

**PoC:** Identificación de patrones de desinformación en medios de comunicación

**0:00–0:20 — Introducción (ambos) (<- ESTO NO SE LEE!!)**

**1. Octavio:**

- Hola, soy Octavio!!!

**2. Luisa Fernanda:**

- Y yo soy Luisa Fernanda!!! Y ambos somos estudiante de Ingeniería en Ciencia de Datos en la Universidad Iberoamericana de Ciudad de México.

**3. Octavio:**

- Ambos participamos en el desarrollo del proyecto de investigación "Identificación de patrones de desinformación en medios de comunicación".

**0:20–0:45 — El problema 4. Luisa Fernanda:** - Hoy en día, las redes sociales son el principal medio por el cual muchas personas se informan... pero también son la vía más rápida para difundir desinformación.

**5. Octavio:**

- Las noticias falsas generan confusión, polarización social y decisiones mal informadas. Es un problema global que afecta la democracia y la cohesión social.

**6. Luisa Fernanda:**

- Si se considera la política seguida por X (anteriormente Twitter), donde la moderación de contenidos dejó de ser prioridad, la detección de patrones de desinformación cobra aún más relevancia.

**0:45–1:10 — La solución 7. Octavio:** - Nuestra propuesta utiliza inteligencia artificial para detectar e identificar patrones lingüísticos, psicológicos y visuales que permiten identificar contenido engañoso.

**8. Luisa Fernanda:**



- Integramos múltiples enfoques: análisis de texto con procesamiento de lenguaje natural, verificación automática mediante grafos de conocimiento, y análisis multimodal que combina texto e imagen.

**1:10–1:30 — Tecnología y ética 9. Luisa Fernanda:** - Usamos machine learning, deep learning y modelos explicables para que el sistema no solo funcione, sino que las personas puedan entender cómo y por qué clasifica algo como desinformación. Cuando las personas entienden el por qué de las decisiones automatizadas, son más propensas a tomar decisiones correctas.

**10. Octavio:**

- Además, el sistema fue diseñado considerando principios de IA responsable: privacidad, equidad, inclusión, transparencia y seguridad.

**1:30–1:50 — Impacto 11. Octavio:** - Este proyecto es útil para periodistas, educadores, medios de comunicación y plataformas sociales que quieren combatir la desinformación. Sobre todo, es especialmente útil para las personas ya que les permitirá discernir entre información verídica y desinformación.

**12. Luisa Fernanda:**

- Queremos empoderar a las personas con herramientas que promuevan el pensamiento crítico y la confianza en la información verificada.

**1:50–2:00 — Cierre 13. Luisa Fernanda:** - Una sociedad informada es una sociedad más libre.

**14.- Octavio:** - La verdad nos hará libres!!!

## **Nivel de Madurez: TRL 2-3**

Buscamos que, a través de este documento podamos entender ¿Qué hace el proyecto?, ¿Por qué es valioso?, ¿Cómo lo hace?, ¿A quién impacta? Podrás encontrar algunas preguntas guía que te pueden ayudar a la redacción de cada sección en caso de que apliquen. Favor de contestar cada sección a modo de párrafo (no contestar a modo de cuestionario), puedes apoyarte con gráficos.

### **1. Generalidades**



## 1.1 Generalidades del proyecto

### 1. *¿Cuál es alcance e impacto de tu proyecto?*

El proyecto "FakeNews" tiene un alcance e impacto significativo en la lucha contra la desinformación, que se ha convertido en un problema crítico en el entorno digital actual. Su objetivo principal es desarrollar métodos basados en inteligencia artificial para identificar y mitigar la desinformación en mensajes electrónicos, utilizando técnicas que analicen patrones lingüísticos y psicológicos. Este enfoque interdisciplinario involucra diversas disciplinas como informática, psicología, lingüística y comunicación, lo que permite una comprensión más amplia de la desinformación y su propagación. Además, el proyecto busca mejorar la detección de noticias falsas a través de una clasificación detallada de las mismas, que incluye categorías como noticias engañosas y rumores. El impacto del proyecto se manifiesta en la mejora de la confianza pública y la estabilidad social al abordar los efectos negativos de la desinformación, que puede generar decisiones erróneas y divisiones sociales. Al desarrollar soluciones comprensibles desde la perspectiva de la explicabilidad, el proyecto pretende hacer que las tecnologías de detección sean accesibles y útiles para diferentes sectores de la sociedad. En resumen, "FakeNews" no solo busca identificar y clasificar la desinformación, sino también contribuir al desarrollo sostenible al mitigar sus efectos adversos en el funcionamiento social.

