



¿Confiamos más en los humanos o en las máquinas?

La confianza en la automatización según género, edad y riesgo situacional.

Autor: Ana Navarro Jaro

Tutor: Dra. Nereida Bueno Guerra

Facultad de Ciencias Humanas y Sociales

Madrid

Mayo de 2024

# ÍNDICE

INTRODUCCIÓN.....	5
1.1 RELEVANCIA Y ESTADO ACTUAL DE LA CUESTIÓN .....	5
1.2 MARCO TEÓRICO.....	6
1.2.1 Confianza .....	6
1.2.2 Automatización .....	7
1.2.3 Confianza en la automatización .....	7
1.3 FINALIDAD Y OBJETIVOS .....	14
MÉTODOS.....	15
2.1 DISEÑO DEL ESTUDIO .....	15
2.2 MUESTRA.....	15
2.3 PROCEDIMIENTO .....	17
2.4 VARIABLES E INSTRUMENTOS DE MEDIDA.....	18
RESULTADOS .....	23
3.1 PREFERENCIA HUMANO VS. MÁQUINA.....	23
3.2 FIABILIDAD ATRIBUIDA AGENTE HUMANO VS. MÁQUINA .....	26
DISCUSIÓN.....	31
4.1 LIMITACIONES.....	36
CONCLUSIÓN .....	37
REFERENCIAS .....	38
ANEXOS.....	43

## **Resumen**

El impacto de las herramientas basadas en Inteligencia Artificial evoluciona a un ritmo vertiginoso. Su creciente protagonismo en diferentes áreas, que tradicionalmente habían sido propias de los humanos, hace necesario estudiar de qué forma las personas nos relacionamos con los sistemas automatizados. La confianza se presenta como un elemento clave en la interacción de los humanos con las máquinas, por lo que en el estudio se expone conceptualmente este constructo y los factores de los que depende. En este contexto, resulta cabal preguntarse: ¿las personas confiamos más en los humanos o en las máquinas? ¿esta elección es la misma para diferentes tipos de tarea? Si no es así ¿de qué depende dicha elección? En una muestra de 644 personas se exploraron factores disposicionales como la edad o el género; otros situacionales como el riesgo percibido de la tarea y otros aprendidos, como las expectativas o la fiabilidad atribuida a cada agente. Todo ello con el fin de arrojar luz acerca de qué aspectos influyen en el grado de confianza depositado en las máquinas y comprender cómo esta variable influye en el uso que se hace de las mismas.

***Palabras clave:** Confianza, automatización, sistemas automatizados, uPAS, interacción humano-máquina, riesgo, Inteligencia Artificial*

## **Abstract**

The impact of tools based on Artificial Intelligence is evolving at a dizzying pace. Their growing prominence in different areas, which traditionally had been the domain of humans, makes it necessary to study the factors that influence human-computer interaction. Trust is presented as a key element in the interaction of humans with machines, so the study conceptually exposes this construct and the factors on which it depends. In this context, it is pertinent to ask: do people trust humans or machines more? Is this choice the same for different types of tasks? If not, what does this choice depend on? In a sample of 644 people, we explored dispositional factors such as age or gender, situational factors such as perceived risk of the task, and learned factors such as expectations or the trustworthiness attributed to each agent. The aim was to shed light on the aspects that influence the degree of trust placed in machines and to understand how this variable influences the use people make of them.

**Key words:** *trust, automation, automated systems, uPAS, human-computer interaction, risk, Artificial Intelligence*

# INTRODUCCIÓN

## ***1.1 Relevancia y estado actual de la cuestión***

En las últimas décadas, tecnologías emergentes como la Inteligencia Artificial (IA) evolucionan a un ritmo vertiginoso. Estas no solo impregnan múltiples áreas de nuestra vida, sino que también juegan un importante papel ayudando a los profesionales de diversas áreas como la justicia, la sanidad, la agricultura o recursos humanos. En este contexto, la Comisión Europea (2020, p. 1) expone que “en la medida en que la tecnología deviene cada vez más central en todos los aspectos de la vida de las personas, estas deberían poder confiar en ella. La confiabilidad es un prerrequisito para su adopción”.

Es tal su auge e impacto que los buscadores y servicios de ayuda basados en inteligencia artificial están presentes también en el día a día influyendo en las decisiones que las personas toman, por ejemplo, al recomendar noticias, ofertas de trabajo e incluso potenciales amigos en redes sociales (Vicente & Matute, 2023). También proporcionan soporte a las instituciones públicas, de forma que estas también confían en los sistemas basados en IA a la hora de tomar decisiones sobre cuestiones que atañen a la esfera social y de las que antes se encargaban las personas, hasta el punto de que algunos autores refieren que estos son quienes dominan actualmente las decisiones y las reglas burocráticas de la sociedad (Lorenz *et al.*, 2021). Como ejemplos, los sistemas de concesiones de becas o préstamos bancarios (Martínez *et al.*, 2022), la concesión de permisos de un interno de prisión con RisCanvi (Andrés-Pueyo *et al.*, 2018), el diagnóstico de una enfermedad (*DxGPT*), el tratamiento psicoterapéutico (aplicaciones como *Appsy*) o la creación de imágenes (*Dall.e*), son procesos que ya se resuelven mediante algoritmos, con los prometedores beneficios y también los potenciales errores y sesgos inherentes asociados a su actuación; lo que disminuye su objetividad, seguridad e imparcialidad (Vicente & Matute, 2023). La presencia de estos sistemas en las últimas décadas ha llevado a la Fundación Española para la Ciencia y la Tecnología (FECYT, 2022) a publicar en un informe los resultados de la última investigación que han realizado acerca de la percepción social de la población española sobre la ciencia y la tecnología. Entre los datos que publican, la Inteligencia Artificial se sitúa como la segunda aplicación de las ciencias capaz de producir mayores beneficios, siendo un 47.4% de personas las que opinan que es una herramienta beneficiosa. En la misma línea, el 39.2% de los encuestados señala como beneficiosa la robotización del trabajo y un 63.8% considera

que los robots u otras tecnologías podrían realizar las tareas que las personas desarrollan en sus puestos de trabajo, al menos parcialmente. Es esta integración de los sistemas inteligentes en el organigrama social lo que abre la puerta a reflexionar acerca de cuál es la percepción y la actitud de los seres humanos respecto a las máquinas; preguntarnos hasta qué punto llega la confianza que depositamos en ella, y si esta está modulada por el tipo de tarea a la que debemos enfrentarnos en un momento dado.

## ***1.2 Marco teórico***

### **1.2.1 Confianza**

En el contexto de la interacción humano-máquina hay un consenso entre los expertos al reconocer el papel significativo que desempeña la confianza como un factor mediador en dicha, por lo que es necesario delimitar conceptualmente este constructo. Sin embargo, la confianza constituye un concepto de naturaleza multidimensional, abstracta y compleja, lo que convierte en un reto su definición y operativización al tratarse de un constructo que no se puede observar ni medir de forma directa (Poornikoo & Ivar, 2023). En la revisión que Muir (1994) hace de las múltiples conceptualizaciones existentes en torno a este término, entendido en un plano interpersonal, expone los aspectos comunes a todas ellas: primero, la confianza se define como la expectativa sobre un otro, connotación que nos sitúa en el futuro en la medida en que permite predecir una gratificación próxima, un determinado comportamiento o un suceso. Segundo, la confianza se deposita sobre un receptor específico, que puede ser algo o alguien concreto. Así, no se considera la confianza en términos absolutos como algo que se tiene en mayor o menor medida de forma global, sino que es algo relativo, al depender de otros factores. Tercero, la confianza se relaciona con distintas características del receptor sobre el que se deposita, como su fiabilidad, honestidad y motivaciones. Estos elementos subrayan la idea de que la confianza se necesita cuando algo se intercambia en una relación de cooperación en la que media la incertidumbre. Las definiciones que los distintos arrojan acerca de la confianza son muy diversas. En un inicio se hablaba de la confianza como algo disposicional, dado de forma natural (Rotter, 1967), aunque con el paso de los años se ha considerado un elemento más dependiente de la situación y del contenido. En este estudio seguiremos la definición de Lee & See (2004, p.54), abordando la confianza como la actitud que se sostiene al pensar que un agente contribuirá a conseguir los objetivos de un individuo en una situación en la que priman la incertidumbre y la vulnerabilidad.

Además, es importante destacar aquí que la confianza está sujeta a un proceso de formación dinámico, por lo que puede cambiar en la medida en que las personas estamos expuestas a nueva información constantemente.

### **1.2.2 Automatización**

Antes de abordar la confianza en la automatización como el eje central del presente trabajo es necesario definir el concepto de automatización. Con dicho término nos referiremos a toda herramienta tecnológica que, de manera activa, posee la capacidad de seleccionar datos, transformar información, tomar decisiones o controlar procesos (Lee & See, 2004, p.50). Concretamente, se abarcan los programas o sistemas informáticos que utilizan algoritmos entrenados previamente con datos reales que permiten generar respuestas o facilitar decisiones en contextos específicos. Como conceptos análogos, a lo largo del trabajo se emplean indistintamente los términos “máquina”, “robot”, “sistemas automatizados” o “Inteligencia Artificial” para referirnos a estas herramientas. Los accidentes relacionados con la automatización pueden deberse a múltiples motivos; bien sean organizacionales, ambientales, contextuales o debidas también al abuso, al mal uso o al desuso que un operador hace de ella (Parasuraman & Riley, 1977). La confianza juega un papel central en los casos de desuso o de uso indebido de la automatización, ya que guía la forma en la que el humano decide delegar o no en la máquina. Entre otros ejemplos, Hoff & Bashir (2014) ilustran en un nivel macro cómo el accidente del crucero Costa Concordia, en el que murieron 32 pasajeros, podría ser el resultado de la falta de confianza que el capitán depositó en el sistema de navegación automática en pro del control manual, cuando este decidió desviar la ruta programada y chocó con el arrecife que causaría el hundimiento. Un grado inapropiado de confianza en la automatización no puede causar accidentes únicamente en el nivel macro, sino que más adelante se exponen ejemplos de cómo ello influye en un nivel más micro.

### **1.2.3 Confianza en la automatización**

Partiendo de esta base, es necesario apuntar que actualmente no existe un modelo que aborde de forma integral y específica la confianza en la automatización o en las máquinas, sino que esta ha sido estudiada desde diferentes perspectivas, de forma que cada una sitúa el foco de interés en distintas características del constructo. Tal y como arguyen los expertos, la confianza que media en la interacción humano-máquina comparte propiedades importantes con la confianza humano-humano, pero difiere en algunos aspectos fundamentales. En el nivel más básico, ambos tipos de confianza son

dependientes de la situación en la que se da, y son relevantes cuando se producen en el seno de una relación de cooperación donde caracterizada por un sentimiento de incertidumbre. Tal y como expone Muir (1994), como la confianza interpersonal, la confianza humano-máquina es un fenómeno continuo que puede depositarse en un agente en su totalidad o atribuirse solo a ciertas partes, capacidades o funciones del mismo (Chiou & Lee, 2023). Según Mayer *et al.*, (1995) la confianza interpersonal se basa en la habilidad, integridad o benevolencia del destinatario de esa confianza (Mayer *et al.*, 1995), mientras que en la interacción con las máquinas, la confianza se basa en el rendimiento, proceso o propósito del sistema (Lee & Moray, 1992). Como un constructo psicológico, la confianza parece ser relativamente estable en el largo plazo, pero también puede cambiar, evolucionar y debilitarse a lo largo del tiempo. La literatura puntualiza la asimetría que existe también en torno a la progresión de la confianza. En el contexto interpersonal, esta se construye lenta y gradualmente ya que los seres humanos tienden a ser cautos en lo que respecta a abrirse a desconocidos, basándose inicialmente en la intuición de poder predecir las acciones de la otra persona. Con el paso del tiempo, la confirmación de dicha integridad deviene la base fundamental de la confianza (See & Lee, 2004). Por otro lado, la confianza entre los humanos y la automatización sigue el orden inverso. Lyons & Guznov (2018) sugieren la existencia de un “sesgo de positividad” en la interacción con las máquinas, con el que se parte de una confianza total en torno a su correcto funcionamiento. Las personas tienden a suponer que las máquinas son perfectas basándose en la fe, sin embargo, la confianza en ellas se diluye rápidamente tras la comisión de un error del sistema, de forma que la fiabilidad y la previsibilidad pasan a ser la base de dicha confianza. A este respecto, algunas líneas de investigación subrayan el riesgo de sobreconfiar en las capacidades de estos sistemas automatizados, lo que podría comprometer la seguridad de las personas (Lee & See, 2004).

