



UNIVERSIDAD  
CATÓLICA  
DE CUENCA

**UNIVERSIDAD CATÓLICA DE CUENCA**

*Comunidad Educativa al Servicio del Pueblo*

**UNIDAD ACADÉMICA POSGRADO**

**Maestría en Psicología Clínica con Mención en Psicoterapia**

**Detección del Engaño en el Discurso Mediante el Monitoreo  
de la Realidad y Chat GPT-4 en una Muestra General  
Española**

**INFORME DE INVESTIGACIÓN PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE MAGISTER EN PSICOLOGÍA CLÍNICA CON  
MENCIÓN EN PSICOTERAPIA**

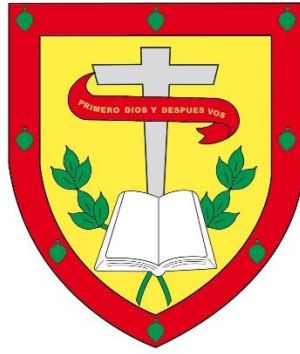
**AUTOR: LCDO. DANIEL IVÁN LITUMA ANDRADE**

**DIRECTOR: DR. ANDRÉS ALEXIS RAMÍREZ CORONEL, PHD.**

**CUENCA - ECUADOR**

**2025**

**DIOS, PATRIA, CULTURA Y DESARROLLO**



**UNIVERSIDAD CATÓLICA DE CUENCA**

*Comunidad Educativa al Servicio del Pueblo*

**UNIDAD ACADÉMICA DE POSGRADO**

**Maestría en Psicología Clínica con Mención en Psicoterapia**

DETECCIÓN DEL ENGAÑO EN EL DISCURSO MEDIANTE EL  
MONITOREO DE LA REALIDAD Y CHAT GPT-4 EN UNA  
MUESTRA GENERAL ESPAÑOLA

**INFORME DE INVESTIGACIÓN PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE MAGISTER EN PSICOLOGÍA CLÍNICA CON  
MENCIÓN EN PSICOTERAPIA**

**AUTOR:** LCDO. DANIEL IVÁN LITUMA ANDRADE

**DIRECTOR:** DR. ANDRÉS ALEXIS RAMÍREZ CORONEL, PHD.

CUENCA – ECUADOR

2025

**DIOS, PATRIA, CULTURA Y DESARROLLO**

## **Certificado del Asesor**

Se certifica que:

El informe de investigación Detección del Engaño en el Discurso Mediante el Monitoreo de la Realidad y Chat GPT-4 en una Muestra General Española, de autoría del Sr. Daniel Iván Lituma Andrade con número de identidad 1104506603, con nacionalidad ecuatoriana, previo a la obtención del Título de Cuarto Nivel o Posgrado correspondiente a Magister en Psicología Clínica con Mención en Psicoterapia, cumple con la caracterización y estructura (parte protocolaria y parte expositiva) y se sujeta a la normativa pertinente exigida por el Consejo de Educación Superior, CES y la Universidad Católica de Cuenca, en consecuencia se autoriza su presentación para los trámites pertinentes.

Santa Ana de los Ríos de Cuenca

Febrero, 2025

---

Dr. Andrés Alexis Ramírez Coronel, PhD.

Asesor Metodológico y Científico

## **Declaratoria de Autoría y Responsabilidad**

**Daniel Iván Lituma Andrade** portador de la cédula de ciudadanía N° **1104506603**. Declaro ser el autor de la obra: **Detección del Engaño en el Discurso Mediante el Monitoreo de la Realidad y Chat GPT-4 en una Muestra General Española**, sobre la cual me hago responsable sobre las opiniones, versiones e ideas expresadas. Declaro que la misma ha sido elaborada respetando los derechos de propiedad intelectual de terceros y eximo a la Universidad Católica de Cuenca sobre cualquier reclamación que pudiera existir al respecto. Declaro finalmente que mi obra ha sido realizada cumpliendo con todos los requisitos legales, éticos y bioéticos de investigación, que la misma no incumple con la normativa nacional e internacional en el área específica de investigación, sobre la que también me responsabilizo y eximo a la Universidad Católica de Cuenca de toda reclamación al respecto.

Cuenca, febrero 2025

F: .....

**Lic. Daniel Iván Lituma Andrade**

**C.I. 1104506603**

## **Agradecimiento**

Agradezco a Dios por haberme guiado en este camino y por darme la oportunidad de culminar esta importante etapa de posgrado.

A la Universidad Católica de Cuenca y a la Unidad Académica de Posgrado y la Maestría de Psicología Clínica con Mención en Psicoterapia, por brindarme una formación integral basada en el conocimiento, la ética y el compromiso profesional.

De manera especial, expreso mi más sincero agradecimiento a la Dra. Sofía Faggioni, Mgt., por su invaluable guía y apoyo incondicional a lo largo de este proceso, así como a la Dra. Jéssica Jaramillo, Mgt., cuya experiencia y dedicación han sido fundamentales en nuestra formación como psicoterapeutas.

Finalmente, extiendo mi gratitud a todos los docentes y colegas que contribuyeron a mi crecimiento académico y profesional, ayudándome a consolidar mi preparación y a desarrollar las competencias necesarias para ejercer esta hermosa profesión con responsabilidad y vocación

## **Dedicatoria**

A mis padres, pilares fundamentales de mi vida, cuyo amor, guía y apoyo incondicional han sido la fuerza que me ha impulsado a superarme cada día. Su confianza en mí ha sido el motor que me permitió alcanzar este logro.

A mi hermana, quien con su fortaleza y compañía me ha inspirado a seguir adelante, recordándome siempre que incluso en los momentos más difíciles, hay razones para continuar.

A mis abuelitos, cuyo amor y sabiduría permanecen en mi corazón como un tesoro invaluable. Aunque ahora son mis ángeles, sé que siguen guiando mi camino con su luz.

Finalmente, a mis amigos, con quienes forjé un lazo inquebrantable a lo largo de estos años. Su apoyo, compañía y ejemplo me han ayudado a crecer no solo profesionalmente, sino también como ser humano.

A todos ustedes, mi más profundo agradecimiento.

## Resumen

La detección del engaño es un tema clave en la Psicología Clínica, Forense y Jurídica, donde la veracidad de los testimonios es esencial para el desarrollo adecuado de los procesos. Tradicionalmente, se han utilizado sistemas como Análisis de Contenidos Basado en Criterios y Análisis Científico del Contenido de Declaraciones, pero el Monitoreo de la Realidad ha demostrado ser un método fiable para distinguir el engaño. Este estudio analiza la efectividad del sistema de Monitoreo de la Realidad combinado con la herramienta de inteligencia artificial *CHAT GPT-4* para la detección del engaño en una muestra española. Se llevó a cabo una investigación cuantitativa no experimental de tipo descriptivo correlacional y transversal. La muestra estuvo conformada por 50 participantes españoles, quienes elaboraron relatos reales e inventados bajo diversas condiciones experimentales. Se empleó un análisis de consistencia interna, para evaluar la fiabilidad de las respuestas según los criterios; un análisis descriptivo de medidas que incluyó la media, mediana, mínimo, máximo, desviación estándar y cuartiles, además se utilizaron pruebas estadísticas como el análisis de varianza (ANOVA) para evaluar diferencias significativas en los resultados del sistema RM en función de las diferentes condiciones experimentales. Los resultados indican que el sistema RM combinado con *CHAT-GPT4* es eficaz para identificar el engaño, especialmente en historias completamente inventadas o reales. No obstante, se observaron limitaciones en la variabilidad de los resultados dependiendo de la complejidad de las narrativas y de las condiciones experimentales. La investigación concluye que la integración de la inteligencia artificial en el análisis de testimonios no solo mejora la precisión en la detección del engaño, sino que también abre nuevas oportunidades para su aplicación en ámbitos clínicos y forenses, donde la veracidad es crucial para la toma de decisiones.

**Palabras clave:** *Detección engaño, Evaluación psicológica, Reality Monitoring, Sistemas embebidos.*

## **Abstract**

Detection of deception is a key issue in Clinical, Forensic and Legal Psychology, where the veracity of testimonies is essential for the proper development of the processes. Traditionally, systems such as Criterion-Based Content Analysis and Scientific Content Analysis of Statements have been used, but Reality Monitoring has proven to be a reliable method to distinguish deception. This study analyzes the effectiveness of the Reality Monitoring system combined with the artificial intelligence tool CHAT GPT-4 for deception detection in a Spanish sample. A descriptive correlational and cross-sectional non-experimental quantitative research was carried out. The sample consisted of 50 Spanish participants, who elaborated real and invented stories under different experimental conditions. An internal consistency analysis was used to evaluate the reliability of the responses according to the criteria; a descriptive analysis of measures including mean, median, minimum, maximum, standard deviation and quartiles; and statistical tests such as analysis of variance (ANOVA) were used to evaluate significant differences in the results of the MRI system according to the different experimental conditions. The results indicate that the MRI system combined with CHAT-GPT4 is effective in identifying deception, especially in completely fabricated or real stories. However, limitations were observed in the variability of the results depending on the complexity of the narratives and experimental conditions. The research concludes that the integration of artificial intelligence in the analysis of testimonies not only improves the accuracy of deception detection but also opens new opportunities for its application in clinical and forensic settings, where veracity is crucial for decision making.

**Keywords:** *Deception Detection, Psychological Assessment, Reality Monitoring, Embedded Systems.*

## Índice de Contenido

<b>Capítulo I: El Problema</b> .....	10
<b>1.1 Situación Problemática</b> .....	11
<b>1.2 Problema Científico</b> .....	13
<b>1.3 Línea de Investigación</b> .....	13
<b>1.4 Objeto del proyecto de investigación NEXUS</b> .....	14
<b>1.5 Campo de Acción</b> .....	14
<b>1.6 Objetivo general</b> .....	15
<b>1.7 Objetivos Específicos (OE)</b> .....	16
<b>1.8 Hipótesis Específicas</b> .....	16
<b>1.9 Variables</b> .....	17
<b>1.10 Justificación de la Investigación</b> .....	17
<b>Capítulo II: Marco Teórico</b> .....	19
<b>2.1 Fundamentación Teórica</b> .....	19
<b>Capítulo III: Metodología</b> .....	29
<b>3.1 Tipo de Investigación</b> .....	29
<b>3.2 Métodos Técnicas e Instrumentos de Investigación</b> .....	31
<b>3.3 Universo de Estudio y Tratamiento Muestral</b> .....	31
<b>3.4 Tratamiento Estadístico y Representación Gráfica</b> .....	32
<b>Capítulo IV: La Propuesta</b> .....	49
<b>4.1 Antecedentes de la propuesta</b> .....	49
<b>4.2 Introducción</b> .....	50
<b>4.3 Objetivo</b> .....	53
<b>4.4 Justificación</b> .....	53
<b>4.5 Alcance</b> .....	54
<b>Conclusiones, limitaciones y prospectiva</b> .....	56
<b>Recomendaciones</b> .....	58

## 1. CAPÍTULO I: EL PROBLEMA

El engaño es una característica inherente a la naturaleza humana y ha estado presente en diversas culturas desde tiempos inmemoriales. Según Soto et al. (2013), a lo largo de la historia y, particularmente en el ámbito de la Psicología forense, la capacidad de descubrir la información que una persona oculta ha sido un tema de gran interés en la investigación. Este interés ha impulsado el desarrollo de múltiples técnicas y herramientas diseñadas para detectar el engaño, con el propósito de administrar justicia y mantener el orden social (Masip et al., 2005).

Existen múltiples métodos para la detección del engaño, como aquellos que incluyen procedimientos basados en la medición, registro y análisis de la actividad psicofisiológica; los centrados en los correlatos no verbales y paraverbales del engaño; y los que examinan el análisis del contenido verbal del discurso (Bogaard et al., 2013). Aunque estos métodos han demostrado ser útiles y populares, a menudo resultan confusos o difíciles de aplicar. Por lo tanto, es imprescindible seguir investigando y perfeccionando estas técnicas para incrementar su precisión y confiabilidad (Sporer y Sharman, 2006).

A pesar de los avances históricos y actuales, la detección del engaño sigue siendo un desafío. Los seres humanos han evolucionado para ser hábiles tanto en engañar como en detectar el engaño, dado que estas capacidades han sido esenciales para la supervivencia y la interacción social a lo largo del tiempo. Esta dinámica creó una "carrera armamentista" entre los engañadores y quienes intentan desenmascararlos, lo que ha impulsado el desarrollo de técnicas y tecnologías cada vez más sofisticadas.

En años recientes, los avances tecnológicos han abierto nuevas posibilidades para la detección del engaño, tales como el análisis computarizado de texto, el análisis de voz y la tecnología de reconocimiento facial. Si bien estas herramientas ofrecen el potencial

de mejorar la objetividad y precisión en la detección del engaño, también presentan desafíos y limitaciones que deben ser superadas para lograr su efectividad total.

La búsqueda por detectar el engaño refleja la fascinación continua de la humanidad por comprender el comportamiento y las complejidades de la comunicación interpersonal. Aunque ningún método es completamente infalible, la investigación y la innovación constantes en este campo prometen mejorar la capacidad de discernir la verdad de la mentira en contextos tan diversos como investigaciones criminales o las interacciones cotidianas.

### 1.1. **Situación Problemática**

El engaño se ha convertido en una limitante significativa en los procesos de evaluación psicológica. Según Vrij (2008), las personas suelen recurrir al engaño durante la ejecución de pruebas, lo que resalta la necesidad de desarrollar modelos eficaces capaces de detectarlo de manera precisa y predictiva.