### 2. *¿Cómo funciona?*

Este proyecto funciona a través de un enfoque multidisciplinario que integra diversas técnicas de inteligencia artificial para identificar y mitigar la desinformación en mensajes electrónicos. Su metodología se basa en la extracción y análisis automático de datos noticiosos, utilizando patrones lingüísticos y psicológicos para detectar noticias falsas. Este proceso implica varias etapas y enfoques: primero, se clasifican las noticias falsas en diferentes categorías, como noticias engañosas o rumores, lo que permite un análisis más específico. La detección se realiza mediante cuatro métodos principales: el enfoque basado en contenido, que examina las características semánticas de los textos; el enfoque basado en redes sociales, que utiliza datos sobre el comportamiento de los usuarios y las rutas de propagación de la información; el enfoque basado en conocimiento, que verifica la autenticidad de las noticias a través de bases de datos y grafos de conocimiento; y el enfoque multimodal, que combina texto e imágenes para mejorar la identificación de contenido engañoso. Cada uno de estos métodos tiene sus propias técnicas y herramientas. Por ejemplo, el análisis del contenido se centra en identificar diferencias lingüísticas entre textos verdaderos y falsos, mientras que el análisis de redes sociales busca patrones en cómo se comparten y propagan las noticias. La verificación automática utiliza procesamiento de lenguaje natural para comparar noticias con hechos conocidos, mientras que el enfoque multimodal considera tanto el texto como los elementos visuales para una detección más robusta. En conjunto, estas estrategias permiten al proyecto abordar la desinformación de manera integral, mejorando la capacidad para identificar noticias falsas y contribuyendo a la restauración de la confianza pública en la información.

### 3. *¿Es un producto o servicio?*

El proyecto "FakeNews" se presenta como un **producto de investigación** más que como un servicio comercial. Su enfoque principal es el desarrollo de métodos y tecnologías basadas en inteligencia artificial para la identificación y mitigación de la desinformación en mensajes electrónicos. Aunque no se detalla un modelo de negocio específico, el objetivo es crear soluciones comprensibles y efectivas que puedan ser aplicadas en diversos contextos, como redes sociales y plataformas de noticias, para combatir la propagación de noticias falsas. El proyecto se centra en la creación de un sistema que utiliza técnicas avanzadas de análisis de datos, incluyendo el procesamiento del lenguaje natural y el aprendizaje automático, para detectar patrones lingüísticos y psicológicos asociados con la desinformación. Esto implica la elaboración de herramientas que podrían ser utilizadas por investigadores, periodistas, plataformas de medios y educadores.



# Acelerando México con IA



intel

para mejorar la calidad de la información y fomentar una mayor confianza pública. En resumen, "FakeNews" es un esfuerzo investigativo que busca generar un impacto significativo en la forma en que se identifica y maneja la desinformación, pero no se define como un servicio comercial específico destinado a usuarios finales.

#### 4. *¿Quiénes son los usuarios potenciales y por qué este proyecto es de valor para ellos?*

Los usuarios potenciales abarcan una amplia variedad de grupos que se beneficiarían significativamente de las soluciones propuestas para la detección de desinformación. En primer lugar, **periodistas y medios de comunicación** son usuarios clave, ya que necesitan herramientas efectivas para verificar la autenticidad de la información antes de su publicación, lo que les permitiría mantener la integridad y la confianza del público en sus reportes. Además, **plataformas de redes sociales** podrían utilizar estas tecnologías para identificar y mitigar la propagación de contenido engañoso en sus plataformas, contribuyendo así a un entorno informativo más saludable. Otro grupo importante son **educadores y académicos**, quienes podrían emplear los métodos desarrollados en el proyecto para enseñar a los estudiantes sobre la identificación de noticias falsas y fomentar habilidades críticas en el consumo de información. Asimismo, **instituciones gubernamentales y organizaciones no gubernamentales** interesadas en la promoción de la alfabetización mediática y la lucha contra la desinformación también se beneficiarían al implementar estas tecnologías en sus programas y políticas. El valor del proyecto radica en su capacidad para ofrecer soluciones comprensibles y accesibles que no solo identifican patrones lingüísticos y psicológicos asociados con la desinformación, sino que también permiten a los usuarios comprender mejor el fenómeno de las noticias falsas. Esto es crucial en un contexto donde la desinformación puede tener efectos perjudiciales en la confianza pública, las decisiones sociales y políticas, así como en el desarrollo económico sostenible. En resumen, "FakeNews" proporciona herramientas valiosas para una amplia gama de usuarios que buscan combatir la desinformación y promover un entorno informativo más veraz y confiable.