Existen tres niveles de confianza (Hoff & Bashir, 2014; Marsh & Dibben, 2003): disposicional, situacional y aprendida. En nuestro estudio, la primera representa la tendencia general de una persona a confiar en la automatización, independientemente del contexto o del sistema específico que sea, que vendría determinado tanto por la genética como por la influencia ambiental. Factores como la cultura, edad, género y personalidad influyen en el grado en que se dispone de este tipo de confianza. Por el contrario, la confianza situacional no se considera estable a lo largo del tiempo, sino que depende del contexto específico de cada situación y se encuentra moderada por el ambiente externo e



interno y por las características dependientes del contexto. Otros autores coinciden al exponer que el nivel de confianza en la automatización es dependiente del contexto de la tarea (Appelganc *et al.*, 2022; Castelo *et al.*, 2019), y afirman que se ve influido por el nivel de riesgo que esta conlleva, de forma que aquellas tareas que conciernen a la salud o la seguridad pública se conciben como más arriesgadas que aquellas que versan sobre asuntos económicos o financieros (Weber *et al.*, 2002). Por último, la confianza aprendida se basa en las experiencias relevantes pasadas respecto de un sistema específico, lo que condiciona la evaluación que se sostiene respecto de la interacción actual con él. Estas experiencias pueden alterar el proceso de formación de confianza en la automatización, determinando las decisiones que se toman acerca de si usarlos.

Por otro lado, autores como Castelo *et al.*, (2019) señalan que las personas presentan cierta aversión a delegar en la actuación de los algoritmos cuando se trata de tareas que son comúnmente realizadas y atribuibles a los humanos, aún a pesar de que estos algoritmos puedan reportar un mejor desempeño. En su estudio, donde contrastaban la confianza para una serie de tareas en función del tipo de agente (siendo este un humano cualificado o un algoritmo) encontraron diferencias significativas a favor del humano en la mayoría de las tareas propuestas (como pilotar un avión o escribir un artículo). Sin embargo, para tareas a las que los participantes atribuían un mayor grado de objetividad, como dar indicaciones o emitir predicciones meteorológicas, mostraban un mayor grado de confianza hacia los algoritmos. Así, concluyen que las personas están menos dispuestas a confiar y a emplear algoritmos para aquellas tareas que conciben como subjetivas, al pensar que estos no disponen de las habilidades necesarias para ejecutarlas. En la misma línea, Roesler *et al.*, (2022) ponen de relieve que los participantes de su estudio, para la tarea de evaluar qué porcentaje de rayos X simulados superaban en brillo a un umbral dado (y, por tanto, determinar el porcentaje de tejido potencialmente maligno en la prueba de un paciente determinado) confiaban en mayor medida en el apoyo que recibían por parte de un humano que en la ayuda proporcionada por el sistema automatizado. Este mismo experimento lo condujeron también Rieger *et al.*, (2022) quienes revelaron una mayor confianza cuando el apoyo era brindado por un humano experto que cuando derivaba de una IA. Sin embargo, cuando el humano era presentado como novel en la materia, los participantes mostraban mayores niveles de confianza en la IA. En consecuencia, concluyeron que el nivel de pericia del agente parece ser una condición clave que modula el nivel de confianza.

En la medida en que estos servicios basados en IA ostentan un papel en lo que respecta la toma de decisiones y la cooperación entre la inteligencia natural y artificial se presenta como una mejora del rendimiento de los profesionales; comprender el cuándo y el cómo los usuarios humanos eligen depender de la automatización puede ayudar a prevenir decisiones incorrectas y consecuencias potencialmente desastrosas, concretamente por el hecho de que estos algoritmos pueden presentar errores en los juicios y la capacidad de decidir en el transcurso del procesamiento e interpretación de los datos, dando lugar a los denominados sesgos (Merrit *et al.*, 2015; Vicente & Matute, 2023). Como se mencionaba previamente, las interacciones defectuosas de las personas y la automatización derivan consecuentemente en un mal uso (por depender en exceso) o, en contrapartida, en el desuso (al rechazar la ayuda de la automatización incluso cuando podría mejorar el desempeño de la persona). En el marco de la salud mental, mientras que la OMS (2021, p. 9) recomienda “el fomento de la autoasistencia mediante el uso de tecnologías sanitarias electrónicas y móviles”, la realidad nos ilustra con ejemplos que se han resuelto trágicamente debido a esta sobreconfianza y falta de cuestionamiento de las herramientas digitales, tal y como ejemplifica uno de los usuarios de un chat de inteligencia artificial, al decidir terminar con su propia vida tras seguir los consejos de *Eliza* (Xiang, 2023). Para estudiar los mecanismos que subyacen a estos distintos niveles de confianza por parte de los usuarios en la automatización, Dzindolet *et al.*, (2002), han desarrollado el *Perfect Automation Schema* (PAS), un instrumento que mide las creencias en relación con la actuación de un sistema automatizado. Por un lado, mide las *expectativas* en el rendimiento de las máquinas, donde una alta puntuación revelaría la confianza en un desempeño casi perfecto, en el que apenas se concibe margen de error por parte del sistema. De esta expectativa deriva la segunda subescala, que mediría *creencias del tipo “todo o nada”*, es decir, pensar que el sistema automatizado puede funcionar perfectamente, o no funcionar en absoluto.

Al igual que los esquemas con los que funcionamos las personas se ven influidos por nuestras experiencias y actúan como un mecanismo cognitivo cuyo fin es facilitar la interacción con el entorno (Fiske, 1998; Rumelhart, 1980), del mismo modo lo hace el *esquema de automatización perfecta*, objeto del presente estudio. Así, considerando los sesgos que operan en relación con los sistemas automatizados, Dzindolet y colaboradores (2001) identificaron dos ideas claras en su estudio. Para una tarea en la que los participantes eran asistidos por un humano o por un sistema automatizado, estos

declararon fiarse en mayor medida de un sistema automatizado en comparación con un ayudante humano antes de realizar la tarea experimental. También es notable destacar que, después de cometer un error, la percepción de utilidad disminuía tanto para los sistemas humanos como para los automatizados. Sin embargo, esta disminución fue más pronunciada en el caso de los sistemas automatizados en comparación con los humanos. Los autores concluyeron así que, dado que las personas esperamos un perfecto desempeño por parte de los sistemas automatizados, la comisión de un solo fallo genera un mayor rechazo a la hora de volver a confiar en ellos en comparación con lo que ocurre con los agentes humanos, cuyos fallos se toleran más. Con todo ello, comprender el PAS y las creencias asociadas a este esquema cognitivo es importante para dilucidar de qué forma se operativiza la confianza en la automatización y conocer uno de los de los constructos diferenciales entre la confianza en la automatización y la confianza interpersonal (Dzindolet *et al.*, 2003).

Teniendo en cuenta que hablar de la confianza en la automatización implica hablar de un constructo multifactorial, cabe anotar que las variables individuales juegan un papel importante, pudiendo influir en el grado en que se desarrolla esta confianza. Es evidente que el mundo tecnologizado en el que vivimos ha convertido los dispositivos digitales en una herramienta de uso diario, pero esta es una realidad aún más acuciada en el caso de los jóvenes, quienes emplean gran parte de su tiempo diario en ellos, de forma que las tecnologías de la información y las comunicaciones constituyen su medio de comunicación, de expresión e incluso de aprendizaje (Saiz, 2021). Así, los jóvenes recurren a Internet como primera opción en la que buscar información, además de utilizarlo como medio de comunicación y socialización (Mayer, 2011). Este factor diferencial respecto de otras franjas etarias justifica el hecho de que, en líneas generales, los jóvenes muestren más confianza en Internet que los mayores. Si bien el 62% de los usuarios afirma confiar “bastante” en Internet, su uso diario va descendiendo conforme se avanza en edad, encontrando el mayor porcentaje de uso en la franja etaria que abarca de los 16 a los 24 años (98.1%) y el menor en la población de 65 años en adelante (59.9%) (INE, 2022). Además de existir diferencias entre chicos y chicas en cuanto a la frecuencia de uso de Internet (siendo superior en ellos) el informe PISA de 2018 recoge las diferencias en relación con las actividades a las que se destina ese tiempo. Según explica, las chicas emplean este tiempo en línea en actividades sociales (como chatear o comunicarse a través de redes sociales), mientras que los chicos lo utilizan

preferentemente en actividades de ocio (como jugar o leer noticias). Dicha tendencia no solo se da en España, sino que se extrapola a la media de la OCDE (2015) y se reproduce también en la Unión Europea.

Otro de los factores diferenciales reside en el sexo, aspecto que se refleja en las trayectorias educativas y las elecciones vocacionales por las que optan los estudiantes, donde se acucia la brecha de género existente en torno a los ámbitos científico-tecnológicos, en los que, de forma persistente en las últimas décadas, las mujeres representan un porcentaje significativamente menor respecto de los hombres (Grañeras-Pastrana *et al.*, 2022). En este contexto surge un modelo educativo que promueve la preparación y el conocimiento del alumnado en línea con los avances tecnológicos. Bajo el acrónimo en inglés STEAM, se recogen las cinco disciplinas académicas clave en la sociedad actual: Science (ciencias), Technology (tecnología), Engineering (ingeniería), Arts (artes) y Mathematics (matemáticas). Esta acepción, acuñada por la Rhode Island School of Design de Estados Unidos en 2010, supone una actualización del término STEM, de forma que ahora da cabida a las artes, y con ello a la innovación y a la creatividad. Desde el Gobierno, a través del plan España Digital 2025 se recomienda encarecidamente la promoción más paritaria posible de las vocaciones circunscritas a cada una de estas áreas. A este respecto, el documental “Picture a Scientist” ilustra los desafíos y los obstáculos que las mujeres experimentan en los campos de la ciencia y la investigación como consecuencia de los sesgos de género, la discriminación y la falta de representación de estas en dichas áreas, ya que en él se vio que niños y niñas dibujaban predominantemente a hombres cuando recibían la directriz que lleva como nombre el cortometraje. Esto variaba en función de la edad, lo que respondería a los mensajes que tradicionalmente se han introyectado desde los medios de comunicación acerca de quiénes se dedican a la ciencia. Cabe destacar aquí el estudio de Chambers (1963), cuyo objetivo era determinar la edad a la que los niños comenzaban a desarrollar imágenes diferentes acerca del perfil de científico. Adaptando el *Draw a Scientist Test* a una muestra de cerca de 5000 escolares, concluyeron que los elementos estereotípicos incluidos en los dibujos (como bata blanca, barba o gafas), aumentaban según aumentaba el curso, pero no estaban apenas presentes en los preescolares. Es llamativo a su vez que únicamente fueron dibujadas 28 mujeres, todas ellas ilustradas por niñas.

En lo que respecta a la actualidad, Grañeras-Pastrana *et al.*, (2022) compilan algunos datos significativos sobre las trayectorias educativas predominantes que se hacen

eco de estas diferencias de género. Como ejemplo, se ha revelado que existen diferencias significativas entre el porcentaje de chicos que proyectan su futuro laboral en torno a la ciencia y la ingeniería, siendo de 8.7 puntos por encima de las chicas que espera desarrollarse profesionalmente en estas áreas. Cabe aquí la excepción de las ciencias de la salud, donde las chicas presentan un porcentaje superior al de los chicos en 12.2 puntos. Las TIC revelan también dicha brecha de género, siendo el porcentaje de chicos de 9.5 frente al 1.2 de las chicas que esperan trabajar en este campo. Esto se traduce en que las mujeres representan el 70% de las plazas en los estudios relacionados con las Ciencias de la Salud, y solo el 25% en las carreras como Ingeniería o Arquitectura (Cáceres *et al.*, 2017). Estos datos adquieren relevancia en el objeto del presente estudio debido a que, tal y como indica la Comisión Europea (2018), estas diferencias de género no se reflejan solo en las trayectorias académicas y laborales, sino también en las actitudes de los ciudadanos hacia la tecnología y la innovación. En su informe, señalan que los hombres sostienen una visión más positiva que las mujeres acerca del impacto digital; idea que ilustran con el dato de que un 70% de hombres, frente a un 63% de mujeres, piensan que las tecnologías digitales más recientes ejercen un impacto positivo en su calidad de vida. Además, afirman que en la medida en que las mujeres tienden a informarse menos que los hombres acerca de estos avances tecnológicos, ellas presentan un mayor grado de desconfianza hacia las herramientas digitales.

Las diferencias individuales que median en el desarrollo de un mayor o menor grado de confianza en la automatización no se circunscriben únicamente a áreas objetivables como son las áreas STEAM. Como se adelantaba al inicio, existe ya un amplio abanico de procesos automatizados cuyas repercusiones se enmarcarían en el espectro de lo subjetivo. En contextos como son los procesos de selección de personal, se ha demostrado que las personas prefieren confiar en la ayuda humana más que en un soporte automatizado (Appelganc *et al.*, 2022; Langer *et al.*, 2021). En contrapartida, como ejemplo del mayor grado de subjetividad a este respecto podemos considerar la esfera de la salud mental, donde las TIC también constituyen una realidad que ha ido ganando entidad desde hace algunos años (López-Santín & Serón, 2018). Ya en 2017 se reportó la existencia de más de diez mil *apps* sobre salud mental disponibles para los usuarios en Apple store o Google Play (Torous & Roberts (2017), y según un estudio, el 72% de los pacientes psiquiátricos ambulatorios manifestaron un alto interés por utilizar sus teléfonos inteligentes para monitorizar su salud mental (Torous *et al.*, 2014). Como

cabría esperar según la información anterior, concluyeron que la generación más joven, representada por los menores de treinta años, presentaba los mayores índices de disposición a recibir un tratamiento de estas características. En esta línea, un estudio australiano acerca de la percepción de las psicoterapias online por parte de los adolescentes reveló que un 72% de ellos accedería a un tratamiento online para tratar un problema de salud mental y el 31.9% preferiría una terapia online frente al formato clínico tradicional (Sweeney *et al.*, 2019). Por su parte, Glasheen y colaboradores (2016) revelaron en su estudio que los adolescentes preferían informarse en Internet sobre cuestiones como la sexualidad, pero optaban por el apoyo tradicional cara a cara para otros aspectos como la resolución de conflictos entre iguales o solicitud de asesoramiento. Tal y como deducen Sweeney *et al.*, (2019), cabe pensar, por tanto, que los adolescentes consideran más apropiados los servicios online para tratar aquellas cuestiones en torno a las que existe cierto estigma social, al encontrar una mayor comodidad.