Dentro de los modelos más relevantes para el análisis del discurso y la detección del engaño, se destacan tres herramientas fundamentales. En primer lugar, el *Criteria-Based Content Analysis (CBCA)*, o Análisis de Contenidos Basado en Criterios, es una técnica forense que busca diferenciar entre declaraciones veraces y falsas, particularmente en casos que involucran a víctimas de delitos (Rassin, 2000). En segundo lugar, el *Scientific Content Analysis (SCAN)*, o Análisis Científico del Contenido de Declaraciones, examina los patrones del habla e identifica señales de engaño mediante el análisis de la estructura y el contenido de las declaraciones (Masip et al., 2002). Finalmente, el *Reality Monitoring (RM)*, o Monitoreo de la Realidad, se utiliza para distinguir entre recuerdos genuinos, adquiridos a través de la percepción de la realidad externa, y recuerdos fabricados internamente mediante la imaginación (Colman, 2009).

En la actualidad, la inteligencia artificial (IA) está revolucionando diversas áreas, incluida la detección del engaño. La IA se presenta como una herramienta poderosa para identificar y prevenir conductas ilícitas, ya que permite realizar análisis de datos avanzados, esenciales para desenmascarar fraudes y corrupción. Con el tiempo, la IA ha transformado la forma en que se abordan estos problemas, proporcionando capacidades analíticas y predictivas que superan a las herramientas tradicionales. A través de algoritmos sofisticados y técnicas de aprendizaje automático, esta tecnología puede procesar grandes volúmenes de datos, detectando patrones y comportamientos sospechosos que podrían indicar engaño.

El Chat *GPT-4*, una herramienta basada en IA desarrollada por *OpenAI*, representa un ejemplo notable del potencial futuro en la detección de patrones lingüísticos y comportamentales asociados con el engaño. No obstante, es importante señalar que aún existen limitaciones que impiden su plena implementación, por lo tanto, la IA debe considerarse como una herramienta complementaria, no como una solución definitiva en sí misma.

La meta a largo plazo es integrar la IA en diversos campos profesionales con el fin de fortalecer la capacidad de identificar y combatir conductas ilícitas. Esta integración no solo beneficiaría a áreas como la Psicología forense y jurídica, ayudando en la verificación de testimonios y prevención de fraudes, sino que también podría mejorar la efectividad de las intervenciones psicológicas.

En este contexto, la presente investigación se centra en el sistema RM y el Chat *GPT-4* de *OpenAI*, con el objetivo de analizar el contenido del discurso para detectar engaños en una muestra general española del Proyecto Nexus. Al combinar el sistema RM con las capacidades de *Chat GPT*, se busca profundizar en cómo la inteligencia artificial puede contribuir a mejorar la evaluación de la credibilidad en los testimonios.

## 1.2. Problema Científico

Con el paso del tiempo, especialmente en el ámbito jurídico y de la Psicología forense, la detección del engaño ha adquirido una importancia crucial. Esto se debe a que los testimonios suelen constituir la principal fuente de evidencia en la investigación de un caso, de forma que, la presencia de un relato falso puede comprometer el desarrollo legal y poner en riesgo los resultados de procesos judiciales. Ante esta necesidad, a finales del siglo XX comenzaron a surgir investigaciones desde la perspectiva psicológica, con el objetivo de distinguir entre declaraciones veraces y engañosas.

Existen diversas herramientas encaminadas a la detección del engaño, como los sistemas CBCA, SCAN y RM, que han generado criterios útiles para evaluar la credibilidad de un testimonio. Sin embargo, en la actualidad, la llegada de las IA y el desarrollo de programas como *Chat GPT* representan una auténtica revolución tecnológica, las cuales podrían mejorar y optimizar las herramientas existentes para la detección del engaño, abriendo nuevas posibilidades en la evaluación de testimonios.

El desafío consiste en integrar ambas herramientas de manera que contribuyan eficazmente a los campos de la Psicología forense y jurídica. Para maximizar sus beneficios y minimizar los riesgos, es fundamental abordar los desafíos legales, éticos y prácticos que su implementación implica. De esta manera, se podrá aprovechar el potencial de la IA para identificar con mayor precisión la veracidad de los testimonios, promoviendo la justicia y fortaleciendo la integridad en nuestras instituciones y sociedades.

## 1.3. Línea de Investigación

*Dominio: 61 Psicología.*

**Línea de Investigación 12:** Salud y Bienestar por ciclo de vida.

**Sublínea 6:** Comportamiento en salud individual, familiar y comunitaria

### ***Ámbito de investigación***

Proyecto Nexus, aprobado por la jefatura de investigación con el código PICCIITT19-17 Cuenca, agosto de 2020. Versión 2.0.

#### **1.4. Objeto del proyecto de investigación NEXUS**

Detección del engaño mediante el estudio del contenido del discurso utilizando criterios del sistema de evaluación inteligente *Reality Monitoring (RM)*.

#### **1.5. Campo de Acción**

El proyecto de investigación *NEXUS* tiene como finalidad diseñar, implementar y validar un sistema avanzado para la detección del engaño, utilizando el análisis de contenido del discurso basado en los criterios del sistema RM. Este sistema busca diferenciar narrativas veraces de aquellas fabricadas, mediante la evaluación de aspectos psicológicos y lingüísticos específicos del discurso.

El alcance del proyecto abarca tanto la investigación como el desarrollo tecnológico apoyado en inteligencia artificial y lingüística. Se realizarán análisis comparativos de los resultados obtenidos en distintas condiciones experimentales para medir la eficacia del sistema. Además, se evaluará la precisión del mismo en la detección de engaños dentro de contextos controlados, con el objetivo de identificar los parámetros que mejor predicen la veracidad de los testimonios.

Cabe destacar que se excluyen del proyecto el uso del sistema en investigaciones criminales reales sin la previa validación ética y legal, así como el desarrollo de aplicaciones comerciales directas sin estudios complementarios que aseguren su fiabilidad. Esto es esencial para garantizar que el sistema se aplique de manera segura y ética en futuras investigaciones o implementaciones.

Entre los entregables del proyecto se incluye un informe sobre la fiabilidad de los criterios utilizados en el sistema RM, un análisis comparativo de los resultados obtenidos en diferentes condiciones experimentales y una documentación detallada de la precisión del sistema RM combinado con Chat GPT-4. Estos entregables servirán como base para próximas investigaciones y aplicaciones del sistema en diferentes ámbitos.

El proyecto asume que se contará con una muestra representativa de la población general española para la recopilación de datos, así como con una colaboración efectiva entre lingüistas, psicólogos y desarrolladores de inteligencia artificial. Estas colaboraciones interdisciplinarias son esenciales para el éxito del sistema (Altamirano y Bernuy, 2022).

No obstante, el proyecto presenta ciertas limitaciones como la variabilidad lingüística y cultural dentro de la muestra, lo cual podría afectar la consistencia de los resultados. Además, existen restricciones técnicas y éticas relacionadas con el uso de inteligencia artificial para el análisis de discurso, lo que puede dificultar ciertas aplicaciones del sistema (Gordon y Turnbull, 2024).

Para que el proyecto sea considerado exitoso, es necesario alcanzar una precisión mínima acordada previamente en la detección del engaño. Así mismo, el cumplimiento de los objetivos específicos debe estar respaldado por evidencia científica y tecnológica sólida. Finalmente, los resultados obtenidos deben ser validados en escenarios controlados antes de ser aplicados en contextos prácticos, garantizando la efectividad y fiabilidad del sistema.

#### 1.6. **Objetivo general**

Analizar la efectividad del sistema de Monitoreo de la Realidad y la herramienta de inteligencia artificial *Chat GPT-4* en la detección del engaño a través del análisis de

contenido del discurso, utilizando una muestra general de participantes españoles, con el fin de mejorar la precisión en la evaluación de la credibilidad de los testimonios.

### 1.7. **Objetivos Específicos (OE)**

**OE1.** Identificar la confiabilidad de las respuestas obtenidas de la muestra española del proyecto NEXUS utilizando los criterios de Monitoreo de la Realidad y el Chat GPT-4 OPEN AI para la detección del engaño.

**OE2.** Comparar los resultados obtenidos de la muestra española del proyecto NEXUS usando chat GPT-4 OPEN AI considerando las diferentes condiciones experimentales y respuestas obtenidas.

**OE3.** Evaluar el nivel de precisión de los criterios del Monitoreo de la Realidad para la detección del engaño al ser aplicados utilizando el Chat *Gtp-4* de *OPEN AI*, de muestra general española del proyecto NEXUS.

### 1.8. **Hipótesis Específicas**

**H1.** Si el uso de los criterios del sistema de Monitoreo de la Realidad aplicados mediante *Chat GPT-4 OPEN AI* en la data de la muestra general española, permitirá identificar de manera fiable la detección del engaño, dependiendo de las condiciones experimentales.

**H2.** Si existen diferencias significativas en los resultados de los discursos analizados mediante el sistema de Monitoreo de la Realidad entre las distintas condiciones experimentales.

**H3.** Si el uso de los criterios del Monitoreo de la Realidad, en combinación con el modelo de Chat GPT-4 de OPEN AI, permitirá detectar con un alto nivel de precisión posibles elementos de engaño de la muestra general española del proyecto NEXUS.

## 1.9. Variables

**Tabla 1**

*Variables*

<i>Independiente</i>	<i>Dependiente</i>
El engaño	Criterios del Monitoreo de la Realidad

## 1.10. Justificación de la Investigación

En un mundo caracterizado por la proliferación de desinformación y noticias falsas, estudios como el presente resultan imprescindibles para integrar la inteligencia artificial en sistemas que permitan la detección del engaño. Tal como destaca Nahari (2018), es fundamental que los profesionales en Psicología clínica, forense y jurídica desarrollen herramientas que faciliten la identificación del engaño, lo que contribuiría significativamente a la prevención de la injusticia y el fraude, evitando su continuo aumento.

Estas herramientas, además, pueden desempeñar un papel crucial en la protección de la verdad y la integridad. Por ejemplo, podrían emplearse para mejorar los mecanismos subyacentes en los trastornos mentales, esto abriría la puerta a nuevas estrategias de intervención y tratamiento. Asimismo, serían útiles para evaluar la credibilidad de testimonios, permitiendo diferenciar entre recuerdos auténticos y fabricados en procesos jurídicos y forenses. Este enfoque, a su vez, posibilitaría el desarrollo de protocolos de entrevista más efectivos, que minimicen la sugestión y maximicen la calidad de la información obtenida (Huacón y Martínez 2021).

Teniendo en cuenta lo expuesto, Soto et al. (2013) mencionan que, aunque existe una amplia cantidad de investigaciones realizadas sobre el sistema de CBCA para la detección del engaño, no debe subestimarse el potencial del sistema RM. Este sistema se utiliza para analizar los recuerdos basados en sus cualidades de contenido, prestando especial atención a los detalles perceptuales, espaciales y temporales. Por lo tanto, la integración de la IA en conjunto de los criterios del RM representaría una innovación clave para mejorar la precisión en la distinción entre una declaración engañosa y una veraz.

Uno de los mayores desafíos para los profesionales en estos ámbitos es percibir la detección del engaño de manera objetiva, sin ser influenciados por prejuicios propios o ajenos. Como afirman De Paulo et al. (2003), es fundamental resaltar que nuestros juicios a menudo se basan en recuerdos, los cuales pueden ser defectuosos debido a la fragilidad de la memoria en el momento de realizar el análisis de un testimonio. Por ello, una de las habilidades cognitivas más importantes es la capacidad de discernir con precisión el origen de la información.

En este sentido, la integración de la IA con los criterios del sistema RM podría mejorar significativamente la precisión en la detección del engaño, dado que una de sus principales ventajas es su capacidad para analizar grandes volúmenes de información de manera rápida y precisa. Además, la IA tiene la habilidad de aprender velozmente, adaptarse y evolucionar en función de lo que ha aprendido (Masip et al., 2005).

## 2. CAPÍTULO II: MARCO TEÓRICO

### 2.1 Fundamentación Teórica

El engaño puede describirse como una conducta deliberada cuyo objetivo es hacer que otras personas crean que una historia contada es verdadera, aun cuando se sabe perfectamente que los detalles son falsos (Carvajal y Barreto, 2022). Este fenómeno puede involucrar aspectos como la inclusión, exclusión o modificación de la información. En este sentido, la Clasificación Internacional de Enfermedades (CIE-10) lo define como el acto de persuadir intencionalmente a alguien sobre algo falso, lo que generalmente se emplea como una estrategia para resolver conflictos (Alazrai et al., 2018). De manera similar, Masip et al. (2016) entienden al engaño como un esfuerzo intencional de ocultar o manipular información, ya sea de manera verbal o no verbal, con el fin de inducir o preservar en otros una creencia que el comunicador conoce que es falsa.

Desde la antigüedad, los seres humanos han buscado procedimientos para detectar mentiras utilizando diversas técnicas. De acuerdo con Feijoo et al. (2018), con el avance de la civilización y el desarrollo científico, han surgido nuevos métodos y herramientas más sofisticadas que permiten un análisis más preciso del fenómeno. Sin embargo, a pesar de estos avances, el engaño sigue siendo una práctica común, de hecho, se considera que la mayoría de las personas engañamos por lo menos una vez en el día, ya sea para calmar o respaldar a otros, influir en las personas o causar una buena impresión (Álvares y Yusti, 2015).