#### 5. *Haz énfasis en la profundidad, escalabilidad y duración del impacto.*

La propuesta tiene un enfoque profundo, escalable y con un impacto duradero en la lucha contra la desinformación. Su profundidad se manifiesta en la integración de múltiples disciplinas, como la informática, psicología, lingüística y neurociencia cognitiva, lo que permite un análisis exhaustivo de los mecanismos que propagan la desinformación. Al utilizar técnicas avanzadas de inteligencia artificial, el proyecto no solo busca identificar patrones lingüísticos y psicológicos asociados con noticias falsas, sino que también se esfuerza por desarrollar soluciones comprensibles que faciliten su aplicación en diversos contextos. En términos de escalabilidad, el sistema propuesto está diseñado para adaptarse a diferentes plataformas y entornos informativos. Esto significa que las herramientas desarrolladas pueden ser implementadas tanto en redes sociales como en medios de comunicación tradicionales, lo que permite una amplia difusión y aplicación en diversas comunidades. La capacidad de abordar la desinformación desde múltiples ángulos—incluyendo el análisis del contenido, el comportamiento en redes sociales y la verificación de hechos—hace que el proyecto sea flexible y capaz de evolucionar con las necesidades cambiantes del entorno digital. Finalmente, el impacto del proyecto se espera que sea duradero. Al proporcionar herramientas efectivas para la detección de noticias falsas, se busca restaurar la confianza pública en la información y mitigar las tensiones sociales que surgen de la desinformación. Este enfoque no solo tiene el potencial de mejorar la calidad del contenido informativo a corto plazo, sino que también contribuye al desarrollo sostenible a largo plazo al fomentar una ciudadanía más informada y crítica. En resumen, "FakeNews" representa una respuesta integral a un problema contemporáneo urgente, con implicaciones significativas para la estabilidad social y económica en el futuro.

#### 6. *Comparte tu reflexión sobre el potencial impacto de este proyecto en la sociedad en general, en la economía, en tu región, en el país y/o global.*

El proyecto "FakeNews" tiene el potencial de generar un impacto significativo en la sociedad, la economía y a nivel global. En el ámbito



# Acelerando México con IA



social, su implementación puede contribuir a restaurar la confianza pública en la información, un aspecto crucial en tiempos de crisis y polarización. Al desarrollar métodos efectivos para detectar y mitigar la desinformación, se espera que las comunidades se vuelvan más resilientes frente a las noticias falsas, promoviendo un diálogo más informado y constructivo entre los ciudadanos. Desde una perspectiva económica, la desinformación puede afectar negativamente las decisiones de consumo y las dinámicas del mercado. Al reducir la propagación de noticias falsas, el proyecto podría ayudar a estabilizar mercados y fomentar un entorno más predecible para los negocios, lo que a su vez podría impulsar el desarrollo económico sostenible. En el contexto regional y nacional, la capacidad de identificar y abordar la desinformación puede fortalecer la cohesión social y mejorar la calidad del debate público, lo que es esencial para el funcionamiento democrático. A nivel global, el fenómeno de la desinformación trasciende fronteras y afecta a múltiples sociedades. El proyecto "FakeNews" tiene el potencial de ser una referencia en la lucha contra este problema a escala internacional, promoviendo colaboraciones entre diferentes países y organizaciones para compartir mejores prácticas y tecnologías. Esto no solo contribuiría a un entorno informativo más saludable en diversas naciones, sino que también podría establecer estándares globales para la detección y gestión de la desinformación. En resumen, el impacto del proyecto "FakeNews" podría ser profundo y duradero, abarcando mejoras en la confianza social, estabilidad económica y cooperación internacional en la lucha contra un fenómeno que amenaza el desarrollo sostenible de las sociedades contemporáneas.