### ***1.3 Finalidad y objetivos***

Por tanto, teniendo en cuenta la naturaleza multifactorial que caracteriza la confianza en la automatización, y los distintos grados en los que esta puede desarrollarse en función de las características individuales de las personas, en este estudio pretende analizarse cuáles son aquellas áreas en las que las personas prefieren optar por recurrir a ayuda humana y en cuáles apuestan por la ayuda automatizada, y en qué medida lo hacen. Así, es esperable encontrar diferencias en función de la frecuencia con la que se utilizan herramientas o buscadores basados en inteligencia artificial, del sexo, de la edad, o de la rama de conocimiento, que expliquen la tendencia a confiar en mayor o menor medida en estas herramientas digitales con las que nos encontramos en continua interacción. Ampliar la perspectiva en lo que respecta la confianza en la automatización es clave para comprender de qué formas nos relacionamos los seres humanos con una realidad tecnológica que impregna nuestra vida, tratando de arrojar luz acerca de la optimización de este tipo de recursos y encontrando el equilibrio en la balanza entre el desuso y el mal uso que se hace de las mismas.

# MÉTODOS

## ***2.1 Diseño del estudio***

Para dar respuesta a los objetivos propuestos se ha optado por un diseño observacional transversal con el fin de recopilar datos de una muestra concreta en un momento específico. Los datos se recogieron a través de un formulario online en el que se medían las distintas variables de interés en relación con el objeto de estudio. Con el fin de valorar los factores que influyen sobre nuestra variable dependiente, la confianza en la automatización, se establecieron tres grupos principales en función de la franja etaria: niños, adolescentes y adultos. A través de un formulario adaptado a cada uno de los subgrupos, se propusieron a los participantes diferentes tareas a resolver, para las que debían decidir si confiarían más en un experto humano o en un sistema automatizado para su realización, así como graduar el riesgo que atribuían o la familiaridad que les merecía cada una de las tareas. Además de esto, se administró un cuestionario estandarizado para medir la variable de la confianza en la automatización, detallado más adelante.

## ***2.2 Muestra***

Con el objetivo de evaluar si existen diferencias respecto de la confianza en la automatización en función de la edad, se seleccionan tres perfiles muestrales: niños (6º Educación Primaria), adolescentes (3º ESO y 2º BTO) y adultos.

En el caso de los adultos, la muestra se seleccionó de forma aleatoria, a través de la difusión por redes sociales de los instrumentos, ambos recogidos en un Microsoft Forms, encabezado por una cláusula de consentimiento informado que los participantes aceptan previamente a comenzar los cuestionarios. En el caso de los menores, se contactó por correo electrónico con distintos centros educativos oficiales de educación primaria y secundaria, siendo elegibles para el estudio todos aquellos alumnos cursando 6º de Primaria, 3º de la ESO y 2º de Bachillerato, y cuyos representantes legales firmaran el consentimiento firmado que les autoriza a responder a las preguntas o la declaración de información (en el caso de los menores entre 14-17 años), ambos documentos facilitados a través del centro. En relación con los criterios de exclusión, con el fin de evitar la posible contaminación de los datos, se prescinde de la participación de aquellas personas con dificultades de comprensión lingüística. Así, en los colegios, este criterio englobaría aquellos estudiantes con discapacidad intelectual, Trastorno del Espectro Autista o similares, quienes responderían el cuestionario, pero cuyas respuestas serían eliminadas

en el caso de que no fueran confiables. Por el mismo motivo, en el caso de los adultos se excluyen además las respuestas relativas a personas con demencias.

El estudio partió de una muestra inicial de 208 alumnos de 6° de Primaria, 261 alumnos de 3° ESO y 2° Bachillerato y 238 adultos. El primer tramo etario se compuso definitivamente de 169 alumnos tras excluir a aquellos participantes que no habían sido autorizados por parte de sus representantes legales; el segundo resultó en 250, tras eliminar los participantes que no pertenecían a los cursos académicos correspondientes y el tercero en 225, al excluir a 8 sujetos que no superaban la educación primaria como formación educativa alcanzada, al no constituir un grupo estadísticamente representativo de la muestra.

Teniendo en cuenta que la finalidad del estudio es estudiar cuáles son las variables que se relacionan con el grado de confianza en la automatización y en qué medida lo hacen, primeramente se recogió información sociodemográfica referente al género, la edad, el nivel de formación alcanzado y el área de conocimiento preferente de cada participante (acceda [aquí](#) para consultar el listado oficial de titulaciones correspondientes a cada una de las ramas de conocimiento). Asimismo, se incluyó una pregunta para explorar la frecuencia con la que se utilizan herramientas de inteligencia artificial, con el fin de analizar si un mayor uso se corresponde con una mayor confianza en la misma. En la Tabla 1 se recopilan las características sociodemográficas de la muestra.

Tabla 1. Descripción de las características sociodemográficas de la muestra

	<b>Frecuencia</b>	<b>Porcentaje</b>
<b>N Total</b>	644	100%
<b>Género</b>		
Hombres	280	43.5%
Mujeres	364	56.5%
<b>Edad</b>		
Entre 9-13	167	25.9%
Entre 14-18	256	39.8%
Entre 19-24	87	13.5%
Entre 25-30	45	7%
Entre 31-40	25	3.9%
Entre 41-50	19	3%



51 o más	45	7%
<b>Curso*</b>		
6° Primaria	169	40.3%
3° ESO	161	38.4%
2° BCTO	89	21.2%
<b>Rama de conocimiento</b>		
Ciencias	85	13.2%
Ciencias de la Salud	158	24.5%
Ciencias Sociales y Jurídicas	123	19.1%
Ingeniería y Arquitectura	101	15.7%
Arte	34	5.3%
Otro (profesión no cualificada)	143	22.2%
<b>Nivel de estudios completado**</b>		
Bachillerato	34	15.1%
Formación profesional	19	8.4%
Carrera universitaria	112	49.8%
Estudios de postgrado	60	26.7%
<b>Frecuencia de uso de herramientas de IA</b>		
Nunca	153	23.8%
Raramente	154	23.9 %
Ocasionalmente	146	22.7%
Frecuentemente	98	15.2%
Muy frecuentemente	50	7.8%
Las desconozco	43	6.7%

*Nota.* \* = Esta pregunta no se realizó al grupo de adultos. \*\* = Esta pregunta sólo se realizó al grupo de adultos. (A excepción de \* y \*\*, todas las preguntas se aplicaron en todos los grupos).

### **2.3 Procedimiento**

Previo a la recolección de datos, se llevó a cabo una revisión bibliográfica de la literatura relevante en la materia, con el fin de seleccionar los instrumentos de medida pertinentes y establecer las variables relacionadas con el objeto de estudio. Una vez establecido el marco teórico y definido el diseño del estudio, se solicitó su revisión por parte del Comité de Ética, garantizando los requisitos correspondientes que verifiquen el bienestar y la salvaguarda de los participantes, particularmente en el caso de los menores.

Tras recibir la aprobación por parte del Comité de Ética (dictamen 25/23-24) para la aplicación de los cuestionarios, se procedió a contactar con la muestra constitutiva del estudio.

En el caso de los adultos, la muestra se seleccionó por un procedimiento de muestreo no probabilístico de “bola de nieve”, difundiendo el formulario a través de los distintos canales digitales, desde aplicaciones como Whatsapp o Instagram, con el fin de maximizar el alcance y el tamaño de la muestra. En el caso de los menores, se contactó con varios centros educativos del territorio estatal, de los cuales participaron seis en total (perteneciendo cuatro de ellos a Madrid, uno a León y otro a Lugo). El intervalo de tiempo dedicado a la recogida de datos se encuentra comprendido entre los meses de febrero y abril de 2024.

Todos los participantes del estudio prestaron consentimiento informado (bien por sí mismos en el caso de los mayores de edad, bien a través de sus representantes legales en el caso de los menores). En dicho consentimiento se proporcionaba información del contenido del estudio y sus objetivos, así como se informaba del carácter anónimo, confidencial y de la libertad para desistir de la cumplimentación del formulario en cualquier momento sin que ninguna consecuencia derivara de ello. En el caso de los menores, se dedicaban unos minutos a explicar en el aula los diferentes tipos de ítems, con ejemplos distintos a los que se incluían en el cuestionario que permitían clarificar posibles dudas, sin sesgar la posterior elección de las respuestas.

El cuestionario se dividió según el siguiente orden de bloques de ítems: preguntas sociodemográficas; escala uPAS (se administró en el grupo de adolescentes y adultos, pero se descartó en el de niños debido a que no está validada para este grupo poblacional y sus ítems podrían ser demasiado complejos para su comprensión); ítems de elaboración propia (preferencia por agente humano vs. máquina, fiabilidad atribuida a cada agente; riesgo percibido de la tarea y familiaridad respecto de cada tarea).

#### ***2.4 Variables e instrumentos de medida***

##### *Datos sociodemográficos y epidemiológicos*

Como se recoge previamente en la Tabla 1, se midieron aquí las variables relativas al género, la edad, curso académico (en el caso de los niños y adolescentes), la rama de conocimientos a la que se pertenece o, en el caso de los menores, en cuál les gustaría

desarrollarse profesionalmente, el nivel de estudios superior alcanzado (en el caso de los adultos), así como la frecuencia con la que utilizan herramientas basadas en IA.

*Confianza en la automatización: Escala uPAS (Gibson et al., 2023)*

En segundo lugar, se administró el cuestionario uPAS (Gibson et al., 2023), cuyos ítems tradujeron al castellano por las propias investigadoras del estudio, siendo cotejados convenientemente a través de herramientas digitales como Deepl o ChatGPT. Este instrumento se desarrolló a partir de la revisión de la escala original de Merrit et al. (2015), debido a la cuestionabilidad de sus propiedades psicométricas, en particular las debilidades reportadas en la consistencia interna de sus subescalas (Merrit et al., 2019). La medición de la confianza en la automatización se ha venido llevando a cabo con el cuestionario *Perfect Automation Schema* (PAS) (Merrit et al., 2015). Para estos autores, la confianza se estructuraba en torno a dos factores: las *expectativas* en el rendimiento de las máquinas, donde una alta puntuación revelaría la confianza en un desempeño casi perfecto en el que apenas se concibe margen de error por parte del sistema; y el *pensamiento dicotómico*, traducido en un pensamiento del tipo “todo o nada” por el que se concibe que el sistema puede funcionar a la perfección, o no funcionar en absoluto (Dzindolet et al., 2001, 2002). No obstante, años después los propios autores pusieron en cuestionamiento las propiedades psicométricas de la escala, en particular las debilidades reportadas en la consistencia interna de estos dos factores (Merrit et al., 2019). Esto dio lugar al desarrollo de una escala actualizada denominada uPAS (de ahí la sigla “u”, *updated*, “actualizado” en inglés), validada por Gibson y colaboradores (2023) y cuyos ítems se basan en los mismos constructos conceptuales que la escala original de Merrit et al. (2015). En el año 2019, Merrit y colaboradores validaron una nueva escala con la que medir el *Potencial de Complacencia Inducida por la Automatización*, constructo que también se relaciona con la confianza en la automatización, objeto del presente estudio, al evaluar cómo varían los niveles de vigilancia/monitorización de la actuación de los sistemas automatizados en función de variables como la carga de trabajo. Si bien esta escala podría constituir el instrumento de medida de este estudio al estar elaborada por los mismos autores que la escala inicial que evaluaba el PAS, y al haber sido ampliamente utilizado por la comunidad científica, la razón por la que se ha optado por la escala uPAS (Gibson et al., 2023) es porque se trata de una escala basada en el mismo constructo que la que elaboraron originalmente Merrit y colaboradores en el año 2015, por lo que mide

con mayor exactitud lo que aquí definimos como confianza en la automatización, respaldada en los términos conceptuales sobre los que versa el trabajo. Además, el hecho de que esta nueva escala haya sido escasamente empleada por la comunidad científica podría deberse a que el estudio data de este mismo año, por lo que dicho criterio no dispone de la suficiente entidad como para descartar el instrumento empleado.

