintas analizando la restrictividad percibida en función de la condición asignada. Dada la heterogeneidad de varianzas (Levene: $F(2, 177) = 7{,}00$; $p = {,}001$), se recurrió a la corrección de Welch, que reveló diferencias significativas entre condiciones ($F(2, 115) = 7{,}87$; $p < {,}001$). Las comparaciones de Games-Howell mostraron que la condición restrictiva se percibió como significativamente más restrictiva ($M = 3{,}23$; $DE = 2{,}13$) que la permisiva ($M = 1{,}94$; $DE = 1{,}46$; $p < {,}001$) y que la difusa ($M = 2{,}16$; $DE = 1{,}54$; $p = {,}006$), mientras que la permisiva y la difusa no difirieron entre sí ($p = {,}701$). El dato es revelador, porque la condición difusa, concebida como un punto intermedio de ambigüedad, fue percibida prácticamente igual que la permisiva, lo que sugiere que ante la ausencia de una prohibición explícita los participantes interpretaron la política como permisiva por defecto, un primer indicio del riesgo moral que la ambigüedad introduce (véase Tabla B3).

Efecto de la política sobre el uso de IA

El uso total declarado de IA, suma del porcentaje generado y parafraseado, se analizó de nuevo con la corrección de Welch ante la fuerte heterocedasticidad observada (Levene: $F(2, 177) =$

28,7; $p < ,001$). El contraste reveló un efecto significativo de la condición ($F(2, 99,5) = 11,1$; $p < ,001$), con un tamaño del efecto moderado ($\omega^2 = ,080$; $\eta^2 = ,091$, obtenidos del ANOVA clásico al no proporcionarlos directamente la corrección de Welch), de manera que la condición asignada explica en torno al 8% de la variabilidad en el uso declarado. Las medias siguen el orden previsto, con el mayor uso en la condición permisiva ($M = 26,21\%$; $DE = 37,0$), un valor intermedio en la difusa ($M = 13,86\%$; $DE = 32,7$) y un uso casi nulo en la restrictiva ($M = 3,67\%$; $DE = 14,6$), aunque las comparaciones de Games-Howell solo alcanzaron significación entre la permisiva y la restrictiva (diferencia = 22,5 puntos; $p < ,001$), quedando las diferencias de la difusa con ambos extremos por debajo del umbral (permisiva: $p = ,132$; restrictiva: $p = ,086$) (véase Tabla B4). En conjunto, la prohibición explícita suprime de forma eficaz el uso declarado, que cae de un 26% a menos de un 4%, mientras que la difusa ocupa una posición que no se distingue estadísticamente de ninguno de los extremos, en coherencia con su percepción equivalente a la permisiva.

El análisis del uso conductual objetivo, es decir, de los caracteres efectivamente insertados mediante el botón de IA, replicó el patrón con la misma corrección (Levene: $F(2, 177) = 30,9$; $p < ,001$; Welch: $F(2, 87,6) = 11,3$; $p < ,001$). Los participantes de la condición permisiva insertaron mucho más texto vía IA ($M = 82,43$; $DE = 142,4$) que los de la restrictiva ($M = 8,38$; $DE = 37,1$; $p < ,001$), pero, a diferencia de lo observado en el uso declarado, aquí la condición difusa ($M = 67,33$; $DE = 150,6$) sí se diferenció de la restrictiva (p = ,015) y resultó indistinguible de la permisiva ($p = ,840$) (véase Tabla B5). Este contraste entre ambas medidas es uno de los hallazgos más reveladores, porque en términos de conducta real los participantes expuestos a una política difusa se comportaron como los de la condición permisiva, pero al declarar su uso lo situaron en niveles próximos a los de la condición restrictiva, una brecha entre lo que se hace y lo que se reconoce que constituye una manifestación empírica directa del desacoplamiento moral, en la medida en que el vacío normativo permite usar la IA sin asumir explícitamente haberlo hecho. Las desviaciones típicas elevadas en las dos condiciones más permisivas responden a un patrón de adopción binaria, en el que una parte de los participantes no usó el botón en ningún momento y otra externalizó una proporción sustancial de la redacción, de modo que las medias capturan la propensión agregada al uso dentro de cada condición y no una distribución continua. El volumen de texto pegado, por su parte, no difirió de forma significativa entre condiciones (Welch: $F(2, 107) = 2,58$; $p = ,081$), lo que indica que el efecto de la política se concentró en el uso de la herramienta integrada y no en otras vías de incorporación de contenido.

Uso de IA externa

Dado que la plataforma solo registraba el uso del botón integrado, se examinó por separado el uso declarado de herramientas de IA externas. Nueve participantes, el 5% de la muestra, declararon haber recurrido a ellas, con una distribución descendente a través de las condiciones, del 9,5% en la permisiva al 3,5% en la difusa y al 1,7% en la restrictiva, un gradiente coherente con el efecto general de la política, aunque la prueba de chi-cuadrado no alcanzara significación dado el reducido número de casos ($\chi^2(2) = 4{,}38$; $p = {,}112$; V de Cramer = ,156) (véase Tabla B6). El escaso número impide cualquier inferencia firme, pero el patrón sugiere que la influencia de la política se extiende, siquiera de forma tendencial, incluso a las herramientas que escapan al control del entorno, lo que enlaza con la asociación entre la oportunidad percibida y el uso externo que se detalla más adelante.

Señales conductuales del uso de IA

Las correlaciones de Spearman entre las variables de esfuerzo, elegidas por su robustez ante distribuciones sesgadas, ofrecen una vía de identificación del uso de IA independiente del autoinforme. El número de ediciones manuales correlacionó negativamente con el uso declarado ($\rho = -{,}521$; $p < {,}001$) y con el uso objetivo del botón ($\rho = -{,}358$; $p < {,}001$), de manera que quienes más teclearon fueron quienes menos IA emplearon, mientras que el uso objetivo y el declarado mostraron entre sí una correlación elevada ($\rho = {,}735$; $p < {,}001$) que confirma la consistencia entre conducta registrada y autorreporte. El número de palabras no se asoció con ninguna medida de uso (todos los $\rho < {,}10$), resultado esperable dado el rango de extensión acotado por el diseño, lo que deja a las ediciones manuales como la señal conductual más discriminante del uso real de IA disponible en los datos (véase Tabla B7).

Predictores psicológicos del uso de IA

Para evaluar la capacidad explicativa de los factores individuales se estimaron dos modelos de regresión sobre el uso declarado, ambos controlando por la condición experimental. El primero incorporó los componentes clásicos de la Teoría del Comportamiento Planificado, esto es, las normas subjetivas descriptiva e injuntiva, el control conductual percibido en sus dos caras de evasión de la detección y de capacidad técnica, y la racionalización utilitaria como aproximación a la actitud, dado que la batería no incluyó un ítem de actitud independiente y este recoge la evaluación favorable del comportamiento que constituye el núcleo de ese

constructo. El modelo resultó significativo ($F(7, 172) = 3{,}76$; $p < ,001$; $R^2 = ,133$), pero su capacidad explicativa recae principalmente en la condición experimental, ya que entre los constructos propiamente psicológicos solo la norma descriptiva alcanzó significación ($\beta = ,208$; $p = ,008$), mientras que la norma injuntiva, ambas facetas del control conductual percibido y la actitud no contribuyeron de forma apreciable (todas $p > ,30$). Este patrón respalda H2, puesto que, una vez aislado el efecto de la política, los componentes clásicos de la TPB apenas predicen el uso no autorizado de IA, con la única excepción de percibir que los iguales la utilizan.