El acto de engañar conlleva una serie de reacciones fisiológicas y comportamentales en la persona que lo realiza, como el aumento del ritmo cardíaco o de la sudoración (Vrij et al., 2004). Estas reacciones se deben a la activación del sistema nervioso autónomo, que responde al estrés o a la necesidad de mantener la coherencia entre lo dicho y lo percibido por los demás. Además, el engaño se manifiesta de múltiples formas, tales como

la falsificación (mentiras), la equivocación (evasión, ambigüedad), la omisión (ocultar información importante), la exageración y la subestimación (Walczyk et al., 2014). Estas estrategias se utilizan con diversos fines, como evitar conflictos, obtener beneficios personales o influir en la percepción de los demás. Así mismo, según Álvarez y Yusti (2015), la habilidad para engañar es considerada valiosa en áreas como la política, las ventas y, especialmente, en el ámbito forense, donde se emplea tanto en la investigación como en el enjuiciamiento de delitos, e incluso con personas involucradas en procesos judiciales. Esto subraya la importancia de entender las bases psicológicas y fisiológicas del engaño para interpretar su papel en contextos complejos.

En este contexto, este fenómeno constituye una limitante en los procesos de evaluación psicológica, ya que, como afirma Vrij (2008), las personas tienden a recurrir a cualquier tipo de engaño durante la ejecución de pruebas. Esto resalta la necesidad de analizar y desarrollar un modelo eficaz que permita identificarlo con precisión y carácter predictivo (Masip, 2017). Desde la perspectiva psicológica, el engaño se ha convertido en tema de estudio clave, abarcando desde la simulación de síntomas de conveniencia situacional hasta aspectos patológicos (Ekman, 2021). Estos síntomas se entienden como comportamientos que siguen un proceso de aprendizaje, lo que facilita su adquisición, mantenimiento o supresión (Díaz, 2014).

La credibilidad, por su parte, desempeña un papel fundamental en la vida diaria y en el ámbito profesional (Martin y Leach, 2013). El engaño puede ocurrir en una variedad de contextos, incluidos los clínicos y forenses. De acuerdo con Blandón et al. (2017) las limitaciones para su detección dificultan la recolección de información durante los procesos psicológicos, lo cual puede resultar en un mal manejo de terapias o casos judiciales, afectando a pacientes o personas inocentes.

Además, en el área psicológica, evaluar la veracidad de la información suministrada por los individuos es crucial, ya que puede impactar directamente tanto en el diagnóstico como en el tratamiento (Huacón y Martínez, 2021). De hecho, las pruebas de evaluación basadas en autoinforme, comúnmente empleadas en contextos clínicos, son particularmente susceptibles de manipulación, bien sea por simulación o exageración, debido a factores como la deseabilidad social y las tendencias en las respuestas.

Por su parte, Carvajal et al. (2021) definen a la detección del engaño como el proceso de determinar si una comunicación contiene información verdadera o falsa. Algunas formas de engaño generan cambios conductuales y fisiológicos, tanto verbales como no verbales que ofrecen indicadores más evidentes para su detección (Constancio et al., 2023). Para este fin, métodos como el polígrafo, el evaluador de estrés psicológico, las huellas digitales cerebrales y la electroencefalografía (EEG) han sido ampliamente aplicados en la actualidad (Bowman et al., 2014).

Ante lo expuesto, Masip et al. (2015) destacan que la persona promedio tiene una tasa de detección de mentiras de aproximadamente 54%, que en raras ocasiones supera el 60%. Sin embargo, algunos individuos muestran una notable capacidad para identificar engaños, con una precisión superior al 90%. Décadas de investigación en el área han demostrado que herramientas como la observación de la conducta o la atención al discurso son insuficientes para garantizar una detección precisa del engaño. De hecho, algunos estudios demuestran que sólo el 54% de las verdades y mentiras se clasifican correctamente (Alvares y Yusti, 2015). Dado que el engaño es una práctica común, la detección del mismo se vuelve una tarea crucial para evitar que personas se aprovechen de otras.

Así mismo, es importante diferenciar entre el acto de engañar y estar desinformado. Una persona que no recuerda un evento o cuyos recuerdos han sido alterados por alguna razón, y por lo tanto proporciona un relato inexacto, no está intentando engañar, más bien estaría cometiendo un “error honesto”. Esto sucede porque, aunque la persona se esfuerce en ofrecer una versión precisa, puede no recordar ciertos detalles o considerar como ciertos datos incorrectos (Masip et al, 2004).

En un estudio llevado a cabo por Hartwing et al. (2007) se encontró que tanto las personas que engañan como las que dicen la verdad dedican un tiempo similar a planificar el contenido verbal de su discurso. Sin embargo, aquellos que engañaron afirmaron haber invertido más tiempo en preparar su comportamiento no verbal en comparación con quienes decían la verdad (Alvares y Yusti, 2015). Al engañar, una persona necesita crear una historia, y debe supervisar cuidadosamente su invención para que resulte lo más creíble posible y se ajuste a la información que el observador o los observadores conocen o podrían conocer.

Los individuos que mienten tienden a no asumir que su credibilidad será aceptada automáticamente, lo que los lleva a esforzarse más que las personas sinceras para parecer honestos ante el entrevistador o investigador (Fernández y Ullah, 2022). Además, al no dar por garantizada su credibilidad, también intentan controlar las reacciones de su interlocutor para evaluar si están siendo descubiertos, lo cual requiere un uso significativo de recursos cognitivos (Huacón y Martínez, 2021).

Así mismo, Vrij et al. (2010) señalan que las personas que engañan deben suprimir la verdad mientras fabrican su relato, lo que demanda un esfuerzo cognitivo considerable. A partir de este enfoque, surge una novedosa estrategia para la detección del engaño que consiste en aumentar la carga cognitiva del sospechoso. Es decir, la persona que miente

debe centrarse en varios aspectos, como controlar sus nervios, cuidar su lenguaje corporal, incluyendo posturas, voz, mirada, y estar atenta al comportamiento del interlocutor para ajustar su discurso y transmitir una apariencia de honestidad (Nortje y Tredoux, 2019). Todo este proceso debe realizarse de manera simultánea, lo que hace que mentir y/o engañar sea más complejo para la toma de decisiones, requiriendo mayor tiempo en su ejecución. Por lo tanto, se puede inferir que durante el engaño los tiempos de reacción tienden a ser más largos en comparación de cuando se dice la verdad (Abe et al, 2007; Markowitz et al., 2023).

Para detectar mentiras, es fundamental considerar las diferencias entre mentir y decir la verdad. Las operaciones mentales realizadas por una persona que miente para construir y relatar su historia son significativamente distintas de las elaboradas por una persona sincera (Blandón et al., 2017). En sus investigaciones, Vrij et al. (2010) señalaron que inventar una historia falsa es más complejo porque la persona que engaña debe crear un relato plausible que no contradiga lo que el receptor sabe o podría descubrir. Además, debe asegurarse de no omitir información importante, y ser capaz de memorizar la historia para repetirla coherentemente en el futuro.

La idea de que mentir es cognitivamente más exigente que decir la verdad, ha llevado a que muchos métodos de entrevista busquen aumentar la carga mental del entrevistado. El razonamiento detrás es que, al incrementar la dificultad cognitiva durante la entrevista, el mentiroso, que ya enfrenta un mayor esfuerzo mental, se verá más afectado que alguien que dice la verdad (Vrij et al., 2010). Uno de estos procedimientos es el Monitoreo de la realidad (RM) que Nahari (2018) describe como un proceso mediante el cual una persona evalúa sus propios recuerdos, basándose en las cualidades del contenido. Estos recuerdos se dividen en externos, caracterizados por atributos perceptivos y contextuales, e internos, definidos por atributos que ayudan en su generación.

De la misma manera, Martínez et al. (2021) describen al sistema RM como una función cognitiva clave para discriminar la fuente de estimulación, ya sea interna o externa. Esta herramienta podría ser útil para identificar marcadores cognitivos que actúen como predictores tempranos en el diagnóstico de ciertas patologías como la esquizofrenia. El RM es un proceso fundamental para distinguir entre la información generada internamente y la que proviene del exterior (Lavallé et al., 2023). Además, es indispensable para mantener una comprensión de uno mismo como un agente consciente interactuando con el mundo.

Es importante señalar también que los errores en el monitoreo de la realidad pueden provocar confusiones entre experiencia reales e imaginarias (Masip y Garrido, 2000). En ese sentido, Lemos et al. (2019) destacan que el sistema RM permite identificar pensamientos, sentimientos e imaginaciones en el discurso, ya que los relatos veraces, a diferencia de los falsos, suelen incluir más detalles perceptuales y contextuales.

En el RM, existen procesos involucrados en hacer atribuciones sobre el origen de la información que viene a la mente. Esto incluye la distinción entre la información generada por funciones cognitivas internas, como pensamiento e imaginación, y aquella derivada del mundo exterior a través de la percepción (Simons et al., 2017). Es más, las observaciones neuropsicológicas proporcionan evidencia sorprendente de cuan dependientes somos los seres humanos del buen funcionamiento de determinadas regiones del cerebro en el seguimiento de la realidad en la vida diaria (Johnson et al., 2000).

Se ha destacado que, generalmente las personas pueden mantener suficiente distinción entre el mundo interno y externo para funcionar en las circunstancias cotidianas, pero ocasionalmente pueden llegar a confundir experiencias reales con aquellas que les han

contado, imaginado o soñado (Tan et al., 2024). Es así como el sistema RM cuenta con criterios adaptados a la detección del engaño, enfocados en evaluar los detalles perceptuales y contextuales de los relatos (Nahari, 2018).

Así mismo, es necesario señalar la estrecha relación encontrada por Martínez et al. (2021) entre la monitorización de la realidad y el paradigma de la memoria. Debido a ello, el sistema RM dirige su atención a características como la sensorialidad, los detalles espaciales, y temporales, la afectividad (carga afectiva), la reconstrucción de la historia, el realismo y las operaciones cognitivas (Virj, 2008). La manera en que estos elementos se organizan en nuestros recuerdos determina su procedencia, de modo que los recuerdos externos presentan una mayor cantidad de información contextual, sensorial y semántica, a diferencia de los recuerdos internos, que son más ricos en operaciones cognitivas (Garrison et al., 2017).

El sistema RM está compuesto por dos pruebas para evaluar la credibilidad: el Cuestionario de Característica de la Memoria (MCQ) y el Cuestionario de Valoración de las Características de la Memoria (JMCQ) (Soto et al., 2013). El objetivo de estas herramientas es identificar los procesos mentales utilizados por las personas para decidir si la información (memoria) tiene un origen interno (imaginario o inventado) o externo (percibido) en sus discursos, discriminando entre información verdadera o falsa (Gancedo et al., 2021).

Para explorar su base teórica, Johnson y sus colaboradores desarrollaron el Cuestionario MCQ con 39 ítems que evalúan la claridad de un recuerdo según la distribución de atributos cualitativos (contextuales, sensoriales, semánticos o cognitivos) (Soto et al., 2013). Más adelante, con la finalidad de mejorar las investigaciones, Bogaard et al. (2013) crearon el Cuestionario JMCQ, compuesto por 8 ítems: claridad, información

perceptual, información espacial, información temporal, afectos, reconstructibilidad de la historia, realismo y operaciones cognitivas.

Este modelo de análisis según criterios ha sido probado en investigaciones que analizan los juicios de memoria en personas, comprobando la veracidad de sus testimonios. Se ha observado que el RM es tan válido como la CBCA para detectar el engaño, dado sus altos índices de precisión, lo que la convierte en una estrategia efectiva para distinguir entre declaraciones engañosas y veraces (Gancedo et al., 2021).

Ahora bien, en un mundo caracterizado por avances tecnológicos sin precedentes, la inteligencia artificial (IA) se ha consolidado como un motor clave de transformación y desarrollo. Su impacto es tan trascendental que puede compararse con la invención del vapor en la revolución industrial, un acontecimiento que cambió el curso de la historia (Cierco, 2024). Por su parte, Omil (2019) señala que la IA es una herramienta que otorga a un ordenador, una red de computadoras o un conjunto de robots controlados por estas, la capacidad de realizar tareas que normalmente se atribuyen a los seres humanos inteligentes. En definitiva, el objetivo final es que la tecnología sea capaz de realizar las mismas tareas que la mente humana (Boden, 2017).

La IA tiene la capacidad de analizar grandes volúmenes de datos en cuestión de segundos, superando así las habilidades humanas. Entre sus ventajas destacables se pueden señalar su accesibilidad y disponibilidad, su rapidez y eficacia, así como la optimización del tiempo (Romo et al., 2023). Así mismo, esta tecnología posee una gran adaptabilidad, lo que facilita su aplicación en tareas como la generación y procesamiento de información. Su destreza para manejar información implícita y generar conclusiones lógicas la convierte en una base ideal para desarrollar sistemas de inteligencia artificial más especializados (Ciprés et al., 2024).

De igual forma, Gómez y Galende (2024) manifiestan que OpenAI, una empresa estadounidense responsable de la creación de la familia de modelos de lenguaje extenso-transformador penetrado generativo (GPT-Generative pre trained transformer, por sus siglas en inglés), y del sistema de chatbot ChatGPT (Chat generative pre trained transformer), ha transformado el campo de la inteligencia artificial. Su tecnología ha sido adoptada por un número creciente de compañías, lo que ha impulsado su implementación a nivel mundial.

En este contexto, Guamán (2023) sostiene que ChatGPT es un modelo de lenguaje natural que emplea técnicas de inteligencia artificial para generar respuestas coherentes y pertinentes a preguntas o mensajes de texto en distintos idiomas. Además, utiliza el aprendizaje profundo, una técnica de aprendizaje automático para generar texto similar al lenguaje humano mediante redes neuronales transformadoras (Guzmán, 2023). De este modo, Chat-GPT fue entrenado inicialmente con texto disponible en línea para aprender el lenguaje humano, y posteriormente se perfeccionó con transcripciones para dominar los conceptos básicos de las conversaciones.