Nuestra adaptación en español del uPAS obtuvo propiedades psicométricas apropiadas en esta muestra. El análisis paralelo reveló dos factores y dos componentes que replican la estructura factorial original. La Tabla 2 presenta la estructura factorial (cargas factoriales estandarizadas) donde ambos factores presentaron una correlación de 0.06, explicando el 43% de la variabilidad de los ítems. El ajuste del modelo fue adecuado ( $\chi^2(13) = 53.30$ ,  $p < 0.001$ , RMSEA = 0.0806, IC del 90% 0.0589 – 0.104, TLI = 0.904). La consistencia interna ( $\alpha$  de Cronbach) de las subescalas fue de 0.82 para “pensamiento dicotómico” y 0.65 para “expectativas”.

**Tabla 2.** Cargas factoriales estandarizadas del análisis factorial exploratorio (*minres solution* y *oblimin rotation*) para la adaptación en español del uPAS.

<b>Item</b>	<b>Factor 1: “Pensamiento dicotómico”</b>	<b>Factor 2: “Expectativas”</b>	<b>h<sup>2</sup></b>
Puedo confiar en que los S.A. tomen decisiones precisas	0.05	<b>0.59</b>	0.35
Los S.A. realizan el trabajo mejor que los humanos	-0.02	<b>0.59</b>	0.35
Los S.A. son más fiables que los humanos	-0.02	<b>0.59</b>	0.35
Acepto las decisiones tomadas por los S.A.	-0.03	<b>0.49</b>	0.24
Los S.A. que cometen errores deberían abandonarse	<b>0.65</b>	0.01	0.43
Me niego a utilizar S.A. imprecisos	<b>0.67</b>	0.07	0.46
Si un S.A. comete un error no lo utilizaré	<b>0.83</b>	-0.02	0.69
Los S.A. imperfectos deberían ser descartados	<b>0.76</b>	-0.03	0.57

*Nota.* h<sup>2</sup> = Comunalidad de los ítems (inverso a la unicidad de los ítems). Se encontró una correlación de 0.06 entre ambos factores.

### *Preguntas ad hoc: ítems de elaboración propia*

Por otro lado, se emplearon una serie de ítems de elaboración propia, por los que se preguntó a los participantes, en formato dicotómico, si preferirían delegar en un experto humano o en una máquina para la ejecución de una serie de procesos. Las tareas a realizar, que constituyen el objeto de cada ítem, se dividieron en dos secciones: la primera de ellas contiene ítems en los que se incluyen procesos automatizados basados en aspectos objetivables (como búsqueda de información o resolución de problemas, recogidos bajo el paraguas de las áreas curriculares STEAM: ciencias, tecnología, ingeniería, arte y matemáticas, respectivamente); el segundo de los bloques plantea la realización hipotética de otra serie de procesos automatizados, de los que derivarían consecuencias subjetivas para la persona, insertos en las áreas fundamentales de la vida de la persona: académico-laboral, económica, social, afectivo relacional y de la salud (como un diagnóstico médico, la resolución de la petición de un préstamo, la selección de su candidatura a un puesto de trabajo o la petición de ayuda para recibir un tratamiento psicológico). Esta división en dos bloques responde a la justificación teórica de la que los expertos en la materia se sirven para defender que la confianza en la automatización depende, entre otros aspectos, del contexto y de la naturaleza del tipo de tarea a realizar (Appelganc *et al.*, 2022; Castelo *et al.*, 2019), así como del nivel de riesgo que se asocia con ella (Weber *et al.*, 2002).

El formato de respuesta para cada uno de estos ítems es de tipo dicotómico, lo que refleja si el participante prefiere recurrir en una máquina o en un experto humano para la realización de cada una de las tareas descritas. Posteriormente, con el fin de afinar el grado de confianza para cada uno de los tipos de procesos automatizados planteados, se incluye una escala tipo Likert que medirá la percepción de fiabilidad/adecuación que el participante atribuye al desempeño de cada uno de los agentes en el que ha confiado para la tarea concreta. Asumiendo que las respuestas acerca de estos procesos automáticos subjetivos pueden verse influidas en función de la proximidad o familiaridad que el participante tenga con respecto de la tarea a resolver, se incluye, para cada uno de dichos ítems, una pregunta con la que se pretende evaluar cómo de cercana le resulta la situación. Ver Tabla 3 para un resumen del formato de estos ítems con sus respectivas opciones de respuestas.

**Tabla 3.** Variables y formatos de respuesta

Preferencia humano vs. máquina	Dicotómico (Opción 1= <i>Humano</i> , opción 2= <i>Máquina</i> )
Fiabilidad atribuida a sendos agentes	Escala likert de 1 ( <i>Nada fiable</i> ) a 5 ( <i>Completamente fiable</i> )
Nivel de riesgo atribuido a la tarea	Escala likert de 1 ( <i>Ninguno o muy bajo riesgo</i> ) a 10 ( <i>Muy alto riesgo</i> )
Familiaridad respecto de la tarea	Escala Likert de 1 ( <i>Nada familiarizado</i> ) a 5 ( <i>Muy familiarizado</i> )

La administración de estos dos instrumentos permitirá evaluar si existe correlación entre las dos subescalas que componen el constructo sobre el que se sustenta el uPAS (Apéndice 1) y las respuestas de la escala de elaboración propia. La relación entre ambos instrumentos que sugieran los resultados obtenidos permitirá plantear la necesidad de actualización del instrumento validado, por si este debiera agudizarse con el fin de evaluar de forma ajustada la confianza en las nuevas formas de desarrollo digital de la era tecnológica en que vivimos.

### **2.5 Análisis de datos**

Todos los análisis se realizaron con el software R v4.3.2. En primer lugar, se analizaron las variables sociodemográficas mediante análisis descriptivo. En segundo lugar, se validó la adaptación en español del uPAS (Gibson *et al.*, 2023) en esta muestra utilizando (1) análisis paralelo para determinar el número de factores subyacentes, y (2) análisis factorial exploratorio para examinar la estructura factorial del test, utilizando la solución de mínimos residuos (*minres*) y rotación *oblimin*. Se extrajeron las puntuaciones factoriales y se utilizaron como variables independientes en el estudio. Todos los análisis factoriales se realizaron utilizando el paquete *psych* de R v2.4.3 (Revelle, 2024). En tercer lugar, se ajustó un modelo lineal generalizado de efectos mixtos (GLMM) con enlace logit (familia binomial) y estimador de máxima verosimilitud para la variable dependiente dicotómica (ítems de preferencia humano vs. máquina) para cada grupo de edad, incluyendo estas variables independientes: género, área de conocimiento, nivel de riesgo atribuido a la tarea, frecuencia de uso de herramientas basadas en IA, factor “pensamiento

dicotómico” de la escala *uPAS* (*uPAS1*) y factor “expectativas” de la escala *uPAS* (*uPAS2*) (estas dos últimas variables solo estaban disponibles para adolescentes y adultos). En cuarto lugar, se ajustó un modelo lineal de efectos mixtos (LMM) con estimador REML para la variable dependiente continua (grado de fiabilidad atribuida al agente humano y a la máquina) para cada grupo de edad, incluyendo las mismas variables independientes más el efecto humano vs. máquina y sus efectos de interacción con el resto de las variables. Este modelo tenía como objetivo examinar posibles efectos moderadores de estas variables sobre la relación entre la fiabilidad percibida y el agente (humano vs. máquina). El modelo lineal generalizado de efectos mixtos y el lineal se ajustaron utilizando el paquete *lme4* de R v1.1-35.3 (Bates et al., 2014) utilizando efectos aleatorios cruzados para individuos e ítems para estimar la variabilidad de los efectos fijos en ambos clústeres y evaluar si los efectos fijos son generalizables a todos los niveles del clúster. Esta es una práctica común en la psicología experimental, incluyendo variables dependientes cuantitativas (Baayen et al., 2008; Martínez-Huertas et al., 2022) y dicotómicas (Quené & van den Bergh, 2008). En particular, utilizamos intersecciones aleatorias para individuos e ítems en ambos modelos, y los predictores continuos se estandarizaron antes de estimarlos para facilitar la interpretabilidad de los resultados.

## RESULTADOS

### *3.1 Preferencia humano vs. máquina*

La Tabla 3 presenta los resultados del GLMM con efectos aleatorios cruzados para individuos e ítems que evalúan las relaciones entre las variables independientes y la preferencia de agente humanos vs. máquina en cada grupo de edad. Los tres grupos presentaron una tendencia general a preferir a los humanos en comparación con las máquinas en el estudio, pero encontramos algunas diferencias entre los grupos de edad.

En el primer grupo, solo encontramos diferencias en cuanto al género ( $b = -.532$ ,  $SE = .206$ ,  $z = -2.582$ ,  $p = .0089$ ), lo que se puede traducir en una razón de probabilidades de 0.587, mostrando que las mujeres generalmente prefieren a los humanos más a menudo que los hombres, que generalmente prefieren a las máquinas más a menudo. En el segundo grupo, las mujeres también prefirieron a los humanos más a menudo que los

hombres ( $b = -.345$ ,  $SE = .145$ ,  $z = -2.380$ ,  $p = .017$ ), con un tamaño del efecto menor ( $OR = .708$ ). El riesgo atribuido a la tarea presentó una relación negativa con la preferencia humano vs. máquina ( $b = -.072$ ,  $SE = .025$ ,  $z = -2.885$ ,  $p = .0039$ ), con una razón de probabilidades de  $0.931$ , indicando que hay un aumento del 7% en la preferencia por humanos por cada desviación estándar del riesgo percibido. Por otro lado, hubo una relación positiva de la variable expectativas con la variable dependiente ( $b = .387$ ,  $SE = .077$ ,  $z = -4.994$ ,  $p < .001$ ) presentando una razón de probabilidades de  $1.473$ , lo que se traduciría en un mayor aumento por cada desviación estándar de la variable expectativas. En el tercer grupo, encontramos el mismo patrón de resultados con tamaños de efecto similares, excepto en el caso del género (que no presentó diferencias estadísticamente significativas). En particular, la variable dependiente presentó una relación negativa con el riesgo ( $b = -.075$ ,  $SE = .029$ ,  $z = -2.534$ ,  $p = .011$ ,  $OR = .928$ ) y una relación positiva con las expectativas ( $b = .480$ ,  $SE = .090$ ,  $z = 5.315$ ,  $p < .001$ ,  $OR = 1.616$ ). Las desviaciones estándar de las intersecciones de los ítems muestran que la selección media de los individuos presentó alguna variabilidad dependiendo de la pregunta específica (ver la siguiente sección para un análisis más detallado sobre este punto en relación con la fiabilidad percibida).



Tabla 4. Resultados del modelo lineal generalizado de efectos mixtos (GLMM) con enlace *logit* (familia binomial) (variable dependiente: 0 = humano, 1 = robot).

Parámetros	NIÑOS			ADOLESCENTES			ADULTOS			
	Estimación	SE	Valor-z	Estimación	SE	Valor-z	Estimación	SE	Valor-z	
Efectos fijos	Intersección	-1.515	.342	-4.435***	-1.231	.328	-3.750***	-2.036	.466	-4.365***
	Género	-.532	.206	-2.582**	-.345	.145	-2.380*	-.207	.207	-1.001
	Rama SS vs. “otra”	.097	.232	.417	-.231	.191	-1.211	.289	.305	.947
	Rama CT vs. “otra”	.066	.250	.264	-.120	.191	-.626	.582	.317	1.839
	Riesgo	-.033	.029	-1.123	-.072	.025	-2.885**	-.075	.029	-2.534*
	Uso IA	.099	.080	1.234	.115	.085	1.358	.167	.106	1.565
	Pensamiento dicotómico	-	-	-	.051	.076	.667	-.170	.098	-1.727
	Expectativas	-	-	-	.387	.077	4.994***	.480	.090	5.315***
Efectos aleatorios	SD intersecciones individuos		.923			.668			1.015	
	SD intersecciones ítems		.915			.915			1.169	
	SD residuos		1.027			1.224			1.753	

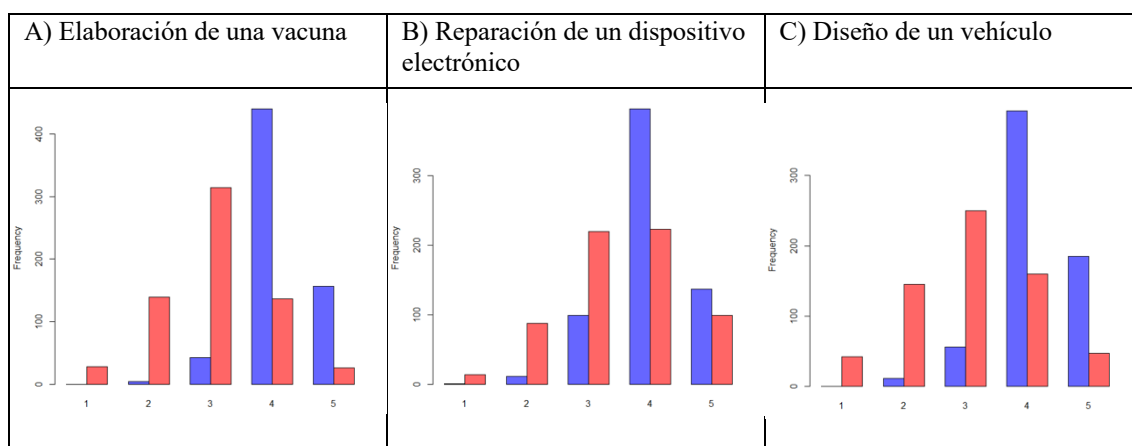
Nota. \*\*\* =  $p < .0001$ . \*\* =  $p < .01$ . \* =  $p < .05$ . SE = Error típico. SD = Desviación típica.