El segundo modelo amplió el anterior con los factores del nivel situacional y del desacoplamiento moral, incorporando la oportunidad percibida, la racionalización normativa y la obligación moral, junto con el comportamiento pasado, medido como frecuencia habitual de uso, y la orientación motivacional. La varianza explicada ascendió del 13,3% al 21,8% ($F(12, 167) = 3{,}88$; $p < ,001$; $R^2 = ,218$), un incremento de 8,5 puntos porcentuales que resultó estadísticamente significativo ($\Delta R^2 = ,085$; $F(5, 167) = 3{,}64$; $p = ,004$), lo que confirma H3 y cuantifica la aportación de la arquitectura multinivel por encima de la TPB clásica. Entre los predictores incorporados resultaron significativos la frecuencia habitual de uso ($\beta = ,191$; $p = ,044$) y la racionalización normativa ($\beta = ,231$; $p = ,009$), mientras que la oportunidad percibida quedó en el umbral ($\beta = ,144$; $p = ,058$) y ni la obligación moral ni la orientación motivacional aportaron peso propio. La racionalización utilitaria, no significativa en el primer modelo (como proxy de actitud), pasó a serlo con signo negativo al introducir la normativa ($\beta = -,186$; $p = ,032$), un efecto de supresión esperable dada la fuerte correlación entre ambos índices ($\rho = ,491$) que aconseja interpretarlo con cautela y no como una contribución sustantiva independiente. En conjunto, el uso de IA queda mejor explicado por el hábito previo y la racionalización normativa que por las actitudes y normas de la TPB clásica, lo que sostiene la lógica acumulativa del modelo teórico (véase Tabla B9).

El Triángulo del Fraude

El examen correlacional de los tres vértices del Triángulo del Fraude muestra que operan de forma selectiva y no equivalente. La presión académica no predijo directamente el uso declarado ($\rho = ,121$; $p = ,105$) ni el del botón ($\rho = ,116$; $p = ,120$), pero sí correlacionó con ambos índices de racionalización (utilitaria: $\rho = ,244$; normativa: $\rho = ,259$; ambas $p < ,001$), lo que la sitúa como antecedente de la justificación antes que como causa directa de la conducta. La oportunidad percibida tampoco predijo el uso declarado ni el del botón, y solo se asoció con

el uso de IA externa (ρ = ,177; p = ,017), sin correlacionar con ninguna otra variable del triángulo (todas p > ,60), de modo que parece operar como un factor independiente ligado al uso menos vigilado y no como parte de una cadena motivacional integrada. De los mecanismos de racionalización, únicamente la normativa (la percepción de que el uso de IA es una norma compartida entre iguales) correlacionó con el uso declarado (ρ = ,207; p = ,005) y con el objetivo (ρ = ,173; p = ,020), mientras que la utilitaria, pese a su fuerte relación con la normativa (ρ = ,491) y con la presión (ρ = ,244), no se tradujo en conducta. Las variables morales, a su vez, actúan como contrapeso sistemático de la racionalización, ya que la internalización, la culpa y los principios correlacionan negativamente con ambos índices (ρ entre −,303 y −,469; todas p < ,001), lo que confirma que el desacoplamiento moral no es arbitrario sino inversamente proporcional al desarrollo moral del individuo, en línea con la teoría de Bandura (véase Tabla B8). El conjunto de estas relaciones dibuja una posible secuencia en la que la presión activa la racionalización normativa y esta se traduce en uso, una hipótesis que el siguiente apartado somete a contraste formal.

Mediación: la racionalización como vía de la presión

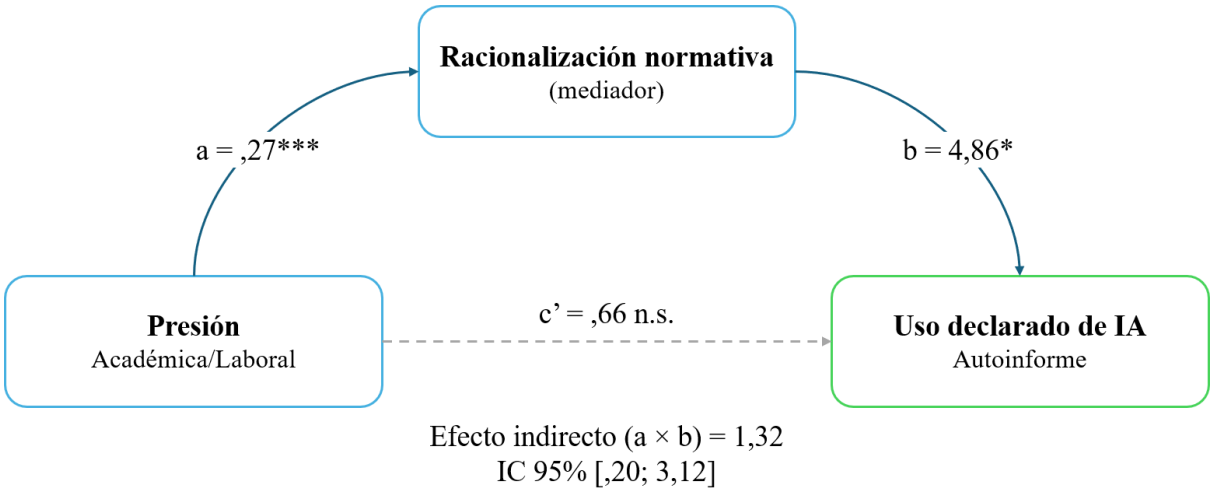
Para contrastar H4 se estimó un modelo de mediación simple con la presión como variable predictora, la racionalización normativa como mediadora y el uso de IA como variable dependiente, con intervalos de confianza obtenidos mediante *bootstrap* de percentiles y 5.000 remuestreos. Sobre el uso declarado, el efecto indirecto de la presión a través de la racionalización resultó significativo según el intervalo de confianza (b = 1,32; IC 95% [0,20; 3,12]), mientras que el efecto directo de la presión sobre el uso no alcanzó significación (b = 0,66; IC 95% [−2,99; 3,93]; p = ,701), lo que configura un patrón compatible con una mediación completa en el que la presión solo se traduce en uso declarado cuando es canalizada por la racionalización. La descomposición de las vías sustenta el mecanismo, ya que la presión predice la racionalización normativa (a = 0,27; p = ,001) y esta, a su vez, predice el uso declarado (b = 4,86; p = ,013), de manera que el producto de ambas explica el efecto indirecto (véase Tabla B10). Conviene precisar que el contraste normal de ese efecto indirecto quedó justo en el umbral (Z = 1,80; p = ,072), por lo que se adopta como criterio el intervalo *bootstrap*, que es el procedimiento de referencia para un efecto cuya distribución muestral es asimétrica.

Cuando la variable dependiente fue el uso conductual objetivo, el patrón se mantuvo en su dirección, pero perdió significación, dado que el efecto indirecto fue de magnitud comparable (b = 3,40) pero su intervalo de confianza incluyó por un margen mínimo el cero (IC 95%

[−0,13; 8,60]) y la vía que une la racionalización con la conducta objetiva quedó también en el umbral (b = 12,48; *p* = ,061) (véase Tabla B11). La mediación se confirma, por tanto, sobre lo que el estudiante declara, pero no sobre lo que la plataforma registra, una asimetría coherente con la idea de que la racionalización opera ante todo como mecanismo de justificación del uso reconocido y no tanto sobre la conducta efectiva. Para descartar que el resultado estuviera confundido por la condición experimental se replicó la mediación controlando por la política mediante un modelo lineal general de mediación, y el efecto indirecto sobre el uso declarado se mantuvo con un intervalo que excluía el cero (b = 1,31; IC 95% [0,22; 2,92]), sin que la política ejerciera efecto alguno a través de la racionalización, lo que confirma que el mecanismo es robusto al control de la variable de tratamiento. El patrón de mediación obtenido se ilustra en la Figura 2.

Figura 2

Modelo de mediación de la racionalización normativa entre la presión y el uso declarado de IA



Nota. Coeficientes no estandarizados. La línea continua indica un efecto significativo y la discontinua un efecto no significativo. El efecto indirecto (a × b) se interpreta a partir del intervalo de confianza *bootstrap*, que excluye el cero. Sobre el uso objetivo de IA la mediación no resultó significativa (véase Tabla B11). N = 180. n.s. = no significativo. * $p < ,05$; *** $p < ,001$.