Por otro lado, una de las principales ventajas de Chat GPT según Radfor et al. (2019) es su capacidad para generar respuestas precisas en una gran diversidad de temas y contextos, así como su habilidad para aprender de forma continua y adaptarse a diferentes usuarios y situaciones. Además, presenta una notable escalabilidad y eficiencia en el procesamiento de lenguaje natural (PNL) lo que le permite ser empleado en áreas como la traducción de idiomas y el análisis de discursos y sentimientos con un alto grado de precisión. Esto es posible gracias a su arquitectura de red neuronal basada en transformadores, que facilita el procesamiento de datos a gran escala (Brown et al., 2020).

La IA ya no es una herramienta exclusiva para científicos de la computación, ingenieros o analistas de datos. Estos avances han tenido un impacto significativo en el campo de la Psicología, tanto en la investigación como en la atención clínica (Alqahtani et al., 2023). Gracias a esta tecnología, los psicólogos pueden explorar con mayor profundidad las complejidades del comportamiento humano y los procesos cognitivos, mediante el análisis de grandes volúmenes de datos, la extracción de información relevante y la creación de modelos predictivos (Gordon & Turnbull, 2024).

Este análisis veloz se lleva a cabo mediante algoritmos de aprendizaje automático que identifican patrones y correlaciones, lo que ha facilitado una comprensión más profunda del comportamiento humano y los procesos mentales (Salah et al., 2023). Por ejemplo, Gordon y Turnbull (2024) mencionan que los algoritmos desempeñan un papel crucial en el análisis de contenido, temas, aprendizaje automático y procesamiento del lenguaje natural. Esta capacidad ha permitido a los profesionales obtener una visión más detallada sobre el discurso y otras fuentes de datos basadas en textos.

Por otro lado, la IA ha adquirido gran relevancia en el ámbito de la Psicología forense, destacándose por su capacidad para reducir el margen de error humano. Gracias a los algoritmos y sistemas de aprendizaje automático, es posible analizar grandes volúmenes de datos, identificar comportamientos sospechosos y contribuir a la toma de decisiones estratégicas orientadas a la prevención de futuros delitos (Rodríguez, 2020). Así mismo, la IA desempeña un papel fundamental en la Psicología cognitiva. El desarrollo de modelos de IA que simulan la cognición humana está permitiendo a los investigadores profundizar en el estudio de procesos cognitivos como la toma de decisiones, la resolución de problemas y la memoria (Gordon & Turnbull, 2024).

No obstante, a pesar de los avances impulsados por la IA en diversos campos, su uso plantea importantes desafíos éticos y metodológicos que requieren atención. Según Iqbal et al. (2023), los investigadores en IA y salud argumentan que la responsabilidad por los errores que puedan surgir sigue siendo una cuestión ambigua, ya que no existe una entidad claramente definida que pueda ser señalada como responsable.

Es decir, las preocupaciones se centran en aspectos como la privacidad, la protección de datos, la transparencia de los algoritmos de la IA y su potencial sesgo. Además, los modelos de IA también enfrentan dificultades para interpretar el contexto social (Gordon & Turnbull, 2024). Por ello, como investigadores y profesionales, es nuestra responsabilidad reflexionar detenidamente sobre las implicaciones del uso de la IA en nuestro trabajo, y garantizar que estas herramientas se utilicen de manera ética y responsable.

Por último, si queremos emplear la IA en nuestra investigación y prácticas psicológicas, como es en este caso, para analizar el discurso en función de la detección del engaño, debemos estar al tanto de las últimas pautas éticas, mejores prácticas y avances tecnológicos (Salah et al., 2023).

### **3. CAPITULO III: METODOLOGÍA**

#### **3.1 Tipo de Investigación**

##### ***Diseño***

Se realizó una investigación con enfoque cuantitativo no experimental descriptivo, correlacional y transversal. Para determinar la efectividad del sistema RM combinado con la herramienta de inteligencia artificial *Chat GPT-4*, se incluyó un análisis comparativo entre dos grupos: un grupo experimental, al que se le aplicaron los criterios

del sistema RM con el apoyo de *Chat GPT-4* y un grupo de control, al que no se le aplicaron estos recursos

### ***Recolección de Datos***

La muestra estuvo compuesta por 50 participantes de la población general española, residentes del Principado de Asturias, con edades comprendidas entre los 18 y los 60 años. Los participantes fueron seleccionados mediante un muestreo por conveniencia, asegurando una distribución equilibrada en cuanto a género y edad. El grupo experimental y el grupo de control se asignaron aleatoriamente.

Los 50 participantes fueron instruidos para elaborar relatos bajo las siguientes condiciones: (1) Historia real de su pasado reciente; (2) Historia inventada de su pasado; (3) Historia construida con los elementos de un viaje en autobús y una persona que sufre un infarto; (4) Historia libre, real o ficticia, a criterio del entrevistado.

### ***Procedimiento***

Se revisaron las transcripciones de los audios previamente obtenidos de la muestra española del proyecto Nexus. Este estudio se enmarca en el enfoque no experimental, dado que no se manipuló ninguna variable, sino que se observó y analizó el comportamiento de los participantes en condiciones naturales. Para el análisis de los datos, se contó con el apoyo de *Chat-GPT-4*, un software de inteligencia artificial reconocido como uno de los sistemas más avanzados en la actualidad.

Las transcripciones de los relatos fueron analizadas en dos etapas. En el grupo experimental, los relatos se procesaron utilizando *Chat-GPT-4*, configurado con los ocho criterios del sistema RM para detectar el engaño. En contraste en el grupo de control, los mismos relatos fueron analizados por *Chat-GPT-4*, sin la configuración basada en los

criterios del sistema RM, limitándose únicamente a determinar si se detectaba indicio de engaño.

El procedimiento de este estudio inicio con la familiarización y manejo del software de inteligencia artificial *Chat-GPT-4*. Posteriormente, se importaron las transcripciones de los audios correspondientes a la muestra y se programó la herramienta para analizar los relatos utilizando los ocho criterios del sistema RM. Los resultados se registraron en una tabla de Excel, asignando valores de 1 cuando se detectaba la presencia de engaño y 0 en caso contrario.

### **3.2 Métodos Técnicas e Instrumentos de Investigación**

Se empleo el sistema RM, el cual es un procedimiento que evalúa la veracidad o falsedad de las declaraciones. Este sistema se basa en la premisa de que la memoria difiere según si los eventos fueron percibidos directamente o imaginados (Fariña y Arce, 2006).

El sistema cuenta con ocho criterios; Claridad (Informe claro, nítido, vívido); Información perceptual (Experiencias sensoriales); Información espacial (Lugares, objetos); Información temporal (Cuándo ocurrió; anclaje); Afectos (Sentimientos); Reconstructibilidad de la historia; Realismo (Historia realista y con sentido); Operaciones cognitivas (Está presente si aparecen inferencias en el momento en que se produjeron los hechos: “a mí me parecía que estaba molesto”) (Bogaard et al., 2013).

### **3.3 Universo de Estudio y Tratamiento Muestral**

La muestra del estudio consistió en adultos sanos residentes del Principado de Asturias, con edades comprendidas entre los 18 y 60 años, seleccionados como parte del Proyecto Nexus. Se incluyeron participantes con capacidad plena para generar narrativas orales en español y sin antecedentes de trastornos psiquiátricos o cognitivos que pudieran interferir el análisis de los relatos.

La selección de la muestra se seleccionó mediante un muestreo no probabilístico por conveniencia, priorizando la accesibilidad de los participantes y su disposición para participar en el estudio. El tamaño de la muestra se estableció en función de la viabilidad del análisis, considerando las limitaciones de tiempo y recursos disponibles para el proyecto.

### **3.4 Tratamiento Estadístico y Representación Gráfica**

Para el OE1, que buscó identificar la confiabilidad de las respuestas obtenidas en la muestra española del proyecto NEXUS utilizando los criterios del RM y el Chat GPT-4 OPEN AI para la detección del engaño, se llevaron a cabo análisis de consistencia interna. Se emplearon los coeficientes alfa de Cronbach y omega de McDonald para evaluar la fiabilidad de los criterios aplicados en narrativas clasificadas como reales, inventadas o mixtas. Los resultados mostraron que las narrativas completamente reales e inventadas alcanzaron coeficientes superiores a 0.80, lo que indicó una alta consistencia interna. Por el contrario, las narrativas mixtas presentaron coeficientes entre 0.60 y 0.70, sugiriendo una mayor heterogeneidad en este tipo de respuestas. Adicionalmente, se calcularon la fiabilidad compuesta (CR) y la validez promedio de extracción (AVE), cuyos resultados confirmaron la robustez de los criterios en las condiciones de narrativas reales e inventadas.

En el OE2, que se centró en comparar los resultados obtenidos entre las diferentes condiciones experimentales, se aplicó un Análisis de Varianza (ANOVA) de una vía. Este análisis permitió identificar diferencias estadísticamente significativas ( $p < 0.05$ ) en los niveles de detección del engaño entre narrativas reales, inventadas y mixtas. Las pruebas post hoc, como el test de Tukey, revelaron que las narrativas mixtas presentaron puntuaciones significativamente diferentes respecto a las reales e inventadas ( $p < 0.01$ ),

destacando que las condiciones puras (reales o inventadas) generaron mayor precisión en la detección.

En relación con el OE3, que evaluó la precisión de los criterios del Monitoreo de la Realidad para la detección del engaño, se implementaron curvas ROC (Receiver Operating Characteristics). Estas curvas permitieron evaluar la sensibilidad (capacidad de identificar narrativas veraces) y la especificidad (capacidad de detectar narrativas falsas). Los resultados mostraron un área bajo la curva (AUC) de 0.85 para narrativas reales y 0.82 para inventadas, lo que indicó un alto nivel de precisión. En cambio, las narrativas mixtas obtuvieron un AUC de 0.72, reflejando una precisión moderada. Paralelamente, se realizó un análisis descriptivo de frecuencias absolutas y relativas, así como de medidas de tendencia central (media y mediana) y dispersión (desviación estándar, valores mínimos y máximos). Este análisis permitió identificar patrones específicos, como una mayor precisión en narrativas reales (90%) frente a narrativas inventadas (85%) y mixtas (70%).

En cuanto a la representación gráfica de los resultados, se utilizaron diversos enfoques para facilitar su interpretación. Los gráficos de barras compararon los porcentajes de detección de engaño según los criterios del RM bajo las diferentes condiciones experimentales, destacando las mayores tasas de precisión en narrativas reales e inventadas. Los diagramas de dispersión ilustraron la relación entre las condiciones experimentales y las puntuaciones obtenidas, lo que permitió identificar agrupaciones claras. Además, los histogramas mostraron la distribución de los niveles de precisión por condición, destacando la variabilidad de las narrativas mixtas. Finalmente, las curvas ROC ofrecieron una representación visual del desempeño del sistema en términos de sensibilidad y especificidad, proporcionando una evaluación integral de su capacidad discriminativa.

## Resultados

**Tabla 2**

*Análisis descriptivo de la muestra general española para la detección del engaño en función a las cuatro condiciones experimentales usando los criterios del RM aplicando el Chat GPT-4 de OPEN AI*

	<b>T1 Inventada</b>	<b>T2 VERDADERA</b>	<b>T3 INVENTADA CON DATOS DE INVESTIGADOR</b>	<b>T4 REAL O INVENTADA</b>
	<b>n (%)</b>	<b>n (%)</b>	<b>n (%)</b>	<b>n (%)</b>
<b>Criterio 1. Claridad (Informe claro, nítido, vívido)</b>				
Verdadero	23 (46,9%)	Verdadero 33 (67,3%)	Verdadero 40 (81,6%)	Verdadero 29 (59,2%)
Engaño	26 (53,1%)	Engaño 16 (32,7%)	Engaño 9 (18,4%)	Engaño 20 (40,8%)
<b>Criterio 2. Información perceptual (Experiencias sensoriales)</b>				
Verdadero	5 (10,2%)	Verdadero 14 (28,6%)	Verdadero 18 (36,7%)	Verdadero 8 (16,3%)
Engaño	44 (89,8%)	Engaño 35 (71,4%)	Engaño 31 (63,3%)	Engaño 41 (83,7%)
<b>Criterio 3. Información espacial (Lugares, objetos...)</b>				
Verdadero	8 (16,3%)	Verdadero 20 (40,8%)	Verdadero 16 (32,7%)	Verdadero 12 (24,5%)
Engaño	41 (83,7%)	Engaño 29 (59,2%)	Engaño 33 (67,3%)	Engaño 37 (75,5%)
<b>Criterio 4. Información temporal (Cuándo ocurrió; anclaje)</b>				
Verdadero	14 (28,6%)	Verdadero 38 (77,6%)	Verdadero 16 (32,7%)	Verdadero 24 (49,0%)
Engaño	35 (71,4%)	Engaño 11 (22,4%)	Engaño 33 (67,3%)	Engaño 25 (51,0%)
<b>Criterio 5. Afectos (Sentimientos)</b>				
Verdadero	13 (26,5%)	Verdadero 27 (55,1%)	Verdadero 13 (26,5%)	Verdadero 22 (44,9%)
Engaño	36 (73,5%)	Engaño 22 (44,9%)	Engaño 36 (73,5%)	Engaño 27 (55,1%)
<b>Criterio 6. Re-constructibilidad de la historia (¿Es posible?)</b>				
Verdadero	20 (40,8%)	Verdadero 43 (87,8%)	Verdadero 43 (87,8%)	Verdadero 38 (77,6%)

Engaño	29 (59,2%)	Engaño	6 (12,2%)	Engaño	6 (12,2%)	Engaño	11 (22,4%)
--------	------------	--------	-----------	--------	-----------	--------	------------

**Criterio 7. Realismo (Historia realista y con sentido)**

Verdadero	25 (51,0%)	Verdadero	47 (95,9%)	Verdadero	45 (91,8%)	Verdadero	44 (89,8%)
Engaño	24 (49,0%)	Engaño	2 (4,1%)	Engaño	4 (8,2%)	Engaño	5 (10,2%)

**Criterio 8. Operaciones cognitivas**

Verdadero	13 (26,5%)	Verdadero	31 (63,3%)	Verdadero	34 (69,4%)	Verdadero	30 (61,2%)
Engaño	36 (73,5%)	Engaño	18 (36,7%)	Engaño	15 (30,6%)	Engaño	19 (38,8%)

---

*Nota:* El análisis descriptivo de los criterios del sistema de Realidad Monitoring (RM) aplicando el Chat GPT-4 de Open AI para la detección del engaño en la muestra general española, se presenta bajo cuatro condiciones: T1 inventada, T2 verdadera, T3 inventada con datos de investigador y T4 real o inventada. A continuación, se desloga el análisis para cada criterio y condición experimental.