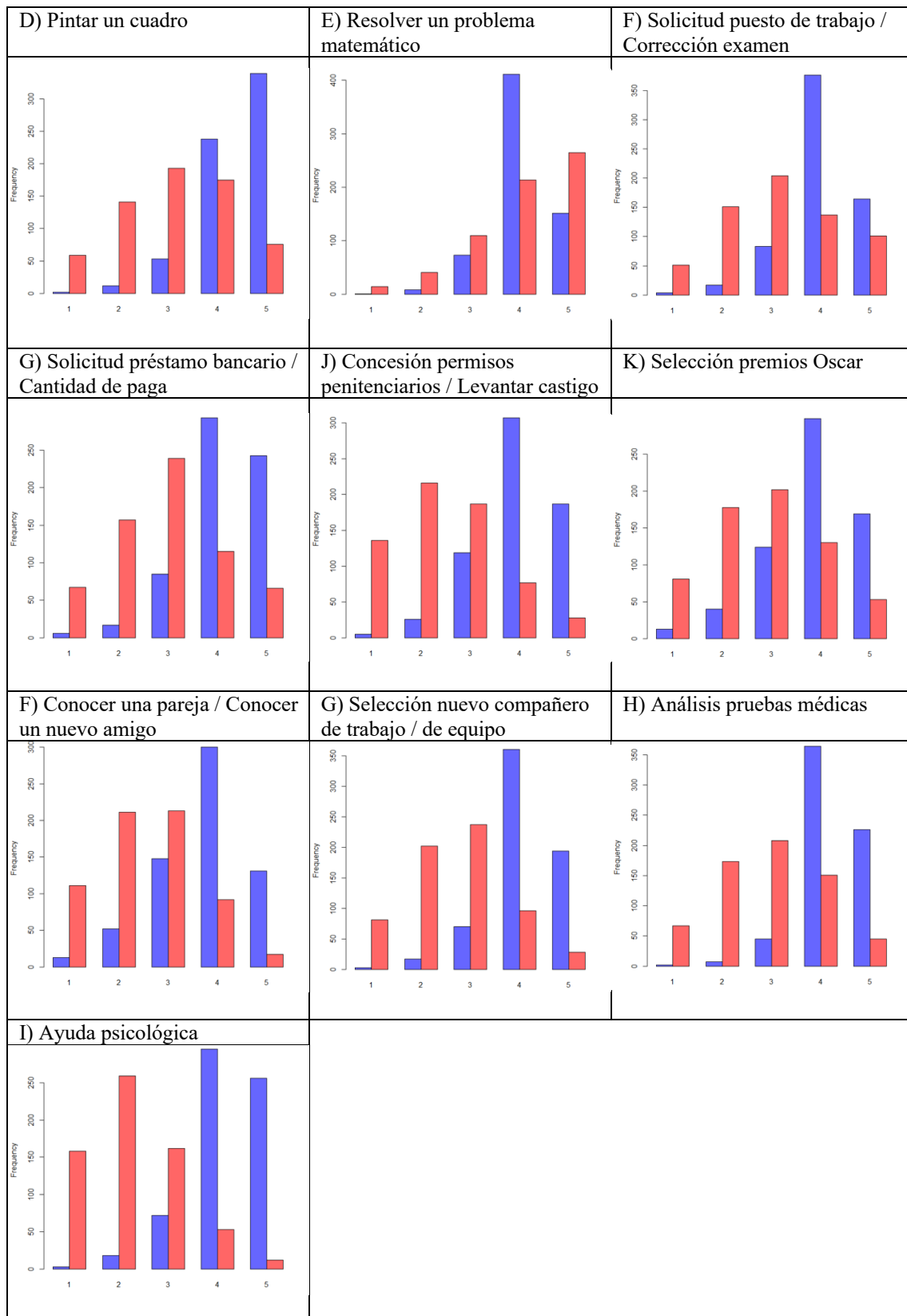
Rama SS= Rama sociosanitaria; Rama CT= rama científico tecnológica; “Otra” = no especialización.

### 3.2 Fiabilidad atribuida agente humano vs. máquina

En la sección anterior, analizamos la preferencia (humano vs. máquina) en diferentes contextos. En esta sección, nuestro objetivo es atribuir un valor cuantitativo para medir esta preferencia, yendo desde el valor “1=Nada fiable” hasta “5=Completamente fiable”. La Figura 1 presenta la distribución de las fiabilidades atribuidas a humanos y robots en los diferentes contextos (ítems) del estudio. En general, podemos ver que los humanos son percibidos como más fiables que las máquinas. Los escenarios planteados a los participantes se circunscriben a diversas áreas: por un lado, uno referente a cada una de las siglas STEAM y, por otro, situaciones respectivas a la esfera académico laboral, económica, social, afectivo-relacional y de la salud. En todas las áreas se ha encontrado una tendencia muy similar en los niveles de fiabilidad atribuida a las máquinas, siendo destacable que en todas ellas los humanos obtienen un mayor nivel de fiabilidad. Es en el caso de “resolver un problema matemático” el único donde se observa que la gráfica es creciente en el caso de la fiabilidad atribuida a las máquinas. Algo similar ocurre en la situación de “reparar un dispositivo electrónico”, tarea para la que las máquinas son consideradas fiables por los participantes. Los ítems relativos al área relacional (“selección de un nuevo compañero de equipo”); social (“concesión de permisos penitenciarios”) y de salud (“recibir ayuda psicológica”) son los que muestran una menor fiabilidad percibida de las máquinas. Es destacable además que, en este último caso, este escenario es el que muestra una mayor diferencia entre la fiabilidad atribuida a cada agente, siendo muy superior la referida al humano respecto de la máquina.

**Figura 1.** Fiabilidad atribuida al agente humano (azul) y a la máquina (rojo) en los ítems del estudio.





La Figura 1 mostró que, en general, los humanos son percibidos como más fiables que los robots en estos contextos. Esto se corroboró utilizando LMMs con efectos aleatorios cruzados para individuos e ítems, donde la fiabilidad atribuida (variable

dependiente) se relacionó con el agente (humano vs. robot), mostrando un efecto estadísticamente significativo a favor de los humanos en comparación con las máquinas en niños ( $b = -1.292$ ,  $SE = .028$ ,  $t = -46.50$ ,  $p < .001$ ), adolescentes ( $b = -1.164$ ,  $SE = .022$ ,  $t = -52.05$ ,  $p < .001$ ) y adultos ( $b = -1.030$ ,  $SE = .021$ ,  $t = -48.27$ ,  $p < .001$ ). Esto significa que, en general, la muestra atribuye una mayor fiabilidad a los humanos que a los robots en las diferentes situaciones del estudio. Luego, se realizó un segundo conjunto de LMMs con efectos aleatorios cruzados para individuos e ítems para evaluar los posibles efectos moderadores de diferentes variables independientes en dicha diferencia percibida (por lo tanto, no interesan los efectos principales de estas variables). La Tabla 4 presenta los parámetros estimados de este modelo.

Los resultados muestran que el género y el riesgo fueron dos variables moderadoras en los tres grupos, lo que significa que la diferencia percibida de fiabilidad atribuida fue mayor para las mujeres que para los hombres y para los individuos que percibieron mayor nivel de riesgo asociado a cada situación. De manera similar, se encontró que la variable expectativas moderaba esta relación, ya que los individuos que puntuaban más alto en la subescala de expectativas en el rendimiento de las máquinas, reportaron diferencias menores en la fiabilidad atribuida entre humanos y máquinas. En el grupo de adultos, también encontramos un efecto moderador relevante de la variable pensamiento dicotómico, lo que significa que los adultos con mayor pensamiento dicotómico atribuyen más diferencias en la fiabilidad entre humanos y máquinas. Otros efectos moderadores estadísticamente significativos, más cercanos a  $\alpha = 0.05$  e idiosincrásicos de algunos de los grupos, fueron un efecto moderador de la diferencia entre la rama sociosanitaria (SS) y la rama no cualificada (“otra”) en adolescentes y un efecto moderador de la diferencia entre la rama científico-tecnológica (CT) y la rama no cualificada en adultos. En ambos casos, la rama no cualificada presentaría una relación más positiva entre la fiabilidad atribuida y el agente (humano vs. robot).

Tabla 5. Resultados del modelo lineal de efectos mixtos (LMM) (variable dependiente: fiabilidad atribuida).

Parámetros		NIÑOS			ADOLESCENTES			ADULTOS		
		Estimación	SE	Valor-t	Estimación	SE	Valor-t	Estimación	SE	Valor-t
Efectos	Intersección	4.104	.095	43.022***	4.054	.099	40.733***	3.884	.123	31.539***
fijos	Humano vs. máquina	-0.984	.076	-	-0.838	.074	-	-0.666	.079	-8.462***
	Género	.039	.074	.527	-.108	.054	-1.990*	.157	.057	2.757**
	Rama SS vs. "otra"	.040	.084	.473	.186	.071	2.597**	.069	.081	.849
	Rama CT vs. "otra"	.020	.0901	.221	.144	.073	1.988*	.077	.085	.906
	Riesgo percibido	.005	.009	.555	-.007	.007	.980	-.004	.007	.556
	Frecuencia de uso IA	-.048	.029	-1.644	-.038	.032	-1.191	-.014	.029	-.502
	Pensamiento dicotómico	-	-	-	.028	.029	.982	.062	.027	2.330*
	Expectativas	-	-	-	-.097	.029	-3.376***	-.006	.024	-.246
	Hum/máq*Género	-.251	.060	-4.202***	-.150	.048	-3.100**	-.110	.048	-2.314*
	Hum/máq*Rama SS vs. otra	.165	.068	2.420*	-.027	.063	-.424	.093	.067	1.382
	Hum/máq*Rama CT vs. otra	-.005	.073	-.063	-.046	.065	-.705	.179	.071	2.520*

	Hum/máq *Riesgo	-0.044	.011	-4.105***	-0.047	.009	-5.094***	-0.065	.008	-8.381***
	Hum/máq *Uso IA	.031	.023	1.341	-.038	.029	-1.334	-.029	.024	-1.197
	Hum/máq *P. dicotómico	-	-	-	.022	.026	.864	-.076	.022	-3.420***
	Hum/máq *Expectativas	-	-	-	.258	.026	10.080***	.252	.020	12.390***
Efectos aleatorios	SD intersecciones individuos		.363			.302			.303	
	SD intersecciones ítems		.201			.243			.322	
	SD residuos		.916			.886			.796	

Nota. \*\*\* =  $p < .0001$ . \*\* =  $p < .01$ . \* =  $p < .05$ . SE = Error típico SD = Desviación típica.

Rama SS= Rama sociosanitaria; Rama CT= rama científico tecnológica; Otra= no especialización.

## DISCUSIÓN

El objetivo del presente trabajo fue estudiar el fenómeno de la confianza de los seres humanos en las máquinas, sabiendo que el uso de herramientas basadas en IA y la implementación de los avances tecnológicos en la vida diaria depende, en gran medida, de la confianza que se deposite en ellas. Las diferencias a nivel individual no son algo baladí a este respecto, de forma que, conforme a la teoría y a los resultados obtenidos en este estudio, podemos decir que existen variables que intervienen en el grado de confianza puesto sobre las máquinas.

La muestra se dividió en tres franjas etarias para dilucidar qué variables moderan el nivel de confianza en las distintas edades. La exposición a las herramientas digitales no es la misma en la niñez, la adolescencia o la adultez; por lo que era esperable encontrar diferencias en la confianza según este criterio. Pese a ello, en los tres grupos de edad la tendencia revela una mayor confianza en los seres humanos que en las máquinas para la realización de las tareas planteadas. Tal y como refleja el FECYT en su investigación (2022), esto podría deberse al acuerdo mayoritario de las personas con la idea de que, ante el desconocimiento de las posibles consecuencias de una nueva tecnología, es preferible actuar con cautela y controlar su uso para proteger la salud.

Es destacable que, en el caso de los niños se encontraron diferencias significativas respecto al género, de forma que los niños confiaban en las máquinas más a menudo que las niñas. Encontrar diferencias significativas de género en este grupo puede ser un indicador de que los roles de socialización ejercen una fuerte influencia ya desde edades muy tempranas, lo que tal vez justificaría parte de la tendencia de los hombres a cursar estudios insertos en el ámbito científico-tecnológico y estudios sociosanitarios en el caso de las mujeres, siendo la presencia de las máquinas muy diferente en sendos contextos académicos (Grañeras-Pastrana et al., 2022).