Moderación: la obligación moral no condiciona el efecto de la política

El contraste de H5 se realizó mediante un modelo lineal general que incorporaba la política, la obligación moral, medida como un compuesto de internalización, culpa y principios con una fiabilidad alta (α = ,872), y la interacción entre ambas. Sobre el uso declarado, el efecto de la política fue significativo (F(2, 174) = 9,00; p < ,001; η^2 = ,093), pero ni la obligación moral (F(1, 174) = 0,96; p = ,329) ni, sobre todo, la interacción entre política y obligación moral (F(2, 174) = 0,09; p = ,912; η^2 = ,001) resultaron significativas. El análisis sobre el uso objetivo reprodujo el mismo patrón, con un efecto significativo de la política (F(2, 174) = 6,49; p = ,002; η^2 = ,069) y una interacción de nuevo no significativa (F(2, 174) = 0,37; p = ,693; η^2 = ,004) (véase Tabla B12). Una comprobación exploratoria adicional, en la que el uso habitual de IA se incorporó como segundo factor, confirmó que el efecto de la política tampoco resultó moderado por el perfil de usuario, ya que la interacción entre política y condición de usuario frecuente no alcanzó significación (F(2, 174) = 1,22; p = ,298; η^2 = ,012), pese a que tanto la política (F(2, 174) = 6,75; p = ,002) como el uso habitual (F(1, 174) = 6,21; p = ,014) mostraron efectos principales significativos (véase Tabla B13), lo que refuerza que la política opera de manera uniforme sobre perfiles individuales distintos. El efecto de la política sobre el uso de IA es, en consecuencia, independiente del nivel de obligación moral del individuo, de modo que incluso los estudiantes con mayor desarrollo moral responden a la política del mismo modo que el resto, un resultado que replica el hallazgo de Heriyati y Ekasari (2020) en el contexto de la IA y que traslada la palanca de la intervención desde el carácter del estudiante hacia el diseño del entorno normativo.

Contraste de hipótesis

En conjunto, los resultados confirman H1, en tanto la política ejerce un efecto sobre el uso de IA y la condición difusa se comporta como la permisiva en lugar de como un punto intermedio; respaldan H2 y H3, dado que la TPB clásica resulta insuficiente por sí sola y el modelo ampliado con hábito y racionalización mejora de forma sustantiva la predicción; sostienen H4 de forma parcial, ya que la mediación de la presión a través de la racionalización se confirma sobre el uso declarado pero no sobre el objetivo; y son compatibles con H5, puesto que la obligación moral no modera el efecto de la política, que se mantiene constante con independencia del nivel moral del individuo.

Discusión

El objetivo de este trabajo era determinar si la política institucional ejerce un efecto causal sobre el uso de inteligencia artificial y qué factores conductuales explican ese uso una vez aislado dicho efecto. El resultado central matiza la intuición de partida, porque la política importa, y mucho, pero no del modo lineal que cabría esperar, dado que solo la prohibición explícita suprimió el uso mientras que la condición difusa, lejos de situarse en un punto intermedio de cautela, fue percibida y respondida prácticamente como la permisiva, de manera que la ausencia de una norma clara no se interpretó como una señal de prudencia sino como una autorización implícita. Este hallazgo, que ya anticipaba la comprobación de la manipulación y que el comportamiento objetivo confirmó al situar a la condición difusa junto a la permisiva, puede interpretarse como una manifestación empírica del riesgo moral que Smit et al. (2025) y Benayoune et al. (2026) atribuyen a la ambigüedad normativa, y sugiere que el vacío regulatorio no es un terreno neutro de indefinición sino un facilitador estructural del uso encubierto.

A esta lectura se añade un matiz que el doble registro de la conducta permite captar y que un diseño basado solo en autoinforme habría pasado por alto. En la condición difusa, el uso real del botón de IA se equiparó al de la condición permisiva, pero el uso declarado se situó en niveles próximos a los de la condición restrictiva, de modo que los participantes expuestos a la ambigüedad hicieron tanto como los de la permisiva pero reconocieron tan poco como los de la restrictiva. Esa brecha entre lo que se hace y lo que se admite es, en los términos de la teoría de Bandura (1999), una manifestación directa del desacoplamiento moral, porque el vacío normativo ofrece al estudiante la cobertura cognitiva necesaria para usar la herramienta sin asumir explícitamente haberlo hecho y lo libera de la disonancia que una prohibición clara habría activado.

El segundo bloque de resultados respalda la decisión teórica de no detenerse en la Teoría del Comportamiento Planificado. Una vez controlada la condición experimental, los constructos clásicos del modelo apenas predijeron el uso de IA, con la única excepción de la norma descriptiva, es decir, la percepción de que los iguales recurren a la herramienta; mientras que la norma injuntiva, el control conductual percibido y la actitud no aportaron capacidad explicativa apreciable. Este patrón es coherente con la tesis, desarrollada en el marco teórico, de que en el uso no autorizado de IA la conciencia ética existe, pero no llega a traducirse en conducta, y de que la predicción mejora cuando se incorporan los mecanismos que median esa

desconexión. La ganancia de ocho puntos y medio de varianza explicada al pasar del modelo clásico al ampliado, sostenida por el hábito previo y la racionalización normativa, cuantifica esa aportación y confirma que la arquitectura multinivel no es una complicación innecesaria sino una condición para explicar el fenómeno, en la línea de lo que Chudzicka-Czupała et al. (2016) observaron al añadir la obligación moral a la TPB.

Dentro de esa arquitectura, los tres vértices del Triángulo del Fraude no operaron de forma equivalente, lo que refina la comprensión del fenómeno más allá de la aplicación mecánica del modelo. La presión académica no se tradujo directamente en conducta, sino que actuó como antecedente de la racionalización y solo a través de ella alcanzó el comportamiento, tal como confirmó el análisis de mediación sobre el uso declarado. La oportunidad percibida, por su parte, operó como un factor independiente que no se integró en la cadena motivacional y que únicamente se asoció con el uso de herramientas externas, las menos vigiladas, lo que concuerda con la redefinición del vértice de oportunidad como indetectabilidad algorítmica que propone Başer et al. (2026). Y de los dos mecanismos de racionalización, solo la normativa, la percepción de que el uso de IA es una norma compartida entre iguales, se tradujo en conducta, mientras que las variables morales actuaron como contrapeso sistemático de la racionalización, lo que confirma que el desacoplamiento no es arbitrario sino inversamente proporcional al desarrollo moral del individuo.

Conviene subrayar, no obstante, que esta mediación se confirmó sobre el uso declarado pero no sobre el objetivo, una asimetría que aconseja prudencia interpretativa y que admite una lectura sustantiva, ya que si la racionalización canaliza la presión hacia lo que el estudiante reconoce pero no necesariamente hacia lo que hace, es razonable entender que opera ante todo como un mecanismo de justificación del uso admitido, una narrativa que el individuo construye para reconciliar su conducta con su autoimagen, antes que como un motor directo de la acción. En cualquier caso, dado que la mediación descansa sobre datos transversales, su interpretación debe entenderse como compatible con la secuencia propuesta y no como una demostración causal, que en este estudio queda reservada en exclusiva al efecto de la política, lo único manipulado experimentalmente.

El resultado relativo a la obligación moral cierra el argumento. El efecto de la política se mantuvo constante con independencia del nivel moral del individuo, sin que la obligación moral lo moderara ni en el uso declarado ni en el objetivo, y la misma robustez se observó frente al uso habitual de la herramienta. Esto significa que incluso los estudiantes con mayor

desarrollo moral respondieron a la política del mismo modo que el resto, un patrón que replica en el contexto de la IA el hallazgo de Heriyati y Ekasari (2020) según el cual el razonamiento moral no modera la materialización del fraude, y que desplaza el foco de la intervención desde una cuestión de carácter individual hacia una cuestión de diseño del entorno, porque si la integridad personal no protege frente a una política permisiva o ambigua, la palanca efectiva no está en apelar a la conciencia del estudiante sino en configurar el marco normativo en el que decide.