## 1.2 Oportunidad Detectada

### 1. *Idea general del proyecto.*

La iniciativa se centra en la urgente necesidad de abordar la desinformación, un fenómeno que ha crecido exponencialmente con el auge de las tecnologías digitales y las redes sociales. En un mundo donde la información se propaga a una velocidad sin precedentes, el proyecto busca desarrollar métodos innovadores basados en inteligencia artificial para identificar y mitigar la difusión de noticias falsas. La esencia del proyecto radica en la creación de un sistema que no solo detecte contenido engañoso, sino que también sea comprensible para los usuarios. Esto es crucial, ya que muchas soluciones actuales en inteligencia artificial carecen de explicabilidad, lo que dificulta su adopción y confianza por parte del público. A través de un enfoque multidisciplinario, el proyecto integra conocimientos de áreas como la psicología, la lingüística y la comunicación para analizar cómo se propaga la desinformación y cómo se puede contrarrestar efectivamente. Además, el proyecto reconoce que la desinformación no solo afecta a individuos, sino que tiene repercusiones más amplias en la cohesión social y el funcionamiento democrático. Al abordar este problema desde múltiples ángulos—incluyendo análisis lingüísticos, comportamientos en redes sociales y verificación de hechos—se busca crear herramientas robustas que puedan ser utilizadas por periodistas, plataformas de redes sociales y educadores para promover un entorno informativo más saludable. En definitiva, esta iniciativa representa un paso crucial hacia la construcción de una sociedad más informada y resiliente frente a los desafíos de la desinformación.

### 2. *¿Qué necesidad satisface? Planteamiento de la oportunidad que este proyecto resuelve.*

La necesidad que satisface este proyecto radica en la creciente problemática de la desinformación, que se ha convertido en un fenómeno omnipresente en la era digital. Con la rápida proliferación de la tecnología y el uso generalizado de las redes sociales, la propagación de información falsa ha alcanzado niveles alarmantes, generando confusión, malentendidos y tensiones sociales. Este contexto ha llevado a una erosión de la confianza pública en las fuentes de información, lo que impacta negativamente en la cohesión social y el funcionamiento democrático. El proyecto aborda esta oportunidad al desarrollar métodos innovadores basados en inteligencia artificial para identificar y mitigar la desinformación en mensajes electrónicos. A través del análisis automático de datos noticiosos, se busca no solo detectar





# Acelerando México con IA



noticias falsas, sino también proporcionar soluciones comprensibles que permitan a los usuarios entender cómo se identifican y clasifican estas informaciones engañosas. Esto es crucial, ya que muchas de las soluciones actuales carecen de explicabilidad, lo que dificulta su aceptación y uso efectivo. Además, el proyecto reconoce que la desinformación tiende a intensificarse en momentos críticos, como crisis sanitarias o campañas electorales, donde la necesidad de información veraz es más apremiante. Al ofrecer herramientas que integran enfoques multidisciplinarios—incluyendo psicología, lingüística y comunicación—se busca abordar el problema desde múltiples ángulos, mejorando así la capacidad para detectar patrones lingüísticos y psicológicos asociados con la desinformación.