La tabla 2 presenta un análisis descriptivo de los criterios del sistema de Monitoreo de la Realidad (RM) aplicados mediante Chat GPT-4 en diferentes condiciones experimentales. Estas condiciones incluyen narrativas completamente inventadas (T1), historias verdaderas (T2), historias inventadas con datos proporcionados por el investigador (T3) y narrativas reales o inventadas a criterio del entrevistado (T4). Los resultados destacan los porcentajes de respuestas clasificadas como "Verdadero" (percepción de veracidad) y "Engaño" (percepción de falsedad) según cada uno de los ocho criterios del sistema RM.

Los criterios como claridad y reconstructibilidad de la historia mostraron un desempeño notable en la detección de veracidad. Por ejemplo, en el criterio de claridad, la percepción de historias "Verdaderas" fue mayor en la condición T3 (81,6%), lo que sugiere que las narrativas con elementos proporcionados por el investigador tienden a estructurarse de manera más clara. Por otro lado, en el criterio de reconstructibilidad, las condiciones T2 y T3 obtuvieron valores elevados de veracidad (87,8% en ambas), lo que refleja que tanto las historias reales como las inventadas con apoyo de datos específicos son percibidas como más coherentes y plausibles.

En contraste, los criterios como información perceptual e información espacial mostraron una tendencia a identificar el engaño. En el caso de la información perceptual, el porcentaje más alto de respuestas categorizadas como Engaño se observó en la condición T1 (89,8%), lo que indica que las narrativas completamente inventadas suelen carecer de detalles sensoriales concretos. De manera similar, el criterio de información espacial reflejó una percepción alta de falsedad en la misma condición (83,7%), evidenciando que las historias ficticias a menudo presentan menos detalles relacionados con lugares y objetos.

Los resultados de los criterios adicionales ofrecen una visión complementaria sobre la capacidad del sistema RM para detectar la veracidad y el engaño en las narrativas. El criterio de información temporal destacó en historias verdaderas (T2), con un 77,6% de percepción de veracidad, indicando que las narrativas reales suelen incluir referencias temporales claras, a diferencia de las historias inventadas que carecen de anclajes temporales consistentes.

De manera similar, el criterio de afectos mostró que las narrativas verdaderas tienen una mayor carga emocional, siendo percibidas como más genuinas en comparación con las inventadas, que suelen ser emocionalmente planas.

Por otro lado, el criterio de realismo fue uno de los más efectivos, con un 95,9% de percepción de veracidad en historias verdaderas, reflejando la coherencia y lógica de las narrativas auténticas. En contraste, el criterio de operaciones cognitivas, aunque mostró resultados intermedios, resaltó que las narrativas basadas en datos reales o ambiguos (T3 y T4) tienden a integrar procesos cognitivos complejos, aumentando su percepción de credibilidad.

Sin embargo, criterios como información perceptual y espacial continuaron asociándose con altos porcentajes de engaño, reflejando la falta de detalles sensoriales y contextuales en narrativas inventadas. En conjunto, estos hallazgos refuerzan que criterios como realismo, afectos e información temporal son esenciales para identificar narrativas veraces, mientras que los demás contribuyen a un análisis más integral del discurso.

Además, un análisis comparativo entre las condiciones experimentales sugiere que las condiciones T3 y T4 presentan una mayor percepción de veracidad en comparación con T1. Esto pone de manifiesto que el contexto narrativo, como la inclusión de datos proporcionados o la ambigüedad en la naturaleza de las historias, influye en la percepción de veracidad. Además, se destaca que criterios como realismo y reconstructibilidad de la historia son más efectivos para identificar narrativas auténticas, mientras que criterios como información perceptual y espacial son útiles para detectar falsedad.

En general, este análisis descriptivo revela variaciones significativas en la percepción de veracidad y engaño a través de los diferentes criterios y condiciones experimentales. Ciertos criterios, como el realismo y la reconstructibilidad de la historia, muestran una fuerte tendencia hacia la percepción de veracidad en historias auténticas o con datos sólidos, mientras que otros criterios, como la información perceptual y espacial, tienden a ser asociados con el engaño independientemente de la condición.

**Tabla 3**

*Fiabilidad total de los resultados de la muestra general española para la detección del engaño en función a las condiciones experimentales considerando los criterios del RM aplicando el Chat GPT-4 de OPEN AI.*

	<b>Media</b>	<b>DE</b>	<b><math>\alpha</math> de Cronbach</b>	<b><math>\omega</math> de McDonald</b>
<b>T1 Inventada</b>	0,691	0,275	0,766	0,769
<b>T2 verdadera</b>	0,355	0,282	0,807	0,807
<b>T3 Inventada con datos de investigador</b>	0,426	0,230	0,654	0,678
<b>T4 Real o Inventada</b>	0,472	0,246	0,673	0,683

*Nota:* Las condiciones utilizadas en la interpretación de los resultados fueron en base a T1 inventada, T2 verdadera, T3 inventada con datos de investigador y T4 real o inventada.

La tabla 3 presenta los resultados de fiabilidad y consistencia interna de los criterios del sistema de Monitoreo de la Realidad (RM) aplicados con Chat GPT-4 bajo diferentes condiciones experimentales. Los resultados con respecto a la fiabilidad reflejan que las condiciones T1 (inventada) y T2 (verdadera) muestran los valores más altos de alfa de Cronbach y omega de McDonald, indicando una mayor consistencia interna y fiabilidad en estos criterios del RM bajo estas condiciones.

Específicamente, la condición T2 (verdadera) presenta los valores más altos de  $\alpha$  y  $\omega$  (0.807), lo cual sugiere que los criterios del RM son particularmente confiables cuando se aplican a narrativas verdaderas.

Por otro lado, la consistencia en la narrativa observó que las condiciones de mayor fiabilidad fueron T1 y T2 lo que indica que los criterios del RM son más efectivos o consistentes al evaluar narrativas completamente inventadas o verdaderas, en comparación con narrativas mezcladas de verdad y ficción (T3 y T4).

Mientras la variabilidad en la aplicación del RM indica que en las medidas de fiabilidad entre las diferentes condiciones sugiere que el contexto de la narrativa (es decir, si es inventada, verdadera, inventada con datos reales, o ambigua) afecta la aplicación de los criterios del RM para la detección del engaño.

De manera general la fiabilidad de los criterios del sistema RM para la detección del engaño varía según el tipo de narrativa evaluada. Los criterios del RM muestran una mayor consistencia interna y fiabilidad cuando se aplican a historias completamente verdaderas o ficticias. Esto destaca la importancia de considerar el contexto de la narrativa al aplicar y evaluar la eficacia de los criterios del RM en la detección del engaño.

#### **Tabla 4**

*Diferencias entre los porcentajes de Engaño en función a la diferentes las condiciones experimentales.*

	<i>Media</i>	<i>Mediana</i>	<i>F</i>	<i>Gl</i>	<i>p</i>
<b>T1 %</b>	69,1	75,0	42,6	3	< .001
<b>T2 %</b>	35,5	37,5			
<b>T3 %</b>	42,6	50,0			
<b>T4 %</b>	47,2	50,0			

*Nota:* El grafico muestra los diferentes porcentajes en la percepción del engaño bajo las condiciones T1 inventada, T2 verdadera, T3 inventada con datos de investigador y T4 real o inventada.

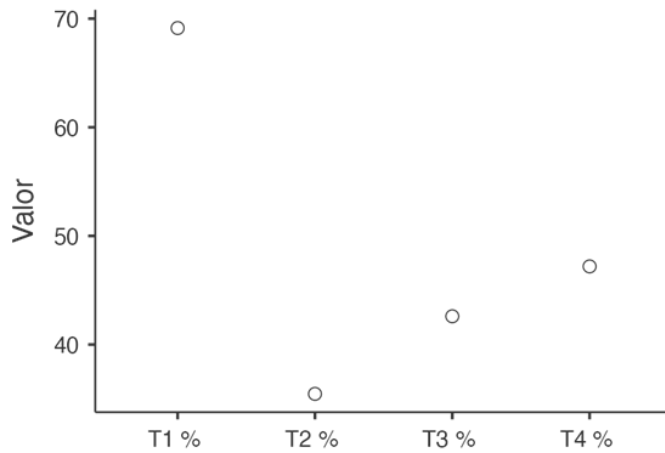
La tabla 4 muestra las diferencias en los porcentajes de detección de engaño bajo las cuatro condiciones experimentales (T1: inventada, T2: verdadera, T3: inventada con datos del investigador, y T4: real o inventada).

Los resultados reflejan una tendencia clara: la condición T1 presenta el mayor porcentaje de detección de engaño (media = 69.1%, mediana = 75.0%), lo que indica que las narrativas completamente inventadas son más fácilmente identificadas como engañosas. En contraste, la condición T2, correspondiente a narrativas completamente verdaderas, tiene el menor porcentaje de engaño (media = 35.5%, mediana = 37.5%), confirmando la mayor coherencia de los relatos auténticos con los criterios del sistema RM.

Las condiciones T3 y T4, que incluyen narrativas mixtas o ambiguas, presentan valores intermedios, lo que evidencia la dificultad del sistema para discernir entre elementos verdaderos e inventados. El análisis estadístico ( $F = 42.6$ ,  $p < 0.001$ ) respalda la existencia de diferencias significativas entre las condiciones, destacando cómo el contexto narrativo afecta la precisión en la detección del engaño.

## Figura 1

*Diferencias del engaño en función de las condiciones experimentales*



*Nota:* En el siguiente gráfico de dispersión con puntos que representan el porcentaje de engaño detectado bajo cuatro condiciones experimentales diferentes (T1 inventada, T2 verdadera, T3 inventada con datos de investigador y T4 real o inventada.) en el eje horizontal y un valor numérico en el eje vertical, probablemente también representando un porcentaje.

La figura 1 indica que el punto para T1 % está significativamente más alto que los demás, lo que indica un mayor porcentaje de engaño detectado bajo esta condición. El punto para T2 % es el más bajo, sugiriendo un porcentaje menor de engaño y los puntos para T3 % y T4 % están aproximadamente a la mitad, indicando niveles intermedios de detección de engaño.

Este patrón sugiere que el sistema de evaluación RM es más propenso a detectar engaño en la condición T1, que se etiquetó como completamente inventada. La condición T2, que era verdadera, mostró el menor porcentaje de engaño, lo cual es esperable si el sistema está funcionando correctamente.

Las condiciones T3 (Inventada con datos de investigador) y T4 (Real o Inventada) muestran resultados intermedios, lo que podría interpretarse como una mayor dificultad

del sistema para discernir entre la verdad y el engaño cuando las narrativas contienen una mezcla de información verdadera y falsa o son ambiguas.

Es decir, el gráfico apoya la idea de que la claridad de la condición (completamente verdadera vs. completamente inventada) afecta la eficacia del sistema de evaluación RM para detectar el engaño. Estos hallazgos refuerzan la capacidad del sistema para distinguir entre narrativas claramente verídicas o fabricadas, destacando la influencia de la claridad narrativa en la efectividad de la detección del engaño.

### **Tabla 5**

*Comparaciones del porcentaje de precisión del engaño entre parejas según la condición.*

			<i>T</i>	<i>P</i>
<b>T1 %</b>	-	<b>T2 %</b>	7,386	< .001
<b>T1 %</b>	-	<b>T3 %</b>	5,442	< .001
<b>T1 %</b>	-	<b>T4 %</b>	4,470	< .001
<b>T2 %</b>	-	<b>T3 %</b>	1,944	0.054
<b>T2 %</b>	-	<b>T4 %</b>	2,915	0.004
<b>T3 %</b>	-	<b>T4 %</b>	0,972	0.333

*Nota:* La tabla X muestra los resultados de pruebas t de Student. Los valores t y p indican la magnitud de la diferencia entre los grupos y si esta diferencia es estadísticamente significativa.

La tabla 5 indica que la comparación entre la condición inventada y la verdadera muestra un valor t de 7,386, con un valor p menor que .001. Esto indica una diferencia significativa y grande en la precisión de la detección del engaño entre estas dos condiciones, con una precisión mucho mayor en la condición verdadera (T2).

La comparación entre la condición inventada y la inventada con datos de investigador también muestra una diferencia significativa, con un valor t de 5,442 y un valor p menor que .001. Esto sugiere que la condición con datos de investigador (T3) tiene una precisión significativamente mayor que la condición completamente inventada (T1).

Al comparar la condición inventada con la condición real o inventada, hay una diferencia significativa, con un valor t de 4,470 y un valor p menor que .001, lo que indica que la condición T4 tiene una precisión mayor en la detección del engaño comparada con T1.