Implicaciones

De este conjunto de resultados se desprende una implicación práctica que se aparta de la respuesta intuitiva ante el Shadow AI. La prohibición explícita suprime el uso declarado, pero la literatura advierte de que tiende a desplazarlo hacia la clandestinidad (Tsigaris y Teixeira da Silva, 2026), y la ambigüedad, lejos de ofrecer una vía intermedia prudente, se interpreta como permiso y reproduce el uso de la condición permisiva al tiempo que reduce su declaración, de modo que ninguna de las dos respuestas habituales, ni prohibir sin más ni dejar la cuestión a criterio del estudiante, resuelve el problema. La evidencia apunta a que la palanca eficaz no reside en la severidad de la norma ni en la apelación a la integridad, sino en la claridad y la previsibilidad de unas reglas que reduzcan la oportunidad percibida rediseñando las propias tareas de evaluación, en la línea de lo que Kangwa et al. (2025) identifican como las políticas que efectivamente disminuyen el fraude. Que la obligación moral no module el comportamiento refuerza esta conclusión, porque sitúa la responsabilidad del resultado en el diseño institucional antes que en la virtud individual del estudiante.

Limitaciones

Los resultados deben interpretarse a la luz de varias limitaciones. La mayoría de los constructos se midieron mediante ítems únicos, una decisión que respondía a la necesidad de un instrumento breve y obliga a leer con cautela los coeficientes asociados a cada predictor, salvo en el caso de la obligación moral, medida con varios ítems y con una fiabilidad alta. A esta cautela se añade que la batería no incorporó un ítem específico de actitud, de modo que ese componente de la Teoría del Comportamiento Planificado se aproximó mediante la racionalización utilitaria, que recoge la evaluación favorable del uso de la herramienta pero que, al pertenecer también al plano de la justificación cognitiva, no constituye una medida pura de actitud, por lo que el contraste del modelo clásico debe entenderse como una aproximación

y no como una prueba estricta de la TPB en su formulación canónica. El tamaño muestral por condición, suficiente para contrastar efectos principales, ofrece además una potencia limitada para detectar interacciones, de modo que el papel moderador no significativo de la obligación moral debe entenderse como ausencia de evidencia y no como prueba concluyente de ausencia de moderación. Salvo el efecto de la política, las relaciones examinadas son correlacionales y se basan en datos transversales, por lo que la mediación describe una asociación compatible con la secuencia teórica y no una cadena causal demostrada. A ello se suma que la medida objetiva capturaba únicamente el uso de la herramienta integrada y no el de herramientas externas; los modelos de regresión se estimaron por MCO sobre una variable dependiente con fuerte concentración en cero, por lo que los coeficientes deben leerse con cautela, aunque su dirección resultó coherente con los contrastes no paramétricos previos; y la tarea, breve y realizada en un entorno controlado, no reproduce las condiciones de presión y de consecuencias reales de un trabajo académico evaluable, lo que acota la validez ecológica de los hallazgos. Por último, la equivalencia observada entre las condiciones difusa y permisiva, aunque teóricamente relevante, depende de la formulación concreta de los textos, de manera que el resultado debe entenderse referido a este tipo particular de ambigüedad, en el que dejar el uso a criterio del estudiante comunica de hecho un permiso, y no necesariamente a cualquier formulación de política ambigua.

Líneas futuras

Estas limitaciones señalan, a su vez, las direcciones más prometedoras para la investigación posterior. Sería valioso replicar el diseño con una muestra de mayor tamaño que proporcione potencia suficiente para contrastar la moderación de los factores conductuales que aquí solo pudieron explorarse, y con instrumentos de varios ítems que permitan estimar la fiabilidad de cada constructo. Resultaría igualmente esclarecedor un seguimiento longitudinal del efecto de políticas institucionales reales, que superase las limitaciones de validez ecológica de una tarea puntual, así como una exploración más fina de la brecha entre uso real y uso declarado. Y dado que la condición difusa se interpretó como permisiva, convendría poner a prueba formulaciones de política genuinamente intermedias, capaces de comunicar incertidumbre sin equivaler a un permiso, para determinar si existe un punto de la escala normativa capaz de moderar el comportamiento sin desplazarlo hacia la clandestinidad ni legitimarlo por defecto.

Conclusión

La política importa, pero no de forma lineal ya que solo la prohibición explícita suprimió el uso, mientras que la ambigüedad fue interpretada y respondida como una autorización implícita, reproduciendo la conducta de la condición permisiva al tiempo que reducía su declaración. Esa brecha entre lo que se hace y lo que se reconoce constituye la manifestación empírica más directa del Shadow AI. Dado que ni la severidad de la norma ni la integridad moral del individuo bastaron para contener el uso encubierto, la palanca eficaz no reside en apelar a la conciencia del estudiante, sino en el diseño de un entorno normativo claro y previsible que reduzca la oportunidad percibida rediseñando las propias tareas de evaluación. El Shadow AI, en definitiva, es menos un problema de carácter que de arquitectura institucional, y es en esa arquitectura, y no en la virtud individual, donde una universidad decide cuánto uso encubierto está dispuesta a generar.

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

Por la presente, yo, Patricia Martín Martínez, estudiante de G-PS + E-2 de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado: Políticas institucionales y *Shadow AI*: evidencia experimental sobre el efecto causal de la regulación en el uso de inteligencia artificial generativa", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
4. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
5. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
6. Redactor de código: Para elaborar código en el contexto de un TFG de ADE cuyo programa es un medio para obtener datos y no un fin en sí.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para qué se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 03/06/2026

Firma: Patricia Martín Martínez

Referencias bibliográficas

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alneyadi, S., y Wardat, Y. (2023). *ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools*. *Contemporary Educational Technology, 15*(4), Article ep448. https://doi.org/10.30935/cedtech/13417

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., y Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy, 19*(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. *Heliyon, 10*(9), Article e30696. https://doi.org/10.1016/j.heliyon.2024.e30696

Askarkyzy, S., y Zhunusbekova, A. (2024). Students' perceptions of artificial intelligence use in higher education and its impact on academic integrity. *Pedagogy and Psychology, 4*(61), 145–155. https://doi.org/10.51889/2960-1649.2024.61.4.008

Ayub, T., Ahmad Malla, R., Khan, M. Y., y Ganaie, S. A. (2024). The art of deception: Humanizing AI to outsmart detection. *Global Knowledge, Memory and Communication*. Advance online publication. https://doi.org/10.1108/GKMC-03-2024-0133

Bahroun, Z., Anane, C., Ahmed, V., y Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability, 15*(17), Article 12983. https://doi.org/10.3390/su151712983

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*(3), 193–209. https://doi.org/10.1207/s15327957pspr0303_3

Başer, M. Y., Kozak, M., y Erdoğan, İ. H. (2026). Do we worry about the use of artificial intelligence and plagiarism? Students' AI-giarism behaviour through the fraud triangle. *The Internet and Higher Education, 69*, Article 101071. https://doi.org/10.1016/j.iheduc.2025.101071

Beck, L., y Ajzen, I. (1991). Predicting dishonest actions using the Theory of Planned Behavior. *Journal of Research in Personality, 25*(3), 285–301. https://doi.org/10.1016/0092-6566(91)90021-H

Becker, D., Connolly, J., Lentz, P., y Morrison, J. (2006). Using the business fraud triangle to predict academic dishonesty among business students. *Academy of Educational Leadership Journal, 10*(1), 37–54.