### 3. *¿Hay evidencia de la existencia de esta necesidad?*

La necesidad de abordar la desinformación se fundamenta en una serie de datos concretos que evidencian su impacto negativo en la sociedad. La proliferación de noticias falsas ha sido documentada extensamente, mostrando que estas se propagan más rápido y más lejos que la información verificada. Por ejemplo, durante la pandemia de COVID-19, estudios revelaron que los rumores y la información errónea sobre el virus se difundieron a una velocidad alarmante, generando confusión y ansiedad en la población. Investigaciones han demostrado que los momentos críticos, como crisis sanitarias y campañas electorales, son contextos propicios para la intensificación de la desinformación. En estos períodos, las personas tienden a buscar información rápidamente, lo que las hace más susceptibles a compartir contenido engañoso. Un estudio de Bordia et al. indica que la necesidad psicológica de encontrar "verdades" en situaciones inciertas impulsa a las personas a aceptar y difundir información falsa. Además, el impacto de la desinformación no solo se limita a la esfera individual; afecta el funcionamiento normal de la sociedad y puede llevar a decisiones erróneas a nivel colectivo. Por ejemplo, en el ámbito político, la desinformación puede influir en elecciones al distorsionar la percepción pública sobre candidatos o políticas. La investigación de Pennycook et al. destaca cómo las motivaciones políticas pueden afectar las creencias de las personas, llevándolas a aceptar noticias falsas que refuercen sus puntos de vista. Desde un punto de vista académico, el enfoque multidisciplinario en el estudio de la desinformación ha crecido significativamente. La combinación de conocimientos en informática, psicología, lingüística y comunicación permite abordar el fenómeno desde múltiples ángulos. Por ejemplo, Choudhary et al. han utilizado análisis lingüísticos para identificar características específicas del vocabulario y la gramática que son comunes en las noticias falsas, lo que proporciona una base sólida para desarrollar métodos automatizados de detección. La categorización de las noticias falsas también ha sido objeto de estudio, con clasificaciones que incluyen noticias engañosas, información rumorística y propaganda deliberada. Según investigaciones recientes, estas categorías ayudan a entender mejor cómo se difunde la desinformación y qué estrategias pueden ser más efectivas para contrarrestarla. Finalmente, el avance en tecnologías como el procesamiento del lenguaje natural y el aprendizaje automático ha abierto nuevas oportunidades para mejorar la detección de noticias falsas. Métodos como la verificación automática mediante grafos de conocimiento han demostrado ser efectivos para juzgar la autenticidad de las noticias al comparar su contenido con hechos verificados.

### 4. *Propósito, la razón por la cual existe el proyecto*

*El propósito del proyecto radica en la necesidad urgente de abordar la creciente problemática de la desinformación, que ha proliferado con el avance de la tecnología digital y el uso generalizado de las redes sociales. Este fenómeno no solo afecta la percepción pública y la confianza en las instituciones, sino que también tiene repercusiones significativas en la estabilidad social y el desarrollo económico sostenible. La difusión de información falsa puede generar confusión, malentendidos y tensiones sociales, lo que impacta negativamente en la cohesión y el funcionamiento adecuado de las comunidades. El proyecto busca desarrollar métodos innovadores basados en inteligencia artificial para identificar y mitigar la desinformación en mensajes electrónicos. A través del análisis automático de datos noticiosos, se pretende crear un sistema que no solo sea efectivo en la detección de noticias falsas, sino que también sea comprensible y accesible para los usuarios. Esto es crucial, ya que muchas soluciones actuales carecen de explicabilidad, lo que dificulta su adopción y confianza por parte del público. Además, el proyecto se fundamenta en la premisa de que existen momentos críticos, como crisis sanitarias o campañas electorales, donde la desinformación tiende a intensificarse. En estos contextos, el riesgo de decisiones erróneas aumenta, lo*



# Acelerando México con IA



*que subraya la necesidad de contar con herramientas efectivas para detectar y contrarrestar la propagación de noticias falsas. La investigación multidisciplinaria que se propone involucra campos como la psicología, lingüística y comunicación, lo cual permite un enfoque integral para entender los mecanismos detrás de la desinformación y desarrollar estrategias más efectivas para su detección.*