Así mismo al comparar la condición verdadera con la inventada con datos de investigador, el valor t es de 1,944 y el valor p es de 0,054, lo cual está justo por encima del umbral común de significancia estadística de 0,05, sugiriendo que no hay una diferencia significativa en la precisión de detección del engaño entre estas dos condiciones o que la diferencia es pequeña.

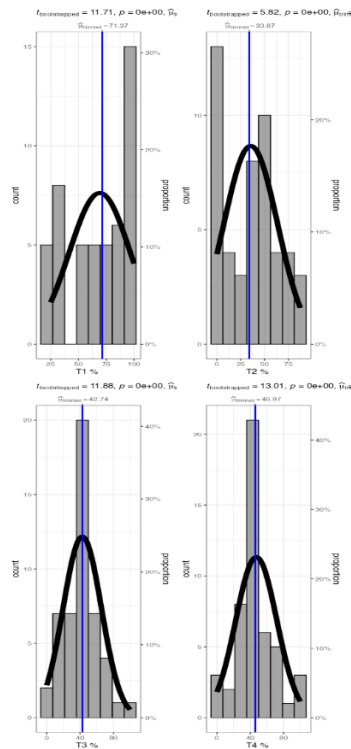
Hay una diferencia estadísticamente significativa entre la condición verdadera y la real o inventada, con un valor t de 2,915 y un valor p de 0,004. Esto indica una mayor precisión en la condición verdadera (T2) en comparación con la real o inventada (T4).

Finalmente, no hay una diferencia estadísticamente significativa entre la condición inventada con datos de investigador y la real o inventada, con un valor t de 0,972 y un valor p de 0,333, lo que sugiere que la precisión de la detección del engaño es similar en estas condiciones. Estos hallazgos destacan que el sistema RM combinado con Chat GPT-4 es particularmente eficaz para detectar engaños evidentes en narrativas completamente inventadas (T1) y demuestra una alta precisión en narrativas verdaderas (T2). Sin embargo, las condiciones intermedias o ambiguas (T3 y T4) presentan un desafío mayor,

lo que sugiere la necesidad de mejorar la capacidad del sistema para manejar narrativas complejas donde la verdad y la mentira están mezcladas.

## Figura 2

*Distribución del Porcentaje de precisión del engaño (Asumiendo que la distribución representa la precisión debido a la discusión previa)*



*Nota:* Histogramas con curva de densidad, para las cuatro condiciones experimentales: T1, T2, T3 y T4. Cada gráfico representa la distribución del porcentaje de precisión del engaño (asumiendo que la distribución representa la precisión debido a la discusión previa) para cada condición experimental.

La Figura 2 muestra la distribución del porcentaje de precisión en la detección del engaño para las cuatro condiciones experimentales: T1 (inventada), T2 (verdadera), T3 (inventada con datos de investigador) y T4 (real o inventada). En esta gráfica, cada histograma está acompañado por una curva de densidad que representa la probabilidad de precisión dentro de cada condición.

La condición T1 (inventada) presenta la mayor dispersión en los porcentajes de precisión del engaño, indicando que el sistema de RM, al aplicarse con Chat GPT-4, detecta más variabilidad en esta categoría. Esto indica que, cuando los participantes elaboraron historias completamente inventadas, el sistema presentó una notable variabilidad al identificar elementos de engaño. Los resultados en esta categoría sugieren que las narrativas completamente fabricadas pueden incluir características ambiguas que dificultan una evaluación uniforme.

Por otro lado, la condición T2 (verdadera) muestra una distribución más concentrada, con menores variaciones en los valores. Esto refleja que el sistema tuvo un desempeño más consistente al evaluar narrativas completamente veraces. La alta precisión alcanzada en esta condición reafirma que los criterios del sistema RM son particularmente eficaces al analizar relatos auténticos.

En las condiciones T3 (inventado con datos proporcionados por el investigador) y T4 (real o inventada a criterio del participante), la distribución de la precisión se sitúa entre las condiciones T1 y T2. Esto sugiere que las narrativas que contienen una mezcla de elementos reales e inventados o que tienen una ambigüedad en su autenticidad representan un desafío mayor para el sistema. La mayor complejidad de estas narrativas puede haber reducido la capacidad del sistema para distinguir con claridad entre verdad y engaño.

La uniformidad en la significancia estadística ( $p=0$ ) para todas las condiciones sugiere que hay diferencias significativas en la precisión de detección del engaño entre las condiciones experimentales o entre las condiciones y un valor o condición de referencia.



<b>T3 %</b>	42,6	50,0	22,9	0.0	100,	25,0	50,0	62,50
<b>menti</b>	0		5		0	0	0	
<b>ra</b>								
<b>T4 f</b>	3,78	4	1,97	0	8	2.00	4,00	5,00
<b>menti</b>								
<b>ra</b>								
<b>T4 %</b>	47,1	50,0	24,6	0.0	100,	25,0	50,0	62,50
<b>menti</b>	9		4		0	0	0	
<b>ra</b>								

---

*Nota:* La tabla refleja el porcentaje de dispersión y el rango en el que se encuentran los datos.

La tabla 6 presenta un análisis descriptivo de los porcentajes y frecuencias relacionados con la detección de mentiras en las diferentes condiciones experimentales. Muestra que las medias y medianas están relativamente alineadas en todas las condiciones, lo que indica distribuciones simétricas con respecto al centro de los datos.

Las amplias gamas de mínimos y máximos, junto con las desviaciones estándar relativamente altas, sugieren una variabilidad significativa en la precisión de la detección del engaño entre individuos o situaciones dentro de cada condición experimental. Los percentiles indican la dispersión y el rango en el que se encuentran los datos, lo que puede ser útil para comprender mejor la distribución de la precisión de la detección del engaño en cada condición.

En cuanto a los valores promedio, la condición T1 muestra la mayor media de detección de mentiras con un 69,13%, lo que confirma que el sistema RM es altamente eficaz al identificar narrativas completamente inventadas. En contraste, la condición T2 presenta la menor media, con un 35,46%, lo que es coherente con la naturaleza verdadera

de las narrativas en esta categoría, reflejando una menor proporción de engaño. Por su parte, las condiciones T3 y T4 tienen valores intermedios (42,60% y 47,19%, respectivamente), lo que indica que el sistema enfrenta una mayor complejidad para discernir entre elementos verdaderos e inventados en estas narrativas.

Por otro lado, la dispersión de los datos, medida a través de la desviación estándar (DE), revela una mayor variabilidad en la condición T1, con un valor de 27,49%. Esto sugiere que la percepción del engaño es menos uniforme en narrativas completamente inventadas. En comparación, las condiciones T3 y T4 presentan una dispersión intermedia (22,95% y 24,64%, respectivamente), mientras que T2 muestra una D.E. de 28,22%, atribuible a errores puntuales del sistema en la clasificación de narrativas verdaderas.

Los valores extremos en las categorías de mínimo y máximo reflejan la naturaleza diversa de las narrativas. Mientras que en todas las condiciones se detectaron casos con un 0% de mentiras, los máximos alcanzan el 100% en T1, T3 y T4, indicando que ciertas narrativas fueron clasificadas completamente como engaño. Estos hallazgos son especialmente relevantes en T3 y T4, donde las narrativas mezcladas o ambiguas plantean un desafío adicional para el sistema.

Los percentiles proporcionan una comprensión más detallada de la distribución de los datos. En T1, el 75% de las narrativas tienen un porcentaje de detección de mentiras igual o inferior al 100%, subrayando la alta proporción de engaño detectado en esta condición. Por el contrario, en T2, los percentiles son significativamente más bajos, con el 50% de las narrativas registrando un porcentaje igual o inferior al 37,5%. Esto es coherente con la expectativa de que las narrativas verdaderas contienen menos elementos engañosos.

## 4. CAPITULO IV: LA PROPUESTA

### 4.1 Antecedentes de la propuesta

La detección del engaño ha sido un área de intereses en diversas disciplinas, particularmente en la Psicología Forense y la Psicología Clínica (Blandón 2017). En el contexto de la Psicoterapia, la veracidad de la información es fundamental para un tratamiento efectivo. Sin embargo, los pacientes ya sea de manera consciente o inconsciente, pueden no ser completamente honestos respecto a sus síntomas o experiencias (Altamirano y Bernuy, 2022).

En el ámbito forense, la detección del engaño es fundamental para garantizar la integridad del sistema judicial, ya que la identificación de declaraciones falsas puede evitar decisiones erróneas que afecten la vida de las personas involucradas (Pallejá et al., 2002). Evaluar la credibilidad de testigos y acusados es esencial en casos críticos como abuso y fraude, donde las pruebas son predominantemente testimoniales. Detectar el engaño contribuye a prevenir injusticias, mejorar las técnicas de interrogación y reducir el estrés asociado con falsas declaraciones

Las investigaciones en este campo se han centrado principalmente en la búsqueda de sus indicios con la finalidad de detectar la mentira de forma eficaz, a través de indicadores fisiológicos, conductuales, paraverbales y del mensaje verbal (Köhnken et al., 2015). No obstante, cuando acudimos a consulta con un psicólogo, es común sentirse vulnerable y expuesto, especialmente si estamos compartiendo información personal o dolorosa.

En los estudios de Jonhson y Raye establecen que el origen de nuestros recuerdos que se codifican y almacenan en nuestra memoria proceden de dos fuentes básicas: La externa generados a partir de la percepción de situaciones reales y la interna producto de

nuestra imaginación. (Soto et al, 2013). Por ello, el proceso de razonamiento que sigue nuestra mente para percibir que información procede de una fuente es lo que se denomina Monitoreo de la Realidad (RM). Las herramientas basadas en RM demuestran su potencial en la diferenciación entre recuerdos reales y fabricados.

Recientemente, la inteligencia artificial ha surgido como una tecnología capaz de mejorar la precisión y la eficiencia en estos sistemas, al procesar grandes volúmenes de datos y detectar patrones complejos que pueden pasar desapercibidos para los humanos (Gordon y Turnbull, 2024). La combinación de la IA con los sistemas RM pueden ofrecer oportunidades prometedoras tanto en la psicoterapia como en la evaluación forense, optimizando la precisión diagnóstica y mejorando la efectividad de los tratamientos.

#### **4.2 Introducción**

En los últimos años, la IA ha comenzado a transformar muchas áreas de la Psicología y la salud mental, desde el diagnóstico hasta la intervención clínica. Uno de los campos más prometedores es la integración de la IA con el sistema RM para la detección del engaño (Alqahtani, 2023).

En la práctica psicoterapéutica, la veracidad de la información proporcionada por el paciente es fundamental para un diagnóstico y tratamiento exitoso. Sin embargo, investigaciones recientes en Psicología clínica han demostrado que los seres humanos son propensos a la distorsión de recuerdos, ya sea de forma consciente o inconsciente. (Soto et al., 2013).

En el contexto de la psicoterapia, comprender por qué un paciente utiliza la mentira como recurso crucial para interpretar su comportamiento y sus procesos cognitivos y emocionales. La mentira puede ser un mecanismo de defensa, una forma de

evitar el dolor emocional o una estrategia para manejar situaciones difíciles (Blandón et al, 2017).

Es importante entender que la experiencia que un paciente comparte en terapia no siempre refleja una verdad objetiva, sino una narrativa influenciada por su percepción y circunstancias personales. Si la terapia se convierte en un esfuerzo por definir una “verdad” absoluta, esto puede resultar contraproducente para el proceso terapéutico (Palacios, 2022). Por lo tanto, en lugar de imponer una visión externa, el terapeuta debe guiar al paciente en la revisión y reformulación de sus interpretaciones de una manera más constructiva, equilibrando la aceptación provisional de sus narrativas con la responsabilidad de ayudarlo a desarrollar un estilo de vida más saludable.

En este contexto, el sistema RM ha sido ampliamente estudiado como una estrategia basada en experiencias reales. Además, es un sistema de medición del proceso cognitivo que permite a las personas evaluar la fuente de sus recuerdos, reconociendo características sensoriales, contextuales y emocionales asociadas a las experiencias reales. (Gancedo et al., 2021). Sin embargo, cuando este sistema se ve alterado, ya sea por factores psicológicos, como trastornos de la memoria o del pensamiento, por intenciones conscientes de engaño o por recuerdos traumáticos, el resultado es una percepción distorsionada de la realidad (Huacón y Martínez, 2021).

La integración de la inteligencia artificial en el sistema del RM representa una solución innovadora para superar estas limitaciones actuales en la evaluación de la veracidad de los recuerdos. Gracias a su capacidad para procesar grandes volúmenes de datos en tiempo real, la IA puede identificar patrones complejos que pueden pasar desapercibidos para los terapeutas (Guzmán, 2023).

Estos sistemas pueden analizar no solo el contenido verbal que el paciente comparte, sino también las señales no verbales, como el tono de voz, las micro expresiones faciales y el lenguaje corporal, que a menudo son indicadores clave de incongruencia entre lo que se dice y lo que realmente se experimenta (Guamán, 2023).

Los avances en los algoritmos de IA, especialmente aquellos relacionados con el procesamiento del lenguaje natural y el aprendizaje profundo, han permitido desarrollar modelos capaces de interpretar patrones de comunicación humana con un alto grado de precisión (Gordon & Turnbull, 2024). La combinación del RM y la IA tiene el potencial de detectar con mayor exactitud las señales sutiles de engaño, ofreciendo a los terapeutas una herramienta valiosa para mejorar el proceso de toma de decisiones clínicas.