Behrens, S. (2009). Shadow systems: The good, the bad and the ugly. *Communications of the ACM, 52*(2), 124–129. https://doi.org/10.1145/1461928.1461960

Bélanger, C. H., Leonard, V. M., y LeBrasseur, R. (2012). Moral reasoning, academic dishonesty, and business students. *International Journal of Higher Education, 1*(1), 72–89. https://doi.org/10.5430/ijhe.v1n1p72

Benayoune, A., Slimi, Z., y Al Habsi, A. (2026). Artificial intelligence policy challenges and institutional readiness in Omani higher education. *Discover Education, 5*, Article 86. https://doi.org/10.1007/s44217-026-01188-4

Chan, C. K. Y. (2023). Is AI changing the rules of academic misconduct? An in-depth look at students' perceptions of "AI-giarism". *arXiv*. https://doi.org/10.48550/arXiv.2306.03358

Chen, P., Fan, Z., Lu, Y., y Xu, Q. (2024). PBChat: Enhance student's problem behavior diagnosis with large language model. En A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, y I. I. Bittencourt (Eds.), *Artificial intelligence in education: 25th International Conference, AIED 2024, Recife, Brazil, July 8–12, 2024, Proceedings, Part I* (Lecture Notes in Computer Science, Vol. 14829, pp. 32–45). Springer. https://doi.org/10.1007/978-3-031-64302-6_3

Chen, S., y Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis. *Educational Research Review, 49*, Article 100737. https://doi.org/10.1016/j.edurev.2025.100737

Chen, S., Lan, Y., y Yuan, Z. (2024). A multi-task automated assessment system for essay scoring. En A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, y I. I. Bittencourt (Eds.), *Artificial intelligence in education: 25th International Conference, AIED 2024, Recife, Brazil, July 8–12, 2024, Proceedings, Part II* (Lecture Notes in Computer Science, Vol. 14830, pp. 276–283). Springer. https://doi.org/10.1007/978-3-031-64299-9_22

Chudzicka-Czupała, A., Grabowski, D., Mello, A. L., Kuntz, J., Zaharia, D. V., Hapon, N., Lupina-Wegener, A., y Börü, D. (2016). Application of the theory of planned behavior in academic cheating research–cross-cultural comparison. *Ethics & Behavior, 26*(8), 638–659. https://doi.org/10.1080/10508422.2015.1112745

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., y Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Galindo-Domínguez, H., Sainz-de-la-Maza, M., Campo, L., y Losada-Iglesias, D. (2026). Influencia de la motivación hacia el aprendizaje y la procrastinación en la dependencia a ChatGPT [The influence of motivation for learning and procrastination on ChatGPT dependence]. *RIED-Revista Iberoamericana de Educación a Distancia, 29*(1). https://doi.org/10.5944/ried.29.1.45497

Györy, A., Cleven, A., Uebernickel, F., y Brenner, W. (2012). Exploring the shadows: IT governance approaches to user-driven innovation. En *ECIS 2012 Proceedings* (Article 222). AIS Electronic Library. https://aisel.aisnet.org/ecis2012/222

Ha, S. T., Phan, T. T. H., Ngo, T. V. N., Duong, C. D., y Ha, N. T. (2025). Integrating artificial intelligence competencies into the theory of planned behavior: Explaining sustainability-oriented entrepreneurial intentions. *Journal of Entrepreneurship, Management and Innovation, 21*(4), 30–53. https://doi.org/10.7341/20252142

Harding, T., Mayhew, M., Finelli, C., y Carpenter, D. (2007). The theory of planned behavior as a model of academic dishonesty in engineering and humanities undergraduates. *Ethics & Behavior, 17*(3), 255–279. https://doi.org/10.1080/10508420701519239

He, R., Cao, J., y Tan, T. (2025). Generative artificial intelligence: A historical perspective. *National Science Review, 12*(5), Article nwaf050. https://doi.org/10.1093/nsr/nwaf050

Heriyati, D., y Ekasari, W. F. (2020). A study on academic dishonesty and moral reasoning. *International Journal of Education, 12*(2), 56–62. https://doi.org/10.17509/ije.v12i2.18653

Ivanov, S., Soliman, M., Tuomi, A., Alkathiri, N. A., y Al-Alawi, A. N. (2024). Drivers of generative AI adoption in higher education through the lens of the theory of planned behaviour. *Technology in Society, 77*, Article 102521. https://doi.org/10.1016/j.techsoc.2024.102521

Jazim, F., Al-Mamary, Y. H., y Abubakar, A. A. (2025). Developing an integrated model to explore key factors influencing university students' behavioral intentions to use ChatGPT in enhancing higher education in the Hail region. *Acta Psychologica, 259*, Article 105445. https://doi.org/10.1016/j.actpsy.2025.105445

Jin, Y., Yan, L., Echeverria, V., Gašević, D., y Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence, 8*, Article 100348. https://doi.org/10.1016/j.caeai.2024.100348

Kangwa, D., Msafiri, M. M., y Fute, A. (2025). Exploring the factors that promote a balance between academic integrity and the effective use of GenAI tools in higher education: A systematic review. *Journal of Computer Assisted Learning, 41*(5), Article e70109. https://doi.org/10.1111/jcal.70109

Kaya-Kasikci, S., Glass, C. R., Chacon Camero, E., y Minaeva, E. (2025). University positioning in AI policies: Comparative insights from national policies and non-state actor influences in China, the European Union, India, Russia, and the United States. *Higher Education Quarterly, 79*, Article e70062. https://doi.org/10.1111/hequ.70062

Lewellyn, P. G., y Rodriguez, L. C. (2015). Does academic dishonesty relate to Fraud Theory? A comparative analysis. *American International Journal of Contemporary Research, 5*(3), 1–6.

Liang, Z., Sha, L., Tsai, Y.-S., Gašević, D., y Chen, G. (2024). Towards the automated generation of readily applicable personalised feedback in education. En A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, y I. I. Bittencourt (Eds.), *Artificial intelligence in education: 25th International Conference, AIED 2024, Recife, Brazil, July 8–12, 2024,*

Proceedings, Part II (Lecture Notes in Computer Science, Vol. 14830, pp. 75–88). Springer. https://doi.org/10.1007/978-3-031-64299-9_6

Liu, J. Q. J., Hui, K. T. K., Al Zoubi, F., Zhou, Z. Z. X., Samartzis, D., Yu, C. C. H., Chang, J. R., y Wong, A. Y. L. (2024). The great detectives: Humans versus AI detectors in catching large language model-generated medical writing. *International Journal for Educational Integrity, 20*, Article 8. https://doi.org/10.1007/s40979-024-00155-6

Lyu, W., Wang, Y., Chung, T. R., Sun, Y., y Zhang, Y. (2024). Evaluating the effectiveness of LLMs in introductory computer science education: A semester-long field study. En *Proceedings of the 11th ACM Conference on Learning @ Scale* (pp. 63–74). Association for Computing Machinery. https://doi.org/10.1145/3657604.3662036

Ma, N., y Zhong, Z. (2025). A meta-analysis of the impact of generative artificial intelligence on learning outcomes. *Journal of Computer Assisted Learning, 41*(5), Article e70117. https://doi.org/10.1111/jcal.70117

Messer, M., Shi, M., Brown, N. C. C., y Kölling, M. (2024). Grading documentation with machine learning. En A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, y I. I. Bittencourt (Eds.), *Artificial intelligence in education: 25th International Conference, AIED 2024, Recife, Brazil, July 8–12, 2024, Proceedings, Part I* (Lecture Notes in Computer Science, Vol. 14829, pp. 105–117). Springer. https://doi.org/10.1007/978-3-031-64302-6_8

Nguyen, A., Kishore, S., Hong, Y., Qutab, S., y Dang, B. (2025). Generative artificial intelligence (AI) in education: From organizing visions to official guidelines. *Information Technology & People, 38*(8), 172–199. https://doi.org/10.1108/ITP-08-2024-1026

Pang, W., y Wei, Z. (2025). Shaping the future of higher education: A technology usage study on generative AI innovations. *Information, 16*(2), Article 95. https://doi.org/10.3390/info16020095

Portilla, J. E. N., Zapa Cedeño, J. K., León Jácome, G. O., y Manzano Gallegos, L. A. (2025). Systematic review: Artificial intelligence (AI) in Education 4.0. *Journal of Educators Online, 22*(3), Article 13. https://doi.org/10.9743/JEO.2025.22.3.13

Puthal, D., Mishra, A. K., Mohanty, S. P., Longo, A., y Yeun, C. Y. (2025). Shadow AI: Cyber security implications, opportunities and challenges in the unseen frontier. *SN Computer Science, 6*(5), Article 405. https://doi.org/10.1007/s42979-025-03962-x

Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., y Vega-Fleitas, E. (2023). ChatGPT challenges blended learning methodologies in engineering education: A case study in mathematics. *Applied Sciences, 13*(10), Article 6039. https://doi.org/10.3390/app13106039