Las diferencias entre hombres y mujeres a este respecto se suavizan a medida que se avanza en edad, de forma que este mismo efecto se encontró también en los adolescentes con un tamaño del efecto menor, pero no se encontró en el grupo de adultos. A pesar de resultar contraintuitivos en un inicio, estos resultados pueden explicarse por la existencia de otra serie de variables que explican las diferencias en la confianza en los adultos en mayor medida que lo hace el género, por ejemplo, la familiaridad con el uso

de herramientas tecnológicas o el nivel de riesgo percibido en la tarea que se va a desempeñar. Este último factor fue significativo tanto en adolescentes como en adultos. A medida que aumenta el riesgo percibido en una tarea, aumenta la preferencia por delegar en seres humanos en lugar de en máquinas. Esto refuerza la hipótesis de que el riesgo situacional es una de las variables contextuales más importantes que afectan la confianza en la automatización (Hoesterey & Onnasch, 2023), de forma que las personas tienden a reducir su confianza en la automatización en tareas de mayor riesgo (Hoff & Bashir, 2014). Sin embargo, esto contradice el estudio de Bogert *et al.* (2021), que afirma que las personas delegan más en algoritmos cuando perciben que la tarea se vuelve más compleja. No obstante, estos autores puntualizan que este efecto ocurre en tareas de tipo intelectual, y en este estudio se han tratado de recoger diferentes áreas y no solo de tipo intelectual, por lo que la forma de juzgar la dificultad o el riesgo de una tarea puede variar según la naturaleza intrínseca de la misma. El nivel de riesgo no ha resultado ser un factor explicativo de la preferencia de los humanos o las máquinas por parte de los niños, pudiendo ser el motivo de ello la ausencia de la percepción del riesgo a esta edad, que se relaciona con la capacidad cognitiva y con la madurez psicosocial y que no alcanzan sus niveles funcionales hasta los 16 años y los 18 años, respectivamente (Icenogle et al., 2019).

Por otro lado, se encontró una relación positiva entre las expectativas en el rendimiento de las máquinas (medida en la escala uPAS) y la tendencia a preferirlas sobre los humanos. Es decir, quienes confían más en el desempeño de los sistemas automatizados y asumen sus decisiones reportan una mayor confianza en las máquinas que quienes tienen expectativas más bajas. Ello apunta a que existe una linealidad en cuanto al planteamiento de los escenarios planteados en el cuestionario y la forma en que la escala uPAS recoge la dimensión “expectativas”, por lo que se han obtenido los resultados esperables a nivel teórico (Dzindolet et al., 2003). En la medida en que nuestras experiencias nos configuran como personas, las interacciones mantenidas con herramientas automatizadas en el pasado sentarán las bases de lo que de ellas cabe esperar (Hoff & Bashir, 2014). Aludiendo a esta confianza aprendida, estos autores indican que las experiencias pasadas con un sistema automatizado o una tecnología similar podría alterar el proceso de formación de la confianza. Teniendo esto en cuenta, futuras investigaciones podrían versar acerca de la influencia que tienen estas interacciones pasadas en la relación actual que se establece con las máquinas. Comprender el potencial



explicativo de esta variable en la interacción de los humanos y la automatización permitiría arrojar luz acerca de qué características favorecen o dificultan el uso de los dispositivos automatizados (Madhavan & Wiegmann, 2005), lo que permitiría orientar en función de ello las políticas de prevención de riesgos o de promoción del uso de la tecnología (Comisión Europea, 2018), cuidando que se empleen en la forma y en la medida apropiada.

En este estudio no solo pretendimos evaluar la confianza en los expertos humanos en comparación con los sistemas automatizados especializados en diferentes áreas, sino que se pretendió cuantificar en qué medida era percibida la fiabilidad de uno y otro agente, analizando cómo los diferentes contextos planteados podrían influir en ello. Según los resultados, son los ítems del ámbito científico-tecnológico los dos en los que las máquinas han obtenido los mayores índices de fiabilidad reportados. En contraposición, el ámbito de la salud mental es el que muestra una menor fiabilidad atribuida a las máquinas, guardando en este ámbito la mayor distancia con respecto de la fiabilidad referida al experto humano. El discurso social actual que acentúa el peligro de que las máquinas terminen por conquistar todos y cada uno de los espacios propios de los seres humanos, cobra aquí especial importancia al hacerse evidente que, en realidad, existen ciertas cuestiones eminentemente humanas en las que parece que a las máquinas aún les queda un largo recorrido hasta alcanzarlas.

El hecho de que prefiramos a las máquinas para aquellas tareas numéricas o automáticas, como resolver un problema de matemáticas, va en línea con el estudio de Bogert et al., (2021), donde afirman que las personas delegan en los algoritmos cuando perciben mayor dificultad intelectual en la tarea. Sin embargo, parece que no ocurre lo mismo en aquellas áreas reservadas al espectro humano como son la capacidad de establecer vínculos relacionales, factor que en el contexto terapéutico se traduce como alianza y que sienta su base, precisamente, en las expectativas del paciente y la confianza que deposita en el terapeuta. Desde aquí, el hecho de que las personas atribuyan un alto nivel de fiabilidad en los humanos en el contexto de la salud mental supone una garantía para los profesionales de este ámbito, al distar mucho la posibilidad de que su función sea fácilmente sustituible por las máquinas. De forma similar se distribuye la fiabilidad atribuida a las máquinas en el contexto social, medido a través de la concesión de permisos penitenciarios, siendo mínima en comparación con la que se atribuye a los humanos. Según los autores, la confianza en la automatización depende del contexto de

la tarea (Appelganc *et al.*, 2022; Castelo *et al.*, 2019), lo que denominaríamos confianza situacional (Hoff & Bashir, 2014), el cual se ve influido por el nivel de riesgo asociado a ella. En línea con los hallazgos de Weber *et al.*, (2002), los resultados confirman que en los escenarios que conciernen a la seguridad pública (permisos penitenciarios) se ha reportado un menor grado de confianza en las máquinas que los que respectan a asuntos financieros (concesión de préstamo bancario). Esto también se replica al comparar el contexto económico con el de salud mental. Al considerarse este último un escenario de mayor riesgo, se confía menos en las máquinas en el caso de la salud mental. Sin embargo, esto no ocurre al comparar el contexto económico con el de salud médica, pues los resultados revelan que se confía más en las máquinas para analizar pruebas médicas y emitir un diagnóstico que para conceder un préstamo bancario. Esto podría deberse a la rápida e incesante evolución de los dispositivos tecnológicos en el contexto sanitario (Raparathi, 2021) y al aumento de la percepción de inseguridad ciudadana actual (Pueyo *et al.*, 2018), de forma que la confianza aumenta en el primero de los contextos, pero disminuye en el segundo.

Pese a que estos datos ponen de relieve la diferencia cuantitativa en los niveles de fiabilidad que se atribuyen a los humanos por encima de las máquinas para todos estos contextos, es cierto que no podemos considerar esta medida como representativa de la confianza en su dimensión más estricta y completa. Al tratarse de un constructo subjetivo, es probable que existan otra serie de factores que influyan en la confianza y que poco tengan que ver con la fiabilidad atribuible a un agente a la hora de realizar una tarea. Así, aspectos relativos a la personalidad, otros de tipo social o incluso a nivel cultural, podrían estar influyendo en los niveles de confianza desarrollados por los sujetos de nuestra muestra hacia los contextos planteados (Lee & See, 2004). Además, pensar en términos de fiabilidad implica situar el foco en el resultado final de las tareas (por ejemplo, en la resolución de un problema de matemáticas) pero en otros muchos contextos lo importante reside más bien en los procesos, tal y como ocurre en el contexto terapéutico. Así, futuras líneas de investigación podrían dirigirse a explorar cuáles son esos esos matices que median en la confianza en la interacción entre los humanos y las máquinas, si son propios de las tareas en sí y de qué manera afectan a ello las diferencias culturales en cuanto a las expectativas y al conocimiento previo sobre el funcionamiento de estas herramientas, con el fin de dilucidar su impacto en la confianza que se deposita sobre ellas y los comportamientos que de ello derivan.

Como factores explicativos de la fiabilidad atribuida a los humanos y a las máquinas se pone de relieve la influencia del género y de la percepción del riesgo. En general, las mujeres percibieron mayores diferencias entre la fiabilidad que les merecían los humanos y las máquinas en comparación con los hombres. También se observó que las personas que atribuyeron un mayor nivel de riesgo a una situación determinada reportaron diferencias mayores en la fiabilidad que atribuían a ambos agentes en cada situación planteada. Estos hallazgos son coherentes con investigaciones previas que sugieren que las mujeres tienden a ser más cautelosas y reticentes en sus actitudes hacia la fiabilidad y el uso de las máquinas (Comisión Europea, 2018), como también ocurre en contextos en los que el riesgo percibido es mayor (Ezer et al., 2008).

En relación con la escala uPAS, se encontró un efecto moderador entre sus dos factores (expectativas y pensamiento dicotómico) y la fiabilidad reportada respecto del agente humano y el robot. Así, quienes tenían expectativas más altas en la automatización, reportaron diferencias menores en la fiabilidad atribuida a humanos y máquinas. Estos hallazgos son coherentes con estudios previos que han identificado que las expectativas sobre las máquinas pueden aumentar la percepción que las personas sostienen respecto de su fiabilidad (Ezer *et al.*, 2007; Merrit et al., 2015). En contraste, el pensamiento dicotómico mostró un efecto moderador relevante en el grupo de adultos, donde aquellos con mayores puntuaciones atribuyeron más diferencias en la fiabilidad entre humanos y máquinas. Siguiendo a Merrit et al., (2015) este hecho podría asociarse con que este tipo de pensamiento todo o nada en la escala se relaciona con la disminución de la confianza en la automatización tras experimentar errores cometidos. El hecho de que los adultos hayan podido tener más experiencias de uso de herramientas automatizadas (y por tanto haber estado expuestos a más posibles fallos), así como a su ya forjada percepción del riesgo, explicaría que esta subescala se relacione con un mayor grado de diferencia en las fiabilidades atribuidas a los humanos y a las máquinas, lo cual no ha resultado ser significativo en el grupo de los adolescentes.

Con todo ello, en los albores del auge de la inteligencia artificial y su diseminación en las diferentes áreas de nuestra vida, en el momento actual parece que los humanos todavía prefieren a los humanos frente a otras formas de ayuda como es el caso de la inteligencia artificial. Frente a otros estudios donde se concluye que los humanos confían excesivamente en la IA, incluso cuando cometen fallos evidentes (Vicente & Matute, 2023), nuestra investigación revela que, al menos en un nivel decisional, las personas

prefieren delegar en la ayuda humana que en la artificial. Sin embargo, es innegable el hecho de que las herramientas digitales impregnan ya todas las áreas de nuestra vida, por lo que facilitar una confianza adecuada en la automatización es clave para mejorar la seguridad y la productividad de los equipos humano-máquina (Hoff & Bashir, 2014). Siendo el uso indebido de estas herramientas uno de los hechos que más preocupa a los teóricos debido a los peligros que ello puede ocasionar (Parasuraman & Riley, 1997), informar y educar a la población sobre las potencialidades y los riesgos de estas herramientas digitales resulta clave para conseguir que jueguen a nuestro favor y que no actúen en detrimento de nuestros propios objetivos. Por tanto, la confianza sí que resulta ser un prerrequisito necesario para la incorporación de los dispositivos automatizados en nuestro día a día. La parte positiva de ello es que, pese a las influencias culturales y sociales que la condicionan, la confianza también se puede moldear y se puede aprender, hasta alcanzar el grado óptimo que nos permita acomodarnos a una realidad en la que logremos poner el progreso tecnológico a nuestro servicio, sin vivir subsumidos en su poder ni tampoco ajenos al mismo.

#### ***4.1 Limitaciones***

Pese a que la división de la muestra por edades juega un rol central tanto en la metodología del estudio como en los resultados obtenidos, el presente estudio no es capaz de discutir la influencia de dicha variable a la luz de la escala uPAS debido a que no existe una versión validada adaptada a población infantil. Además de esto, los ítems de elaboración propia podrían no guardar un paralelismo exacto en la versión de adultos y la versión de menores. Sin embargo, el objetivo primordial fue explorar qué preferencias tienen los sujetos y qué fiabilidad atribuyen a las máquinas en distintas situaciones de su vida diaria, ajustadas según su franja etaria, por lo que los matices diferenciales que puedan albergar los enunciados planteados en ambos grupos son mínimos y no impiden que puedan compararse las tendencias encontradas en niños, adolescentes y adultos en los resultados del estudio.