Además, esta tecnología no solo serviría para identificar el engaño de manera directa, sino que también podría proporcionar una visión más profunda de las posibles causas subyacentes del mismo (Boden, 2017). Por ejemplo, los pacientes que sufren de trastorno de estrés postraumático o la esquizofrenia suelen experimentar alteraciones en su percepción de la realidad asociados a la experiencia negativa. Estas alteraciones no siempre son conscientes, pero afectan profundamente su bienestar emocional y psicológico.

Con el apoyo de la IA, sería posible rastrear las conexiones entre estas distorsiones y sus desencadenantes, lo que permitiría al terapeuta adaptar las intervenciones terapéuticas de manera más precisa y efectiva (Altamirano y Bernuy, 2022). Esta integración representa una evolución en el campo de la salud mental y la psicoterapia, al ofrecer una solución avanzada para la detección del engaño y la mejora de los resultados terapéuticos.

### **4.3 Objetivo**

Analizar la efectividad del sistema de Monitoreo de la Realidad y la herramienta de Inteligencia Artificial Chat GPT-4 en la detección del engaño a través del análisis de contenido del discurso, utilizando una muestra general de participantes españoles, con el fin de mejorar la precisión en la evaluación de la credibilidad de los testimonios.

### **4.4 Justificación**

La efectividad de cualquier intervención terapéutica depende en gran medida de la calidad de la información que el paciente proporciona. El engaño, tanto consciente como inconsciente, puede interferir en el proceso terapéutico, limitando los resultados y afectando la toma de decisiones clínicas.

Herramientas como el RM permiten a los terapeutas evaluar la autenticidad de los recuerdos y experiencias relatadas por los pacientes. Sin embargo, la complejidad de los patrones cognitivos y emocionales relacionados con el engaño requiere una mayor capacidad de procesamiento y análisis más avanzado (Nahari, 2018). En ese sentido la integración de la IA en este campo ofrece un enfoque innovador al proporcionar a los terapeutas análisis más detallados y precisos que no solo detecten incongruencias en los relatos, sino que también arrojan luz sobre las posibles causas subyacentes del engaño.

La aplicación de criterios basados en el RM, combinada con algoritmos de IA, permite un análisis objetivo y multidimensional de la veracidad de los relatos de los pacientes, evaluando no solo el contenido verbal sino también señales como el lenguaje corporal, las expresiones faciales y el tono de voz (Boden, 2017). Esto reduce el sesgo terapéutico y permite una intervención más oportuna y efectiva.

Esta tecnología resulta especialmente útil en casos donde los pacientes presentan trastornos de memoria, como en el caso del trastorno de estrés postraumático (TEPT), la

esquizofrenia, o el trastorno de personalidad esquizoide, donde la distorsión de la realidad es común (Brito et al., 2021).

El camino de esta integración de sistemas no solo optimiza el proceso terapéutico, sino que también mejora la eficiencia, permite alertas tempranas y contribuye al avance del conocimiento clínico y científico. No obstante, es importante garantizar su implementación respete la ética y la confidencialidad del paciente.

#### 4.5 Alcance

La integración de la IA y RM puede tener un impacto considerable en diversas áreas de la Psicología forense y la clínica. Los beneficios abarcan desde el nivel clínico individual hasta aplicaciones en salud mental comunitaria y preventiva.

En primer lugar, tenemos una mejora en la detección de engaños cognitivos; en trastornos como el TEPT o la esquizofrenia, los pacientes suelen presentar dificultades para distinguir entre recuerdos reales e imaginados. (Huacón y Martínez., 2021). La integración de la IA y RM permite analizar grandes volúmenes de datos relacionados con los patrones de comportamiento y lenguaje de los pacientes, detectando señales sutiles de engaño o falso que un terapeuta podría pasar por alto.

Esto facilitaría una identificación temprana de recuerdos alterados, ayudando a los terapeutas a abordar estos problemas más rápidamente. Además, en caso de trastornos donde el retraimiento emocional y la falta de conexión con la realidad son comunes, el sistema RM integrado con la IA podría diferenciar entre una interpretación sincera de la realidad y una distorsión cognitiva (Gordon y Turnbull., 2024).

Esta integración permite crear nuevas formas de realizar intervenciones basadas en análisis de datos. La IA puede transformar la forma en que se realiza la psicoterapia, al proporcionar un análisis de datos en tiempo real que permita detectar micro expresiones

faciales, cambios en el tono de voz, pausas lingüísticas y otros indicadores que revelen engaño (Guamán, 2023).

Al combinar estos indicadores con un análisis emocional del discurso del paciente, los terapeutas pueden obtener un retrato más completo de la experiencia subjetiva del paciente y ajustar su enfoque terapéutico según los resultados (Salah, 2023). Esto puede ser especialmente útil en terapia cognitivo conductuales, donde las distorsiones cognitivas juegan un papel central o en terapia centradas en el trauma, donde el paciente puede tener dificultades para procesar recuerdos dolorosos de manera precisa.

Otro alcance está en que los terapeutas en formación podrían beneficiarse del uso de la IA en su proceso de aprendizaje, utilizando el sistema RM como herramientas de simulación para evaluar cómo lidiar con situaciones donde el paciente presenta información falsa o engañosa. La tecnología podría ayudar a entrenar a los terapeutas para identificar patrones de engaño, proporcionando retroalimentación en tiempo real y simulando escenarios clínicos complejos (Iqbal et al., 2023).

Esto mejoraría las habilidades para detectar y manejar recuerdos falsos o distorsiones en un entorno controlado antes de enfrentar estos desafíos en la práctica clínica real. Finalmente, contribuiría en la investigación clínica avanzada, la recolección de datos a gran escala facilitada por la IA podría abrir nuevas áreas de investigación en Psicología y salud mental.

Los datos obtenidos a través de las evaluaciones con RM podrían permitir a los investigadores estudiar cómo los pacientes distorsionan la realidad en función de diferentes variables, por ejemplo, tipo de trastorno, edad, género. A largo plazo, estas investigaciones podrían conducir al desarrollo de nuevas estrategias terapéuticas más

eficaces para abordar los problemas relacionados con el engaño y en pacientes de diversas poblaciones.

## **5. CONCLUSIONES, LIMITACIONES Y PROSPECTIVA**

El estudio logra un avance significativo en la comprensión y aplicación de los criterios del RM para la detección del engaño, resaltando su potencial en diversas aplicaciones prácticas y abriendo caminos para futuras investigaciones en el campo.

Demostó la fiabilidad de los criterios del sistema Reality Monitoring para la detección del engaño, especialmente en narrativas verdaderas y completamente inventadas, utilizando valores de alfa de Cronbach y omega de McDonald. Este logro es fundamental para validar el uso de estos criterios en futuras aplicaciones prácticas y estudios.

A través de un análisis detallado, se identificaron diferencias estadísticamente significativas en la percepción del engaño entre diferentes condiciones experimentales. Este hallazgo subraya cómo la composición y el contexto de las narrativas afectan la detección del engaño, proporcionando una guía valiosa para interpretar testimonios o declaraciones en contextos forenses y psicológicos.

El estudio mostró que los criterios del RM tienen diferentes niveles de precisión en la detección del engaño, dependiendo de si la narrativa es verdadera, inventada, o contiene elementos de ambos. Este logro destaca la importancia de adaptar y aplicar el RM de manera flexible según el contexto de la narrativa para maximizar su efectividad.

El análisis proporciona evidencia empírica sobre la utilidad del RM, aplicado a través de Chat GPT-4, en contextos donde es crucial determinar la veracidad de las narrativas. Estos hallazgos pueden ser aplicados en la investigación forense y la

Psicología del testimonio, ofreciendo una base sólida para futuras investigaciones y prácticas en la detección del engaño.

Al aplicar tecnología avanzada como el Chat GPT-4 en la evaluación de la veracidad y el engaño, el estudio contribuye al avance del conocimiento en el campo de la detección automatizada del engaño. Esto podría llevar al desarrollo de herramientas más sofisticadas y precisas para la evaluación de testimonios y declaraciones.

Con respecto a las limitaciones encontradas durante el estudio fueron; en la muestra limitada a una población, se centró en una muestra general española. Esto podría limitar la generalización de los resultados a otras poblaciones con diferentes características culturales, lingüísticas o demográficas.

Las condiciones experimentales (inventada, verdadera, inventada con datos de investigador, y real o inventada) son representativas de un rango de posibilidades en narrativas. Sin embargo, existen muchas otras formas de engaño y veracidad que pueden no estar cubiertas por estas condiciones, limitando la aplicabilidad de los resultados a situaciones fuera de estos escenarios.

El uso del *Chat GPT-4* como herramienta para aplicar los criterios de RM implica una dependencia de las capacidades actuales de la inteligencia artificial, que, aunque avanzadas, tienen limitaciones en comprensión, interpretación y generación de lenguaje natural que podrían influir en la detección del engaño.

La aplicación y la interpretación de los criterios de RM pueden estar sujetas a la subjetividad del investigador o al algoritmo de inteligencia artificial, lo que podría introducir un sesgo en la evaluación de la veracidad y el engaño. La detección del engaño puede estar influenciada por factores individuales (como habilidades de comunicación o niveles de estrés) y contextuales (como la naturaleza del evento descrito o la relación

entre el narrador y el evaluador) que no se controlan completamente en condiciones experimentales.

Finalmente, la tecnología de IA, incluido el Chat GPT-4, está en constante evolución, lo que significa que los resultados obtenidos podrían ser superados por avances futuros en la tecnología, afectando la relevancia a largo plazo de las conclusiones del estudio.

## **6. RECOMENDACIONES**

Realizar estudios con muestras más amplias y diversas culturalmente para evaluar la generalización de los resultados. Incluir participantes de diferentes grupos de edad, antecedentes culturales y lingüísticos para comprender mejor cómo estos factores pueden influir en la percepción y detección del engaño.

También explorar una gama más amplia de condiciones experimentales, incluyendo diferentes tipos de engaños y contextos narrativos, para evaluar la robustez y flexibilidad de los criterios de RM en la detección del engaño.

Además de combinar el uso de criterios de RM con otras técnicas de detección del engaño, como análisis de micro expresiones, lenguaje corporal, y otros indicadores psicológicos, para mejorar la precisión y fiabilidad de la detección

Por otro lado, a medida que la tecnología de IA continúa evolucionando, es crucial desarrollar y actualizar los modelos de IA utilizados para la detección del engaño, asegurando que se adapten a las complejidades del lenguaje humano y sean sensibles a las sutilezas del engaño.

Finalmente Investigar y trabajar en las limitaciones actuales de la IA en la comprensión y generación del lenguaje natural, especialmente en contextos complejos y

ambivalentes, para mejorar su capacidad de detección del engaño. Además de realizar estudios longitudinales para evaluar la eficacia de los criterios de RM en el tiempo, especialmente según evolucionan las herramientas de IA y cambian los contextos socioculturales.

## Referencias

- Abe, N., Suzuki, M., Mori, E., Itoh, M., & Fujii, T. (2007). Deceiving Others: Distinct Neural Responses of the Prefrontal Cortex and Amygdala in Simple Fabrication and Deception with Social Interactions. *Journal Of Cognitive Neuroscience*, 19(2), 287-295. <https://doi.org/10.1162/jocn.2007.19.2.287>
- Alazrai, R., Alqasem, F., Alaarag, S., Ahmad Yousef, K. y Daoud, M. (2018). A bispectrum-based approach for detecting deception using EEG signals. *2018 IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom)*, 1–6. [10.1109/HealthCom.2018.8531183](https://doi.org/10.1109/HealthCom.2018.8531183)
- Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A., Alghamdi, S., Saleh, K., Alowais, S., Alshaya, O., Rahman, I., Yami, M., & Albekairy, A. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research In Social and Administrative Pharmacy*, 19(8), 1236-1242. <https://doi.org/10.1016/j.sapharm.2023.05.016>
- Altamirano, J., & Bernuy, A. (2022). La inteligencia artificial en la Psicología: nuevos enfoques para la detección de las declaraciones falsas. *RISTI - Revista Ibérica de Sistemas E Tecnologías de Informação*, 46, 100-111. <https://doi.org/10.17013/risti.46.100-111>
- Álvarez, J., & Yusti, I. (2015). Personalidad, mentira y engaño. *Behavior & Law Journal*, 1(1), 17-26. <https://doi.org/10.47442/blj.v1.i1.12>
- Blandón, I., López, R., Masip, J., & Fenn, E. (2017). Cognición, emoción y mentira: implicaciones para detectar el engaño. *Anuario de Psicología Jurídica*, 27(1), 95-106. <https://doi.org/10.1016/j.apj.2017.02.004>

- Boden, A. (2017). *Inteligencia artificial*, Madrid: Turner (1.<sup>a</sup> ed.).  
[https://www.turnerlibros.com/wp-content/uploads/2022/07/Adelanto\\_InteligenciaArtificial.pdf](https://www.turnerlibros.com/wp-content/uploads/2022/07/Adelanto_InteligenciaArtificial.pdf)
- Bogaard, G., Meijer, H. & Vrij, A. (2013). Using an example statement increases information but does not increase accuracy of CBCA, RM, and SCAN. *Journal of Investigative Psychology and Offender Profiling*, 11(2), 151-163. <https://scihub.se/https://doi.org/10.1002/jip.1409>
- Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., & Su, L. (2014). Countering Countermeasures: Detecting Identity Lies by Detecting Conscious Breakthrough. *PLOS ONE*, 9(3). <https://doi.org/10.1371/journal.pone.0090595>
- Brito, T., Martínez, P., & Ramírez, A. (2021). Revisión de instrumentos psicométricos para la detección del engaño en la evaluación psicológica. *Pro Sciences: Revista de Producción, Ciencias e Investigación*, 5(39), 86–107.  
<https://journalprosciences.com/index.php/ps/article/view/359/423>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. *Neural Information Processing Systems*, 33, 1877-1901.  
<https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf>
- Carvajal, J., & Barreto, I. (2022). Capítulo 14: Detección del engaño a partir del estilo lingüístico: Una revisión sistematizada. *Psicología Jurídica: Género, contexto y enfoques para la comprensión del comportamiento antijurídico*, 173-183. Manual