Silic, M., y Back, A. (2014). Shadow IT: A view from behind the curtain. *Computers & Security, 45*, 274–283. https://doi.org/10.1016/j.cose.2014.06.007

Silic, M., Silic, D., y Kind, T. K. (2025). From Shadow IT to Shadow AI: Threats, risks and opportunities for organizations. *Strategic Change*, 1–16. https://doi.org/10.1002/jsc.2682

Smit, M., Wagner, R. F., y Bond-Barnard, T. J. (2025). Ambiguous regulations for dealing with AI in higher education can lead to moral hazards among students. *Project Leadership and Society, 6*, Article 100187. https://doi.org/10.1016/j.plas.2025.100187

Smith, K., Emerson, D., Haight, T., y Wood, B. (2023). An examination of online cheating among business students through the lens of the Dark Triad and Fraud Diamond. *Ethics & Behavior, 33*(6), 433–460. https://doi.org/10.1080/10508422.2022.2104281

Stadler, M., Bannert, M., y Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, Article 108386. https://doi.org/10.1016/j.chb.2024.108386

Stone, T., Jawahar, I. M., y Kisamore, J. (2010). Predicting academic misconduct intentions and behavior using the Theory of Planned Behavior and personality. *Basic & Applied Social Psychology*, *32*(1), 35–45. https://doi.org/10.1080/01973530903539895

Strzelecki, A., y ElArabawy, S. (2024). Investigation of the moderation effect of gender and study level on the acceptance and use of generative AI by higher education students: Comparative evidence from Poland and Egypt. *British Journal of Educational Technology, 55*(3), 1209–1230. https://doi.org/10.1111/bjet.13425

Tbaishat, D., Amoudi, G., y Elfadel, M. (2025). Adapting teaching and learning with existing generative AI by higher education students: Comparative study of Zayed University

and King Abdulaziz University. *Computers and Education: Artificial Intelligence, 8*, Article 100421. https://doi.org/10.1016/j.caeai.2025.100421

Trialih, R. (2023). The challenge in neutralizing Shadow IT: A literature review. En *2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 1169–1173). IEEE. https://doi.org/10.1109/IEEM58616.2023.10406586

Tsigaris, P., y Teixeira da Silva, J. A. (2026). AI detecting AI in academic writing: Why most AI detector findings are false. *Next Research, 7*, Article 101396. https://doi.org/10.1016/j.nexres.2026.101396

Wang, K. D., Burkholder, E., Wieman, C., Salehi, S., y Haber, N. (2024). Examining the potential and pitfalls of ChatGPT in science and engineering problem-solving. *Frontiers in Education, 8*, Article 1330486. https://doi.org/10.3389/feduc.2023.1330486

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., y Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*. https://openreview.net/forum?id=yzkSU5zdwD

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education, 39*(3), 235–274. https://doi.org/10.1023/A:1018724900565

Xia, Q., Li, W., Yang, Y., Weng, X., y Chiu, T. K. F. (2025). A systematic review and meta-analysis of the effectiveness of generative artificial intelligence (GenAI) on students' motivation and engagement. *Computers and Education: Artificial Intelligence, 9*, Article 100455. https://doi.org/10.1016/j.caeai.2025.100455

Zhai, C., Wibowo, S., y Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11, Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, L., y Xu, J. (2025). The paradox of self-efficacy and technological dependence: Unraveling generative AI's impact on university students' task completion. The Internet and Higher Education, 65, Article 100978. https://doi.org/10.1016/j.iheduc.2024.100978

Anexo A. Instrumento y condiciones experimentales

Tabla A1

Batería de ítems del cuestionario, constructo medido y fuente teórica

Nº	Ítem	Constructo medido	Función en el modelo y justificación operacional	Fuente
Bloque 1: Tu entorno y la IA (escala Likert 1-5, salvo ítem 8, formulado como escala de frecuencia)				
1	La mayoría de las personas de mi entorno usan IA regularmente en sus tareas o proyectos.	Norma subjetiva descriptiva	Percepción del uso de IA en el entorno inmediato; componente descriptivo de la norma subjetiva (lo que hacen los demás).	Ajzen (1991); Stone et al. (2010)
2	Las personas importantes para mí desaprobarían que usara IA sin declararla en un trabajo académico o profesional.	Norma subjetiva injuntiva	Aprobación o desaprobación percibida de los referentes significativos; componente injuntivo de la norma subjetiva.	Ajzen (1991); Beck y Ajzen (1991)
3	Si quisiera, podría usar IA en una tarea sin que nadie lo detectara.	Control conductual percibido	Control percibido sobre la **propia acción** de usar IA sin ser detectado. Se diferencia del ítem 5 en que evalúa la capacidad del individuo para eludir el control, no la debilidad del entorno; por eso se asigna al control conductual percibido y no a la oportunidad.	Ajzen (1991); Başer et al. (2026)
4	Tengo los conocimientos necesarios para usar IA de forma eficaz en una tarea.	Control conductual percibido	Autoeficacia técnica para emplear IA de forma eficaz; faceta de capacidad del control conductual percibido.	Ajzen (1991); Ha et al. (2025); Tbaishat et al. (2025)

5	Creo que es fácil detectar cuando alguien ha usado IA sin declararlo.	Oportunidad percibida	Percepción de la facilidad general de detección en el contexto, es decir, de la debilidad del sistema de control; operacionaliza el vértice de oportunidad del Triángulo del Fraude y se distingue del ítem 3 en que no se refiere a la conducta propia sino a una condición situacional externa.	Başer et al. (2026); Becker et al. (2006); Smith et al. (2023)
6	Las normas sobre el uso de IA en trabajos académicos o profesionales están bien definidas y son claras.	Ambigüedad normativa	Claridad percibida de las normas sobre uso de IA; ítem de redacción directa que se recodifica de forma inversa.	Kangwa et al. (2025); Smit et al. (2025)
7	Siento que las exigencias de mis tareas o responsabilidades actuales me superan con frecuencia.	Presión académica/laboral	Sobrecarga percibida de exigencias; vértice de presión del Triángulo del Fraude.	Becker et al. (2006); Smith et al. (2023)
8	¿Con qué frecuencia usas herramientas de IA en tus tareas actualmente? (escala de frecuencia)	Comportamiento pasado	Frecuencia habitual de uso de IA; indicador de hábito incorporado al modelo ampliado.	Harding et al. (2007); Stone et al. (2010)
Bloque 2: Sobre tus valores y motivaciones (Escala Likert 1-5)				
9	En general, me importa más obtener un buen resultado que el proceso de llegar a él.	Orientación motivacional al rendimiento	Prioridad del resultado sobre el proceso de aprendizaje; rasgo motivacional disposicional.	Becker et al. (2006); Smith et al. (2023)
10	Respetaría las normas éticas sobre uso de IA aunque nadie pudiera comprobarlo.	Obligación moral: internalización	Compromiso con las normas éticas con independencia de la vigilancia; faceta de internalización.	Bandura (1999); Bélanger et al. (2012); Chudzicka-Czupała et al. (2016)

11	Me sentiría culpable si usara IA en una tarea sin declararlo.	Obligación moral: culpa	Anticipación de culpa ante el uso no declarado; faceta afectiva.	Bandura (1999); Chudzicka-Czupała et al. (2016)
12	Usar IA sin declararlo iría en contra de mis principios.	Obligación moral: principios	Incompatibilidad del uso no declarado con los principios propios; faceta de principios.	Bandura (1999); Bélanger et al. (2012)
13	Si la IA mejora el resultado de un trabajo, usarla está justificado aunque no esté permitido.	Racionalización: utilidad percibida	Evaluación favorable del uso de IA porque mejora el resultado. Al captar la valoración positiva del comportamiento, se emplea además como aproximación a la actitud de la TPB ante la ausencia de un ítem de actitud independiente; se interpreta con cautela por su solapamiento conceptual con la racionalización.	Bandura (1999); Başer et al. (2026)
14	Si otros usan IA sin consecuencias, no veo por qué yo no debería.	Racionalización normativa (normalización social)	Justificación del uso por comparación con los pares; mecanismo de difusión de responsabilidad del desacoplamiento moral.	Bandura (1999); Başer et al. (2026)

Nota. Los ítems 1 a 14 se respondieron en una escala Likert de 1 a 5, salvo el ítem 8, formulado como escala de frecuencia. La restrictividad percibida de la política, empleada como comprobación de la manipulación, se midió con un ítem independiente en escala de 1 a 7. El ítem 6 (ambigüedad normativa percibida) se recogió como variable de contexto para caracterizar la muestra; dado que la manipulación experimental operacionalizó directamente la ambigüedad como tratamiento, no se incluyó en los modelos principales.