Por otro lado, no existe una versión adaptada al castellano del instrumento, sino que este ha sido aplicado tras la traducción por parte de las investigadoras del estudio. Sin embargo, teniendo en cuenta el ajuste del análisis factorial y sus similitudes con el instrumento original replicadas en este estudio, no creemos que haya diferencias respecto de la interpretación de las subescalas del test original en nuestra muestra con su versión en inglés. Ahora bien, resulta cabal proponer la validación de la escala al castellano en

aras de realizar estudios multiculturales que permitan generalizar los resultados a muestras poblacionales más amplias, de forma que podría aplicarse en contextos socioculturales con características diferenciales, dotando así a los resultados de mayor fundamento y fiabilidad.

## CONCLUSIÓN

Pese a la concepción imperante de que las máquinas son más confiables que los humanos, este estudio revela que los humanos todavía prefieren confiar en la pericia humana para el desempeño de las tareas que comprometen diferentes áreas de la vida diaria. Incluso los niños, quienes se han desarrollado en el seno de la era digital, otorgan una mayor fiabilidad a la ayuda humana que a la automatizada. Esto implica una garantía para las figuras de referencia de los más pequeños, quienes todavía parecen ser los máximos influyentes en los procesos de socialización. Tanto hombres como mujeres reportan un grado similar de confianza en la tecnología, pero no ocurre lo mismo en el caso de niños y niñas, siendo los primeros quienes más confían en la ayuda automatizada. Trabajar para reducir la presencia de los roles de género en este ámbito es crucial para que exista una representación ecuánime de niños y niñas en los diferentes ámbitos académicos. Los estudiantes de hoy serán los profesionales del mañana, por lo que alcanzar la utopía de un desarrollo profesional igualitario en lo que al género respecta es un trabajo que comienza desde las edades más tempranas, lo que incluye fomentar de forma paritaria la motivación por el ámbito digital.

Es muy importante conocer, conceptualizar y delimitar las características que modulan la confianza en la automatización. Factores como las expectativas que se depositan sobre estas herramientas, la medida en que se toleran sus posibles errores o el riesgo percibido respecto de la tarea que está en juego, son aspectos determinantes en las actitudes y en las decisiones que se toman acerca de recurrir al soporte automatizado o no. Tal vez haya áreas donde la asistencia digital termine por relegar a los humanos a un segundo plano, pero por lo que en este estudio ha podido comprobarse, existe aún un margen importante donde los humanos ostentan la práctica totalidad del dominio. Mientras tanto, lo más inteligente reside en integrar el apoyo automatizado para propulsar, mejorar y enriquecer las potencialidades humanas para dotar de eficiencia, precisión y creatividad la confianza que tradicional y actualmente se deposita en las personas.

## REFERENCIAS

- Appelganc, K., Rieger, T., Roesler, E., & Manzey, D. (2022). How much reliability is enough? A context-specific view on human interaction with (artificial) agents from different perspectives. *Journal of Cognitive Engineering and Decision Making* 16(4), 207-221.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. <https://dx.doi.org/10.1016/j.jml.2007.12.005>.
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://dx.doi.org/10.18637/jss.v067.i01>.
- Bogert, E., Schechter, A. & Watson, R.T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports* (11). <https://doi.org/10.1038/s41598-021-87480-9>
- Cáceres, M.; Raso, F., Rodríguez, A., & Romero, J. (2017). La elección de carrera desde un enfoque de género. Factores clave e implicaciones socioeducativas. Disponible en: <https://www.educaweb.com/noticia/2017/11/29/eleccion-carrera-enfoque-genero-factores-clave-implicaciones-socioeducativas-16203/>.
- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5), 809–825. <https://doi.org/10.1177/0022243719851788>
- Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science education*, 67(2), 255-265.
- Cheney, I., Shattuck, S. (2020). *Picture a Scientist* [Documental]. Wicked Delicate Films.
- Chiou, E., & Lee, J. (2023). Trusting automation: Designing for responsivity and resilience. *Human factors* 65(1), 137-165. DOI: <https://doi.org/10.1177/00187208211009995>
- Comisión Europea (2020). Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza. <https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>
- Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, (2018). *Women in the digital age: executive summary*, Publications Office. <https://data.europa.eu/doi/10.2759/517222>
- Dzindolet, M., Dawe, L., Beck, H., & Pierce, L. (2001). *A Framework of Automation Use (Report No. ARL-TR-2412)*. Aberdeen Proving Ground, MD: Army Research Laboratory. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a390225.pdf>.
- Dzindolet, M., Pierce, L., Beck, H., & Dawe, L. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors*, 44, 79–97.

- Dzindolet, M., Peterson, S., Pomranky, R., Pierce, L., & Beck, H. (2003). The role of trust in automation reliance. *International Journal of Human- Computer Studies*, 58, 697-718. 10.1016/S1071-5819(03)00038-7.
- Ezer, N., Fisk, A. D., & Rogers, W. A. (2007). Reliance on automation as a function of expectation of reliability, cost of verification, and age. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 6–10). Santa Monica, CA: Human Factors and Ergonomics Society.
- Ezer, N., Fisk, A. D., & Rogers, W. A. (2008). Age-related differences in reliance behavior attributable to costs within a human decision aid system. *Human Factors*, 50, 853–863.
- FECYT (2022). Encuesta de Percepción Social de la Ciencia y la Tecnología. Informe de resultados. Ministerio de Ciencia e Innovación. Retrieved from: <https://www.fecyt.es/es/publicacion/percepcion-social-de-la-ciencia-y-la-tecnologia-en-espana-2022>
- Fiske, S. 1998. “Stereotyping, Prejudice, and Discrimination.” In *The Handbook of Social Psychology*, edited by D. Gilbert, S. Fiske, & G. Lindzey. 4th ed., Vol. 2, 257–411. New York: McGraw-Hill.
- Gibson, A. M., Capiola, A., Alarcon, G. M., Lee, M. A., Jessup, S. A., & Hamdan, I. A. (2023). Construction and validation of an updated perfect automation schema (uPAS) scale. *Theoretical Issues in Ergonomics Science*, 24(2), 241-266. Doi: 10.1080/1463922X.2022.2081375
- Glasheen, K.J., Shochet, I., Campbell, M.A. (2016). Online counselling in secondary schools: Would students seek help by this medium? *British Journal of Guidance & Counselling*, 44 (1), 108–122. [http:// dx.doi.org/10.1080/03069885.2015.1017805](http://dx.doi.org/10.1080/03069885.2015.1017805).
- Grañeras Pastrana, M., Moreno Sánchez, M.E., & Isidoro Calle, N. (2022). *Radiografía de la brecha de género en la formación STEAM. Un estudio en detalle de la trayectoria educativa de niñas y mujeres en España*. Ministerio de Educación y Formación Profesional.
- Hoesterey, S., & Onnasch, L. (2023). The effect of risk on trust attitude and trust behavior in interaction with information and decision automation. *Cognition, Technology & Work*, 25(1), 15–29.
- Hoff, K. A.; Bashir, M. (2014). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407-434. Doi:10.1177/0018720814547570
- Icenogle, G., Steinberg, L., Duell, N., Chein, J., Chang, L., Chaudhary, N., ... & Bacchini, D. (2019). Adolescents’ cognitive capacity reaches adult levels prior to their psychosocial maturity: Evidence for a “maturity gap” in a multinational, cross-sectional sample. *Law and human behavior*, 43(1), 69-85.
- Instituto Nacional de Estadística. (2022). Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación (TIC) en los Hogares. [https://transparencia.gob.es/transparencia/transparencia\\_Home/index/MasInformacion/Informes-de-interes/Ciencia\\_y\\_tecnologia/EncuestaEquipamientoyusoTICenLosHogares.html](https://transparencia.gob.es/transparencia/transparencia_Home/index/MasInformacion/Informes-de-interes/Ciencia_y_tecnologia/EncuestaEquipamientoyusoTICenLosHogares.html)

- Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2021.106878>
- Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50-80.
- Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human machine systems. *Ergonomics*, 22, 671–691.
- López-Santín, J. M., & Álvaro Serón, P. (2018). La salud mental digital. Una aproximación crítica desde la ética. *Revista de la Asociación Española de Neuropsiquiatría*, 38(134), 359-379. DOI: 10.4321/S0211-57352018000200002
- Lorenz, L., Meijer, A., & Schuppan, T. (2021). The algocracy as a new ideal type for government organizations: Predictive policing in Berlin as an empirical case. *Information Polity*, 26(1), 71–86. <https://doi.org/10.3233/IP-200279>
- Lyons, J. & Guznov, S. (2018). Individual differences in human-machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science* 1(19). 10.1080/1463922x.2018.1491071.
- Madhavan, P., & Wiegmann, D. A. (2005). Effects of information source, pedigree, and reliability on operators' utilization of diagnostic advice. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* 49(3), 487-491. doi:10.1177/154193120504900358
- Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, 37, 465–498.
- Martínez-Huertas, J.A., Olmos, R., & Ferrer, E. (2022). Model selection and model averaging for mixed-effects models with crossed random effects for subjects and items. *Multivariate Behavioral Research*, 57(4), 603-619. <https://doi.org/10.1080/00273171.2021.1889946>.
- Martínez, N., Agudo, U., & Matute, H. (2022). Human cognitive biases present in artificial intelligence. *Revista Internacional de los Estudios Vascos* 67(2).
- Mayer, M. (2011). La utilización de Internet entre los adolescentes, riesgos y beneficios. *Atención primaria*, 43(6), 287-288.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709–734.
- Merritt, S., Unnerstall, J., Lee, D., & Huber, K. (2015). Measuring Individual Differences in the Perfect Automation Schema. *Human factors*, 57. 10.1177/0018720815581247.
- Merritt, S., Ako-Brew, A., Bryant, W., Staley, A., McKenna, M., Leone, A., & Shirase, L. (2019). Automation-Induced Complacency Potential: Development and Validation of a New Scale. *Frontiers in Psychology*, 10, 225. Doi: 10.3389/fpsyg.2019.00225
- Ministerio de Asuntos Económicos y Transformación Digital. (). España Digital 2025. Agenda 2030. Gobierno de España.



- Ministerio de Educación y Formación Profesional. (2019). PISA 18. Programa para la Evaluación Internacional de los Estudiantes. Informe Español. (*Volumen I*). Secretaría General Técnica.
- Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11), 1905-1922.
- OECD (2015), *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, PISA, OECD Publishing. DOI: <http://dx.doi.org/10.1787/9789264229945-en>
- Organización Mundial de la Salud. (2021). Plan de Acción Integral sobre Salud Mental 2013-2030. <https://www.who.int/es/publications/i/item/9789240031029>
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413-425. <https://doi.org/10.1016/j.jml.2008.02.002>.
- Parasuraman, R., Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230–253. doi:10.1518/001872097778543886
- Poornikoo, M. & Ivar Ø. K. (2023). Model evaluation in human factors and ergonomics (HFE) sciences; case of trust in automation. *Theoretical Issues in Ergonomics Science*. DOI: 10.1080/1463922X.2023.2233591
- Pueyo, A., Arbach-Lucioni, K., & Redondo, S. (2018). The RisCanvi: a new tool for assessing risk for violence in prison and recidivism. *Handbook of recidivism risk/needs assessment tools*, 255-268.
- Raparathi, M. (2021). AI-Driven Decision Support Systems for Precision Medicine: Examining the Development and Implementation of AI-Driven Decision Support Systems in Precision Medicine. *Journal of Artificial Intelligence Research*, 1(1), 11-20.
- Revelle, W. (2024). *Package 'psych' v2.4.3*. CRAN.
- Rieger, T., Roesler, E., & Manzey, D. (2022). Challenging presumed technological superiority when working with (artificial) col-leagues. *Scientific Reports*, 12(1), 3768. <https://doi.org/10.1038/s41598-022-07808-x>
- Roesler, E., Rieger, T., & Manzey, D. (2022). Trust towards human vs. Automated Agents: using a multidimensional trust questionnaire to assess the role of performance, utility, purpose and transparency. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66(1), 2047-2051. DOI: 10.1177/1071181322661065
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651–665.
- Rumelhart, D. E. 1980. "Schemata: The Building Blocks of Cognition." In *Theoretical Issues in Reading*, edited by R. Spiro, B. Bruce, and W. Brewer, 33–58. Mahwah: Lawrence Erlbaum Associates.
- Saiz, V. (2021). Adolescentes y el uso de Internet: resumen del estudio uso y riesgos en el uso de Internet de adolescentes escolarizados entre 12 y 17 años con enfoque de explotación

- Sweeney, G. M., Donovan, C. L., March, S., & Forbes, Y. (2019). Logging into therapy: Adolescent perceptions of online therapies for mental health problems. *Internet interventions, 15*, 93–99. <https://doi.org/10.1016/j.invent.2016.12.001>
- Torous, J., Chan, SR., Yee-Marie Tan, S., Behrens, J., Mathew, I., Conrad, EJ., Hinton, L., Yellowlees, P., Keshavan, M. (2014) Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: A survey in four geographically distinct psychiatric clinics. *JMIR Mental Health. 1:e5*.
- Torous, John; Roberts, Laura Weiss (2017). The Ethical Use of Mobile Health Technology in Clinical Psychiatry. *The Journal of Nervous and Mental Disease, 205*(1), 4–8. doi:10.1097/NMD.0000000000000596
- Vicente, L., Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports, 13*(1), 15737.
- Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*(4), 263–290. <https://doi.org/10.1002/bdm.414>
- Xiang, C. (2023). 'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says. *Vice*. Retrieved from: <https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says>

# ANEXOS

## Apéndice 1. Cuestionario

### Preguntas sociodemográficas (\*Solo para *menores*; \*\*Solo para adultos)

Indica a continuación tu Código Postal\*\*

Señala tu género

- Masculino
- Femenino
- Otro

¿Cuántos años tienes?