Moderno. [https://www.researchgate.net/profile/Juan-Carvajal-8/publication/365127315\\_Deteccion\\_del\\_engano\\_a\\_partir\\_del\\_estilo\\_linguistico\\_Una\\_revision\\_sistematizada/links/6365781b431b1f53006d7210/Deteccion-del-engano-a-partir-del-estilo-lingueistico-Una-revision-sistematizada.pdf](https://www.researchgate.net/profile/Juan-Carvajal-8/publication/365127315_Deteccion_del_engano_a_partir_del_estilo_linguistico_Una_revision_sistematizada/links/6365781b431b1f53006d7210/Deteccion-del-engano-a-partir-del-estilo-lingueistico-Una-revision-sistematizada.pdf)

Carvajal, J., Gómez, E. & Barreto, I. (2021). Comportamiento visual y engaño: Una revisión sistematizada. *Anuario de Psicología*, 51(1), 120-129. <http://doi.org/10.1344/ANPSIC2021.51.14>

Cierco, D. (30 de enero de 2024). *La revolución de la inteligencia artificial*. El Español. [https://www.elespanol.com/invertia/disruptores-innovadores/opinion/20240130/revolucion-inteligencia-artificial/828797114\\_13.html](https://www.elespanol.com/invertia/disruptores-innovadores/opinion/20240130/revolucion-inteligencia-artificial/828797114_13.html)

Ciprés, M., Seguer, R., & Fernández, Ó. (2024). ChatGPT como herramienta de apoyo al aprendizaje en la educación superior: una experiencia docente. *Revista Tecnología Ciencia y Educación*, 7-44. <https://doi.org/10.51302/tce.2024.19083>

Colman, A. (2009). *Un diccionario de psicología*. Oxford Reference. <https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100407823#:~:text=Reality%20monitoring%20is%20a%20form,real%20testing%20in%20clinical%20psychology>.

Constâncio, S., Tsunoda, F., Silva, N., Silveira, D. & Carvalho, R. (2023). Deception detection with machine learning: A systematic review and statistical analysis. *PLOS ONE*, 18(2). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9910662/>

DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003).

Cues to deception. *Psychological Bulletin*, 129(1), 74-

118. <https://doi.org/10.1037/0033-2909.129.1.74>

Ekman, P. (2021). *Cómo detectar mentiras: Una guía para utilizar en el trabajo, la política y la familia*. Ediciones Paidós.

<https://www.marcialpons.es/media/pdf/9788449337093.pdf>

Feijoo, M. & Halty, L. (2018). La intención es lo que cuenta. Una revisión sobre la detección del engaño en intenciones. *Papeles del psicólogo*, 39(1), 51-61.

<https://www.redalyc.org/journal/778/77854690006/77854690006.pdf>

Fernandes, S., & Ullah, M. (2022). A Comprehensive Review on Features Extraction and Features Matching Techniques for Deception Detection. *IEEE Access*, 10, 28233-

28246. <https://doi.org/10.1109/access.2022.3157821>

Fariña, F & Arce, R. (2006). *Psicología del testimonio: Evaluación de la credibilidad y de la huella Psíquica en el contexto penal*. Consejo General del Poder Judicial.

[https://www.researchgate.net/publication/277047252\\_Psicologia\\_del\\_testimonio\\_Evaluacion\\_de\\_la\\_credibilidad\\_y\\_de\\_la\\_huella\\_psiquica\\_en\\_el\\_contexto\\_penal](https://www.researchgate.net/publication/277047252_Psicologia_del_testimonio_Evaluacion_de_la_credibilidad_y_de_la_huella_psiquica_en_el_contexto_penal)

Gancedo, Y., Fariña, F., Seijo, D., Vilariño, M., & Arce, R. (2021). Reality Monitoring: A Meta-analytical Review for Forensic Practice. *The European Journal Of Psychology Applied To Legal Context*, 13(2), 99-110. <https://doi.org/10.5093/ejpalc2021a10>

Garrison, R., Bond, R., Gibbard, E., Johnson, K. & Simons, S. (2017). Monitoring what is real: The effects of modality and action on accuracy and type of reality monitoring error. *Cortex: a journal devoted to the study of the nervous system and behavior*, 87,

108–117. [10.1016/j.cortex.2016.06.018](https://doi.org/10.1016/j.cortex.2016.06.018)

- Gómez, F., & Galende, J. (2024). ChatGPT y GPT-4: utilidades en el sector jurídico, funcionamiento, limitaciones y riesgos de los modelos fundacionales. *Revista Tecnología Ciencia y Educación*, 45-88. <https://doi.org/10.51302/tce.2024.19081>
- Gordon, S., & Turnbull, B. (2024). Adopción de la inteligencia artificial en el campo de la psicología. *Psicología Iberoamericana*, 31(2). <https://doi.org/10.48102/pi.v31i2.547>
- Guamán, M. (2023). ChatGPT, ventajas, desventajas y el uso en la Educación Superior. *Killkana Social*, 7(1), 3-8. <https://doi.org/10.26871/killkanasocial.v7i1.1270>
- Guzmán, A. (2023). ChatGPT, el nuevo y asombroso chatbot de inteligencia artificial. *Ciencia*, 74, 3. [https://www.revistaciencia.amc.edu.mx/images/revista/74\\_3/PDF/14\\_74\\_3\\_1524.pdf](https://www.revistaciencia.amc.edu.mx/images/revista/74_3/PDF/14_74_3_1524.pdf)
- Hartwig, M., Granhag, P., & Strömwall, L. (2006). Guilty and innocent suspects' strategies during police interrogations. *Psychology Crime And Law*, 13(2), 213-227. <https://doi.org/10.1080/10683160600750264>
- Huacón, C., & Martínez, P. (2021). Tecnologías aplicadas a la medición del Reality Monitoring (RM). *Revista Innovación y Desarrollo Sostenible*, 2(1), 57-71. <https://doi.org/10.47185/27113760.v2n1.45>
- Iqbal, J., Jaimes, D., Makineni, P., Subramani, S., Hemaída, S., Thugu, T., Butt, A., Sikto, J., Kaur, P., Lak, M., Augustine, M., Shahzad, R., & Arain, M. (2023). Reimagining Healthcare: Unleashing the Power of Artificial Intelligence in Medicine. *Cureus*. <https://doi.org/10.7759/cureus.44658>
- Johnson, K., Raye, L., Schacter, L. & Scarry, E. (2000). Memory, brain, and belief. *DL Schacter & E. Scarry (Eds.)*, 35-86. <https://psycnet.apa.org/record/2000-07142-001>

- Lavallé, L., Brunelin, J., Jardri, R., Haesebaert, F. & Mondino, M. (2023). The neural signature of reality-monitoring: A meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, 44(11), 4372–4389.  
<https://doi.org/10.1002/hbm.26387>
- Lemos, S., Rizzatti, M., Pereira, T., Sangoi, B., Nascimento, V. & Machado, A. (2019). Understanding Lies Based on Evolutionary Psychology: A Critical Review. *Trends in Psychology*, 27(1), 141-153. <http://doi/10.9788/TP2019.1-11>
- Markowitz, M., Hancock, T., Woodworth, T. & Ely, M. (2023). Contextual considerations for deception production and detection in forensic interviews. *Frontiers in psychology*, 14, 1134052. [10.3389/fpsyg.2023.1134052](https://doi.org/10.3389/fpsyg.2023.1134052)
- Martin, K. & Leach, A. M. (2013). Psychopathy and deception detection. *Personality and mental health*, 7(2), 154–159. <https://doi.org/10.1002/pmh.1215>
- Martínez, C., Andrade, A., Coronel, A., Mejía, B., Carrasco, E. & Carpio, M. (2021). Reality monitoring measurement using Balls Control Test (BCT) Software. *Archivos Venezolanos de Farmacología y Terapéutica*, 40(5), 539-546.  
<https://www.redalyc.org/journal/559/55969711016/55969711016.pdf>
- Masip, J. (2017). Deception detection: State of the art and future prospects. *Psicothema*, 29(2), pp. 149-159. <https://www.psicothema.com/pdf/4376.pdf>
- Masip, J. & Garrido, E. (2000). La Evaluación de la credibilidad del testimonio en Contextos Judiciales a Partir de Indicadores Conductuales: *Anuario de Psicología Jurídica*, 10, 93-132. <https://journals.copmadrid.org/apj/archivos/64159.pdf>
- Masip, J. & Herrero, C. (2015). Nuevas aproximaciones en detección de mentiras I. Antecedentes y marco teórico. *Papeles del Psicólogo*, 36(2), 83-95.  
<https://www.papelesdelpsicologo.es/pdf/2564.pdf>

- Masip, J., Blandón, I., Martínez, C., Herrero, C. e Ibabe, I. (2016). Strategic interviewing to detect deception: Cues to deception across repeated interviews. *Frontiers in Psychology*, 7, 1702. <https://doi.org/10.3389/fpsyg.2016.01702>
- Masip, J., Garrido, E. & Herrero, C. (2002). La detección de la mentira mediante la técnica SCAN. *Psicopatología Clínica Legal y Forense*, 2(2), 39-62. <https://www.masterforense.com/pdf/2002/2002art10.pdf>
- Masip, J., Garrido, E., & Herrero, C. (2004). Defining deception. *Annals of Psychology*, 20(1), 147–172. <https://revistas.um.es/analesps/article/view/27631>
- Nahari, G. (2018). Reality monitoring in the forensic context: Digging deeper into the speech of liars. *Journal Of Applied Research In Memory And Cognition*, 7(3), 432-440. <https://www.sciencedirect.com/science/article/pii/S2211368117301286>
- Nortje, A., & Tredoux, C. (2019). How good are we at detecting deception? A review of current techniques and theories. *South African Journal Of Psychology*, 49(4), 491-504. <https://doi.org/10.1177/0081246318822953>
- Omil, J. (2019). Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde? *Mercados y Negocios*, 40, 5-22. <https://doi.org/10.32870/myn.v0i40.7403>
- Palacios P. (25 de agosto de 2022). *Psicoterapia, mentira y autoengaño*. Fundación Instituto Spiral. <https://fispiral.com.es/psicoterapia-mentira-autoengano/>
- Pallejá, J., Martín, E., & Alonso, M. (2002). La detección del engaño sobre la base de sus correlatos conductuales: la precisión de los juicios. *Anuario de Psicología Jurídica*, 12(12), 37-56. <https://journals.copmadrid.org/apj/archivos/80162.pdf>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog.

[https://cdn.openai.com/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf)

- Rassin, E. (2000). Criteria based content analysis: The less scientific road to truth. *Expert Evidence*, 7, 265-278. <https://link.springer.com/article/10.1023/a:1016627527082>
- Rodríguez, A. (2020). El impacto de la inteligencia artificial en el proceso penal. *Anuario de la Facultad de Derecho Universidad de Extremadura*, 36, 695-728. <https://doi.org/10.17398/2695-7728.36.695>
- Romo, V., Garcia, J., Özdemir, A., & Leiros, R. (2023). ChatGPT ha llegado ¿Y ahora qué hacemos? La creatividad, nuestro último refugio. *Revista de Investigación En Educación*, 21(3), 320-334. <https://doi.org/10.35869/reined.v21i3.4973>
- Salah, M., Halbusi, H., & Abdelfattah, F. (2023). May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research. *Computers In Human Behavior Artificial Humans*, 1(2), 100006. <https://doi.org/10.1016/j.chbah.2023.100006>
- Simons, S., Garrison, R. & Johnson, K. (2017). Brain Mechanisms of Reality Monitoring. *Trends in cognitive sciences*, 21(6), 462–473. <https://pubmed.ncbi.nlm.nih.gov/28462815/>
- Soto, V., Hernández, R. & Estéban, L. (2013). Valoración de la credibilidad del testimonio: aplicación del modelo Reality Monitoring. *Revista internacional de Psicología*, 12(2), 1. <https://www.revistapsicologia.org/index.php/revista/article/view/68/65>
- Sporer, L., & Sharman, J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied*

*Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 20(6), 837-854. <https://doi.org/10.1002/acp.1234>

Tan, S., Jia, Y., Jariwala, N., Zhang, Z., Brent, K., Houde, J., Nagarajan, S. & Subramaniam, K. (2024). A randomised controlled trial investigating the causal role of the medial prefrontal cortex in mediating self-agency during speech monitoring and reality monitoring. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-55275-3>

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*. [https://www.researchgate.net/publication/226793702\\_A\\_Vrij\\_Detecting\\_Lies\\_and\\_Deceit\\_The\\_Psychology\\_of\\_Lying\\_and\\_Implications\\_forProfessional\\_Practice](https://www.researchgate.net/publication/226793702_A_Vrij_Detecting_Lies_and_Deceit_The_Psychology_of_Lying_and_Implications_forProfessional_Practice)

Vrij, A., Akehurst, L., Soukara, S. & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human communication research*, 30(1), 8-41. <https://doi.org/10.1111/j.1468-2958.2004.tb00723.x>

Vrij, A., Granhag, P., & Porter, S. (2010). Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection. *Psychological Science In The Public Interest*, 11(3), 89-121. <https://doi.org/10.1177/1529100610390861>

Walczyk, J., Harris, L., Duck, T., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas In Psychology*, 34, 22-36. <https://www.sciencedirect.com/science/article/pii/S0732118X14000142?via%3Dihub>

[b](#)