Tabla A2

Texto de las tres condiciones experimentales

Condición	Descripción
Condición permisiva	Para esta tarea, el uso de herramientas de inteligencia artificial generativa está permitido. Puedes apoyarte en ellas libremente durante todo el proceso.
Condición difusa	Para esta tarea, el uso de herramientas de inteligencia artificial generativa queda a tu criterio personal. Tú decides si apoyarte en ellas.
Condición restrictiva	Para esta tarea, el uso de herramientas de inteligencia artificial generativa no está permitido. El texto debe ser redactado íntegramente por ti.

Nota. En la aplicación, la cláusula central de cada política aparecía resaltada en mayúsculas y en color para asegurar su saliencia. Las tres políticas se presentaron en un recuadro destacado, con la cláusula central enfatizada, manteniendo constantes la longitud y la estructura para que la única variación entre condiciones fuera el grado de permisividad.

Subapartado A3

Arquitectura y funcionamiento de la aplicación experimental

La recogida de datos se realizó a través de una aplicación web desarrollada específicamente para este estudio con el apoyo de un asistente de programación basado en inteligencia artificial (Claude Code), cuyo código completo está disponible en el repositorio público (Repositorio GitHub). Su arquitectura se compone de un *frontend*, la parte visible que el participante recorre como una secuencia de pantallas, y un *backend* alojado en el servidor, invisible para él, que asigna la condición, gestiona la herramienta de IA y almacena la información generada. Esta separación entre *frontend* y *backend* resulta relevante porque garantiza que las decisiones de diseño del estudio, esto es, qué política se muestra, qué se registra y cómo, queden fuera del alcance y de la influencia del participante.

La manipulación se concretó en la asignación aleatoria de cada participante, en el momento de entrar, a una de las tres políticas sobre el uso de IA, cuyos textos coinciden literalmente con los recogidos en la Tabla A2. Esta asignación se resuelve en el *frontend* mediante una selección aleatoria entre las tres condiciones, en la que cada política conserva además la propiedad que mantiene visible el botón de ayuda de IA en los tres casos, como muestra el siguiente fragmento:

Fragmento de código 1. Asignación aleatoria de condición (JavaScript)

```
const policies = [
  { key: 'permisiva',   description: '… está permitido …',      showAIButton: true },
  { key: 'difusa',      description: '… queda a tu criterio …', showAIButton: true },
  { key: 'restrictiva', description: '… no está permitido …',   showAIButton: true }
];
const assignedPolicy = policies[Math.floor(Math.random() * policies.length)];
```

La política asignada se presenta primero en una pantalla de introducción, que obliga a permanecer en ella un mínimo de tres segundos antes de continuar para asegurar su lectura, y se mantiene después visible durante la propia tarea. Que el botón siga disponible también en la condición restrictiva es una decisión deliberada, puesto que solo conservándolo accesible en todos los casos es posible distinguir si la diferencia de uso entre condiciones procede de la norma y no de una restricción técnica impuesta por la propia aplicación.

La tarea central consistía en redactar un texto de entre 60 y 120 palabras, un intervalo controlado por un contador en tiempo real que solo permitía avanzar una vez alcanzado el mínimo, neutralizando la extensión como fuente de variación. Durante la redacción, el participante podía solicitar la ayuda de la IA y, en su caso, incorporar con un solo clic el fragmento sugerido, y en ese instante el *frontend* contabilizaba de forma automática los caracteres procedentes de la IA, que constituyen la medida objetiva del uso empleada en el análisis y que se envían al servidor junto al resto de eventos.

Esta medida se registra en el momento mismo de la acción y, por tanto, con total independencia de lo que el participante reconozca después en la pantalla de declaración, donde se le pedía indicar por separado qué porcentaje de su texto consideraba generado y qué porcentaje parafraseado por IA. Conviene precisar que ambos porcentajes se solicitaban como campos independientes, sin exigir que su suma fuera coherente, lo que explica el origen de los 10 casos cuya suma superaba el 100% y que se excluyeron en la depuración de los datos.

La información se recogió en dos niveles complementarios que se corresponden con las dos hojas del documento de datos (registro de eventos y registro consolidado por participante), cuya estructura se detalla a continuación. El primero es un registro conductual continuo que el *frontend* va anotando, sin interferir en la experiencia del participante, cada vez que ocurre una

acción relevante, entre ellas la entrada y la salida de cada pantalla, los clics, el tiempo de permanencia, los pegados de texto y, de manera destacada, las aperturas de la ayuda de IA y las inserciones de contenido generado. El segundo es un registro consolidado por participante que reúne en una única fila, al finalizar la sesión, los datos demográficos, las métricas de la tarea y todas las respuestas a las escalas, con una correspondencia exacta entre cada columna y cada variable del cuestionario (Tabla A1). La existencia de estos dos niveles es precisamente la que hace posible la validación cruzada entre la conducta que el sistema registra y lo que el participante declara.

Las sugerencias de la herramienta de ayuda se generaron mediante un modelo de lenguaje (gpt-4o-mini de OpenAI) al que la aplicación recurría siempre desde el backend y nunca desde el frontend, de modo que las credenciales de acceso quedaran protegidas, como ilustra la llamada que el servidor dirige al modelo:

Fragmento de código 2. Llamada al modelo de lenguaje desde el backend (Python)

```
requests.post("https://api.openai.com/v1/chat/completions",
  headers={"Authorization": f"Bearer {OPENAI_API_KEY}"},
  json={ "model": "gpt-4o-mini",
        "messages": [{"role": "system", "content": system_prompt},
                     {"role": "user",   "content": prompt}],
        "max_tokens": 80, "temperature": 0.7 }, timeout=10)
```

El modelo se configuró para devolver texto directamente utilizable y de forma acotada, ya que reescribía el fragmento que el participante hubiera seleccionado con un máximo de 40 palabras o, en ausencia de selección, generaba una frase de hasta 30 palabras coherente con lo ya redactado. Esta limitación deliberada evitaba que la herramienta resolviera la tarea por completo y mantenía el uso de IA como una decisión gradual del participante, observable en su intensidad.

Por último, la aplicación incorporaba mecanismos orientados a preservar la integridad de los datos, entre ellos el envío de la información en segundo plano, que conserva el registro aun cuando el participante cierra la ventana antes de terminar, lo que redujo al mínimo la pérdida de respuestas. El conjunto del código y las instrucciones para su ejecución figuran en el repositorio citado.

Anexo B. Resultados

Tabla B1

Estadísticos descriptivos de las variables continuas de la muestra

	Edad	**Nota Media**
N	180	180
Perdidos	0	0
Media	29,2	7,70
Mediana	23,0	7,85
Desviación estándar	13,2	1,04
Mínimo	18	2,00
Máximo	66	10,0

Nota. N = 180. Nota Media corresponde a la calificación académica autodeclarada en escala 0-10.

Tabla B2

Distribución de frecuencias de las variables sociodemográficas de la muestra

Variable		
Sexo	***n***	**% del Total**
Mujer	107	59,4
Hombre	73	40,6
Nivel de estudios		
Formación Profesional Superior	14	7,8
Bachillerato	4	2,2
Grado	101	56,1
Máster	57	31,7
Doctorado	4	2,2

Nota. N = 180

Tabla B3

Comprobación de la manipulación: restrictividad percibida por condición

Condición	***n***	***M***	***DE***	**Comparación (Games-Howell)**	**Diferencia**	***p***
Permisiva	63	1,94	1,46	Permisiva-Difusa	0,22	,701
Difusa	57	2,16	1,54	Permisiva-Restrictiva	1,30	< ,001
Restrictiva	60	3,23	2,13	Difusa-Restrictiva	1,08	,006

Nota. N = 180. Los participantes fueron asignados aleatoriamente a una de tres condiciones de política institucional de IA: permisiva, difusa y restrictiva. Escala 1–7. ANOVA de Welch: $F(2, 115) = 7{,}87$; $p < ,001$.