- 10-15
- 16-20
- 21-25
- 26-30
- 31-40
- 41-50
- 51 o más

Indica tu curso\*

- 6º Primaria
- 1º ESO
- 2º ESO
- 3º ESO
- 4º ESO
- 1º Bachillerato
- 2º Bachillerato

Indica el nombre de tu centro educativo\*

¿En qué ámbito de conocimiento se encuentran tus estudios o trabajo? / ¿En qué área se encuentran los estudios o el trabajo que quieres realizar en el futuro? (Si dudas entre varias opciones, escoge la que ahora mismo refleje mejor lo que más te gusta).

- Ciencias
- Ciencias de la Salud
- Ciencias Sociales y Jurídicas
- Ingeniería y Arquitectura
- Arte
- Otro (si desempeñas un trabajo no cualificado o no relacionado con ninguna de las áreas mencionadas) / Si quieres dedicarte a alguna profesión no cualificada

Indica tu mayor nivel de estudios alcanzado\*\*

- Educación Primaria
- Educación Secundaria
- Bachillerato o Formación Profesional de Grado Medio
- Formación Profesional de Grado Superior
- Carrera universitaria
- Postgrado y/o doctorado

¿En qué medida recurre a consultar información a chats basados en inteligencia artificial, como ChatGPT o Bard?

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy frecuentemente

**Escala uPAS** (Gibson *et al.*, 2020).

*\*Esta escala se administró al grupo de adolescentes y de adultos, no al grupo de niños.*

		Completamente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Completamente de acuerdo
1.	Los sistemas automatizados tienen un rendimiento perfecto.	1	2	3	4	5
2.	<b>Puedo confiar en que los sistemas automatizados tomen decisiones precisas.</b>	1	2	3	4	5
3.	<b>Los sistemas automatizados realizan el trabajo mejor que los humanos.</b>	1	2	3	4	5
4.	<b>Los sistemas automatizados son más fiables que los humanos.</b>	1	2	3	4	5
5.	Los sistemas automatizados pueden realizar tareas difíciles sin errores.	1	2	3	4	5
6.	<b>Acepto las decisiones tomadas por los sistemas automatizados.</b>	1	2	3	4	5
7.	A menudo cuestiono las decisiones de los sistemas automatizados.	1	2	3	4	5
8.	Me sorprende cuando un sistema automatizado comete un error.	1	2	3	4	5
9.	Tengo pocas expectativas cuando utilizo un sistema automatizado por primera vez.	1	2	3	4	5
10.	<b>Los sistemas automatizados que cometen errores deberían abandonarse.</b>	1	2	3	4	5
11.	<b>Me niego a utilizar sistemas automatizados imprecisos.</b>	1	2	3	4	5
12.	<b>Si un sistema automatizado comete un error no lo utilizaré.</b>	1	2	3	4	5
13.	Si los sistemas automatizados tienen problemas, son inutilizables	1	2	3	4	5

14.	<b>Los sistemas automatizados imperfectos deberían ser descartados.</b>	1	2	3	4	5
15.	Para que yo utilice sistemas automatizados deben funcionar perfectamente.	1	2	3	4	5
16.	Cuando los sistemas automatizados cometen errores dejo de utilizarlos.	1	2	3	4	5
17.	Sigo utilizando los sistemas automatizados después de que cometan errores.	1	2	3	4	5
18.	Hay utilidad en utilizar sistemas automatizados que cometen errores.	1	2	3	4	5

*Nota.* Los ítems que van del 1 al 9, constituyen la subescala de “Expectativas”; del 10 al 18 encontramos los ítems de la subescala “Pensamiento todo o nada”. Siguiendo a Gibson et al., (2020), los análisis del estudio se han realizado en torno a los ítems en negrita.

### Preguntas *ad hoc*:

#### - Preguntas áreas STEAM:

Para las siguientes preguntas, responde **a quién preferirías** confiar la realización de las siguientes tareas (bien a un humano o a una máquina), teniendo en cuenta cuál crees que actuaría con un **menor margen de error**.

#### ¿En quién confiarías más para investigar en la elaboración de una vacuna? (Science)

- Un científico experto en investigación
- Un robot diseñado para realizar tareas de investigación

¿Cómo de fiable consideras al experto para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al robot para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

#### ¿En quién confiarías más para resolver un problema técnico en tu móvil u ordenador? (Technology)

- En un experto humano especializado en reparar dispositivos digitales
- A un sistema automático de diagnóstico y reparación de dispositivos digitales

¿Cómo de fiable consideras al experto para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al sistema automático para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

**¿En qué vehículo confiarías más para comprarlo? (Engineering) / ¿En qué vehículo confiarías más para montarte en él? (versión menores)**

- En un coche diseñado por un experto ingeniero mecánico
- Un coche diseñado íntegramente por un robot

¿Cómo de fiable consideras al experto para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al robot para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

**¿En quién confiarías más para que diseñara y pintara un cuadro para decorar tu vivienda? (Art) / ¿En quién confiarías más para que diseñara y pintara un cuadro para decorar tu habitación/clase? (versión menores)**

- En un prestigioso artista
- En un robot diseñado para tareas artísticas

¿Cómo de fiable consideras al experto para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al robot para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

**¿En quién confiarías más para resolver un problema de matemáticas? (Maths)**

- En un experto matemático
- En un robot entrenado en cálculo

¿Cómo de fiable consideras al experto para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al robot para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

**Preguntas de elaboración propia áreas de funcionamiento (versión adultos y adaptación menores)**

**¿En quién confiarías más que valorase tu solicitud/candidatura a un puesto de trabajo? / ¿En quién confiarías más para que te corrigiese un examen?**

- En un experto en recursos humanos / En un profesor experto en la asignatura
- En una máquina entrenada en selección de personal / En una máquina entrenada para detectar aciertos y fallos

¿Cómo de fiable consideras al experto / profesor para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras a la máquina para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de solicitar un puesto de trabajo / **esperar la corrección de un examen?** (Siendo “no me resulta nada familiar” si tú ni nadie de tu entorno lo ha hecho nunca, y “muy familiar” si tú mismo o varias personas de tu entorno se han enfrentado a esta situación).

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------

**¿En quién confiarías más para que resolviese tu solicitud de un préstamo bancario? / ¿En quién confiarías más para que decidiese la paga que te dan?**

- En un experto economista / **En el criterio de papá/mamá/tutor legal**
- En una máquina especializada / **En el criterio de una máquina**

¿Cómo de fiable consideras al experto / **papá/mamá/tutor legal** para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras a la máquina para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de pedir un crédito al banco / **recibir paga?**

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------



**¿En quién confiarías más para que aprobase los permisos de salida de los internos de una cárcel? / ¿En quién confiarías más para decidir cuándo quitar un castigo por una acción indebida?**

- En la junta de tratamiento de prisión / En la misma persona que me ha puesto el castigo
- En un algoritmo que evalúe si se cumplen los requisitos por parte de los internos / En una máquina que evalúe cuándo quitar el castigo

¿Cómo de fiable consideras a la junta de tratamiento / persona que te ha puesto el castigo para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al algoritmo / máquina para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de esperar la resolución de un permiso penitenciario / esperar a que te quiten un castigo?

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------

**¿En quién confiarías más para seleccionar a los ganadores de los premios Oscar a la mejor película?**

- En la votación de los expertos de la industria del cine
- En las recomendaciones de un sistema de análisis y evaluación automatizado

¿Cómo de fiable consideras a los expertos para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al sistema automatizado para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de ver/asistir a una entrega de premios?

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------

¿En quién confiarías más para presentarte a alguien si estuvieses interesado en encontrar una pareja? / ¿En quién confiarías más para presentarte a alguien si estuvieses interesado en hacer un amigo nuevo?

- En un experto en relaciones humanas
- En una aplicación digital de citas / En una aplicación digital de tu Tablet o móvil

¿Cómo de fiable consideras al experto para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras a la aplicación para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de estar interesado en encontrar pareja / conocer un amigo nuevo?

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------

¿En quién confiarías más para que decidiese a quién incorporar como nuevo miembro a tu equipo de trabajo? / ¿En quién confiarías más para que decidiese a quién incorporar como nuevo miembro a tu equipo de deporte o a un juego?

- En un profesional experto / En un entrenador experto
- En un algoritmo de recomendación

¿Cómo de fiable consideras al profesional / entrenador para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras al algoritmo para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de formar parte de un equipo de trabajo / **participar en un deporte/juego de equipo?**

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------

**¿En quién confiarías más para que analizase tus pruebas médicas y emitiese un posible diagnóstico de enfermedad? / ¿En quién confiarías más para que analizase tus pruebas médicas y te recetase las medicinas que necesitas cuando estás malo/a?**

- En un médico experto
- En una máquina especializada

¿Cómo de fiable consideras al médico para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras a la máquina para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de esperar unos resultados/diagnóstico médico / **esperar unos resultados médicos o que te receten medicinas?**

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------

**¿En quién confiarías más para pedir ayuda si tuvieses un problema psicológico? / ¿En quién confiarías más para pedir ayuda si tuvieses un problema o necesitas ayuda personal?**

- En un psicólogo con años de experiencia / **en papá/mamá/un profesor**
- En una aplicación digital de salud mental

¿Cómo de fiable consideras al psicólogo / **papá/mamá/un profesor** para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo de fiable consideras a la aplicación para resolver esta tarea?

Nada fiable	Poco fiable	Algo fiable	Bastante fiable	Completamente fiable
-------------	-------------	-------------	-----------------	----------------------

¿Cómo calificarías el nivel de riesgo asociado con fallar en esta tarea? (siendo 1=Ninguno o muy bajo riesgo y 10= Muy alto riesgo).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

¿Cómo de familiarizado estás con la situación de solicitar ayuda psicológica / **solicitar ayuda personal?**

Nada familiarizado	Poco familiarizado	Algo familiarizado	Bastante familiarizado	Muy familiarizado
--------------------	--------------------	--------------------	------------------------	-------------------

**Apéndice 2.** Tabla comparativa ítems versión adultos y adaptación menores

<b>Área de evaluación</b>	<b>Versión adultos</b>	<b>Adaptación menores</b>
<b>Ciencias (s_steam)</b>	¿En quién confiarías más para investigar en la elaboración de una vacuna?	<i>Ídem</i>
<b>Tecnología (t_steam)</b>	¿En quién confiarías más para resolver un problema técnico en tu móvil u ordenador?	<i>Ídem</i>
<b>Ingeniería (e_steam)</b>	¿En qué vehículo confiarías más para comprarlo?	¿En qué vehículo confiarías más para montarte en él?
<b>Arte (a_steam)</b>	¿En quién confiarías más para que diseñara y pintara un cuadro para decorar tu vivienda?	¿En quién confiarías más para que diseñara y pintara un cuadro para decorar tu habitación/clase?
<b>Matemáticas (m_steam)</b>	Resolver un problema matemático	<i>Ídem</i>
<b>Académico/laboral</b>	¿En quién confiarías más que valorase tu solicitud/candidatura a un puesto de trabajo?	¿En quién confiarías más para que te corrigiese un examen?
<b>Económica</b>	¿En quién confiarías más para que resolviese tu solicitud de un préstamo bancario?	¿En quién confiarías más para que decidiese la paga que te dan?
<b>Social 1</b>	¿En quién confiarías más para que aprobase los permisos de salida de los internos de una cárcel?	¿En quién confiarías más para decidir cuándo quitar un castigo por una acción indebida?
<b>Social 2</b>	¿En quién confiarías más para seleccionar a los ganadores de los premios Oscar a la mejor película?	<i>Ídem</i>
<b>Afectivo relacional 1</b>	¿En quién confiarías más para presentarte a alguien si estuvieses interesado en encontrar una pareja?	¿En quién confiarías más para presentarte a alguien si estuvieses interesado en hacer un amigo nuevo?
<b>Afectivo relacional 2</b>	¿En quién confiarías más para que decidiese a quién incorporar como nuevo miembro a tu equipo de trabajo?	¿En quién confiarías más para que decidiese a quién incorporar como nuevo miembro a tu equipo de deporte o a un juego?

**Salud 1**

¿En quién confiarías más para que analizase tus pruebas médicas y emitiese un posible diagnóstico de enfermedad?

¿En quién confiarías más para que analizase tus pruebas médicas y te recetase las medicinas que necesitas cuando estás malo/a?

**Salud 2**

¿En quién confiarías más para pedir ayuda si tuvieses un problema psicológico?

¿En quién confiarías más para pedir ayuda si tuvieses un problema o necesitas ayuda personal?