Tabla B4

Uso declarado de IA por condición

Condición	***n***	***M* (%)**	**DE**	**Comparación**	**Diferencia (pp)**	***p***
Permisiva	63	26,21	37,0	Permisiva-Difusa	12,4	,132
Difusa	57	13,86	32,7	Permisiva-Restrictiva	22,5	< ,001
Restrictiva	60	3,67	14,6	Difusa-Restrictiva	10,2	,086

Nota. La variable refleja el porcentaje total de contenido generado o parafraseado por IA en la tarea escrita. ANOVA de Welch: $F(2, 99{,}5) = 11{,}1$; $p < ,001$. Tamaño del efecto del ANOVA clásico: $\eta^2 = ,091$; $\omega^2 = ,080$. Las desviaciones estándar elevadas reflejan alta variabilidad intragrupo consistente con una distribución de uso de IA con concentración en cero. pp = puntos porcentuales

Tabla B5

Uso objetivo de IA (caracteres insertados) por condición

Condición	*n*	*M*	*DE*	Comparación	Diferencia	*p*
Permisiva	63	82,43	142,4	Permisiva-Difusa	15,1	,840
Difusa	57	67,33	150,6	Permisiva-Restrictiva	74,0	< ,001
Restrictiva	60	8,38	37,1	Difusa-Restrictiva	58,9	,015

Nota. ANOVA de Welch: $F(2, 87{,}6) = 11{,}3$; $p < ,001$.

Tabla B6

Uso declarado de IA externa por condición

Condición	Sin uso externo	Con uso externo	% uso externo
Permisiva	57	6	9,5
Difusa	55	2	3,5
Restrictiva	59	1	1,7
Total	171	9	5,0

Nota. $\chi^2(2) = 4{,}38$; $p = ,112$; V de Cramer = ,156.

Tabla B7

Correlaciones de Spearman entre variables conductuales

Variable	1	2	3	4
1. Ediciones manuales	—			
2. Uso objetivo (caracteres)	−,358***	—		
3. Uso declarado (%)	−,521***	,735***	—	
4. Palabras	−,097	,058	,045	—

Nota. *** $p < ,001$.

Tabla B8

Correlaciones de Spearman entre presión, oportunidad, racionalización, variables morales y uso de IA

Variable	Uso declarado	Uso objetivo	Presión	Racion. utilitaria	Racion. normativa	Oportunidad
Presión	,121	,116	—	,244***	,259***	−,029
Racion. utilitaria	−,022	,009	,244***	—	,491***	,034
Racion. normativa	,207**	,173*	,259***	,491***	—	,012
Internalización moral	,002	−,050	−,148*	−,389***	−,303***	,070
Culpa moral	−,098	−,052	−,126	−,469***	−,442***	,027
Principios morales	−,086	−,095	−,185*	−,411***	−,431***	,018

Nota. * = $p < ,05$; ** = $p < ,01$; *** = $p < ,001$. La columna "Uso objetivo" recoge la correlación con el número de caracteres insertados mediante el asistente. La oportunidad percibida no correlacionó con el uso declarado (ρ = ,058; p = ,438) ni con el uso objetivo (ρ = −,080; p = ,285), y se asoció únicamente con el uso de IA externa (ρ = ,177; p = ,017).

Tabla B9

Coeficientes estandarizados de los modelos de regresión sobre el uso declarado de IA

Predictor	**Modelo 1 (β)**	***p***	**Modelo 2 (β)**	***p***
Política: difusa–permisiva	−,434	,015	−,343	,049
Política: restrictiva–permisiva	−,766	< ,001	−,769	< ,001
Norma descriptiva	,208	,008	,124	,119
Norma injuntiva	−,021	,780	,018	,824
Control conductual percibido: evasión	,040	,624	,022	,795
Control conductual percibido: capacidad	−,026	,762	−,076	,407
Racionalización utilitaria	−,077	,329	−,186	,032

Predictor	Modelo 1 (β)	*p*	Modelo 2 (β)	*p*
Oportunidad percibida	—	—	,144	,058
Frecuencia de uso (hábito)	—	—	,191	,044
Racionalización normativa	—	—	,231	,009
Obligación moral	—	—	,034	,720
Orientación motivacional	—	—	−,034	,656

Nota. Variable dependiente: uso declarado de IA. Modelo 1: R^2 = ,133; R^2 ajustada = ,097; $F(7, 172) = 3{,}76$; $p < ,001$. Modelo 2: R^2 = ,218; R^2 ajustada = ,162; $F(12, 167) = 3{,}88$; $p < ,001$. El incremento respecto al Modelo 1 fue estadísticamente significativo: ΔR^2 = ,085; $F(5, 167) = 3{,}64$; p = ,004. Los valores en negrita indican predictores significativos ($p < ,05$).

Tabla B10

Mediación de la racionalización normativa sobre el uso declarado

Efecto / vía	b	*EE*	IC 95%	*p*
Indirecto: presión → racionalización → uso	1,324	0,737	[0,20; 3,12]	,072
Directo: presión → uso	0,662	1,721	[−2,99; 3,93]	,701
Total	1,986	1,562	[−1,14; 5,15]	,204
Vía a: presión → racionalización	0,273	0,085	[0,11; 0,44]	,001
Vía b: racionalización → uso	4,857	1,957	[1,00; 8,92]	,013

Nota. IC = intervalo de confianza. *EE* = error estándar. Los intervalos de confianza se estimaron mediante *bootstrap* de percentiles con 5.000 remuestreos. El IC 95% se reporta como criterio principal para interpretar el efecto indirecto.

Tabla B11

Mediación de la racionalización normativa sobre el uso objetivo de IA

Efecto / vía	b	*EE*	IC 95%	*p*
Indirecto: presión → racionalización → uso	3,400	2,240	[−0,13; 8,60]	,129
Directo: presión → uso	4,480	6,000	[−7,57; 15,91]	,455

Efecto / vía	**b**	***EE***	**IC 95%**	***p***
Total	7,880	5,830	[−3,64; 18,85]	,177
Vía a: presión → racionalización	0,273	0,085	[0,11; 0,44]	,001
Vía b: racionalización → uso	12,478	6,666	[−0,47; 26,07]	,061

Nota. IC = intervalo de confianza. EE = error estándar. Los intervalos de confianza se estimaron mediante bootstrap de percentiles con 5.000 remuestreos. El IC 95% se reporta como criterio principal para interpretar el efecto indirecto.

Tabla B12

Modelo lineal general de moderación entre política y obligación moral (pruebas ómnibus)

Término	***gl***	***F***	***p***	**η²**
Uso declarado				
Política	2, 174	9,00	< ,001	,093
Obligación moral	1, 174	0,96	,329	,005
Política × Obligación moral	2, 174	0,09	,912	,001
Uso objetivo				
Política	2, 174	6,49	,002	,069
Obligación moral	1, 174	1,33	,251	,007
Política × Obligación moral	2, 174	0,37	,693	,004

Nota. Obligación moral centrada en la media. η² = eta cuadrado.

Tabla B13

Modelo lineal general de moderación entre política y uso habitual de IA (pruebas ómnibus)

Término	***gl***	***F***	***p***	**η²**
Uso declarado				
Política	2, 174	6,75	,002	,069
Uso habitual (Usuario frecuente)	1, 174	6,21	,014	,032
Política × Uso habitual	2, 174	1,22	,298	,012

Nota. Variable dependiente: uso declarado de IA. El uso habitual se operacionalizó como variable dicotómica (usuario frecuente frente a no frecuente) derivada de la frecuencia previa de uso de IA. La prueba de Levene indicó heterogeneidad de varianzas ($F(5, 174) = 13,6$; $p <$,001), por lo que las pruebas *F* deben interpretarse con cautela. η² = eta cuadrado.