## 1.3 Generalidades Técnicas

### 1. ¿Qué componentes utilizas? (HW, SW, FW, Materiales, Procesos industriales, Programas)

El proyecto se apoya en una variedad de componentes técnicos que son esenciales para su funcionamiento efectivo. En términos de **hardware**, se utilizarán servidores y unidades de procesamiento gráfico (GPU) para manejar el procesamiento intensivo de datos y el entrenamiento de modelos de inteligencia artificial. Estos recursos son cruciales para ejecutar algoritmos complejos que analizan grandes volúmenes de información en tiempo real. En cuanto al **software**, se emplearán herramientas y bibliotecas específicas para el desarrollo de modelos de aprendizaje automático y procesamiento del lenguaje natural. Esto incluye frameworks como TensorFlow o PyTorch, que permiten la creación y entrenamiento de redes neuronales, así como bibliotecas como NLTK o spaCy para el análisis lingüístico. Además, se utilizarán plataformas de gestión de datos como Apache Hadoop o Spark para el manejo eficiente de grandes conjuntos de datos. En cuanto a **materiales**, se requerirán conjuntos de datos relevantes que contengan ejemplos de noticias verdaderas y falsas, así como metadatos asociados que ayuden a entrenar los modelos. Estos conjuntos pueden provenir de diversas fuentes, incluyendo redes sociales, sitios web de noticias y bases de datos académicas. Los **procesos industriales** involucrados incluyen técnicas avanzadas de minería de datos, análisis semántico y modelado predictivo. Estos procesos permitirán extraer características significativas del contenido textual y visual, facilitando la identificación de patrones que pueden indicar desinformación. Finalmente, los **programas** implementados abarcarán desde algoritmos específicos para la detección de noticias falsas hasta sistemas más amplios que integren diferentes enfoques—como análisis basado en contenido, redes sociales y verificación automática—para ofrecer una solución robusta y comprensiva en la lucha contra la desinformación. Todo esto se unirá en un marco que permita una evaluación continua y mejora del sistema a medida que se recopilan más datos y se desarrollan nuevas técnicas.

### 2. ¿Cuál es la solución de diseño teórico? O sea, ¿cómo es tu proyecto en caso de ser un producto tangible?

La solución de diseño teórico del proyecto se concibe como una plataforma integral para la detección de desinformación, que combina múltiples tecnologías y enfoques metodológicos en un producto tangible. Esta plataforma incluiría una interfaz de usuario intuitiva, accesible tanto en formato web como móvil, permitiendo a los usuarios ingresar noticias o enlaces para su análisis. En el núcleo del sistema, se integrarían varios módulos especializados: el módulo de análisis de contenido utilizaría técnicas de procesamiento del lenguaje natural para extraer características semánticas y lingüísticas de los textos, identificando patrones que suelen estar presentes en las noticias falsas. Simultáneamente, un módulo de redes sociales analizaría el comportamiento del contenido en diversas plataformas, evaluando cómo se comparte y propaga la información, lo que ayudaría a identificar tendencias en la difusión de desinformación. Además, un módulo de verificación automática utilizaría grafos de conocimiento para comparar las afirmaciones contenidas en las noticias con hechos verificados, proporcionando un puntaje de veracidad. Para complementar estos enfoques, un módulo multimodal integraría tanto datos textuales como visuales, analizando imágenes y videos asociados con las noticias para detectar contenido engañoso. Este sistema no solo se diseñaría para ser efectivo en la identificación de desinformación, sino también para ser comprensible y accesible para los usuarios, facilitando así su adopción y uso. El contexto en el que se desarrolla esta plataforma es crucial, dado que la desinformación tiende a proliferar en momentos críticos como crisis sanitarias o campañas electorales, lo que hace que la necesidad de herramientas efectivas sea más apremiante. En resumen, esta solución teórica propone un enfoque integral que



# Acelerando México con IA

combina diversas disciplinas y tecnologías para abordar el problema de la desinformación de manera efectiva y comprensible.

3. *Comparte los procesos utilizados para generar esta PoC (Scrum, Resolución estructurada de problemas, trabajo en equipo, etc.) que valga la pena resaltar, puedes apoyarte de diagramas de flujo para representarlos.*

Para la generación de la prueba de concepto (PoC) del proyecto, se ha implementado la metodología **CRISP-DM** (Cross-Industry Standard Process for Data Mining), que proporciona un marco estructurado para el desarrollo de proyectos de minería de datos y análisis. Esta metodología es especialmente adecuada para el contexto del proyecto, dado que se centra en la comprensión del problema, la preparación de los datos y la implementación de modelos analíticos. El proceso comienza con la **definición del negocio**, donde se identifican los objetivos específicos del proyecto en relación con la detección de desinformación. Esto implica comprender las necesidades de los usuarios y cómo el sistema puede ofrecer soluciones efectivas. A continuación, se lleva a cabo la **comprensión de los datos**, que incluye la recolección y exploración de conjuntos de datos relevantes, como ejemplos de noticias verdaderas y falsas, así como metadatos que puedan ayudar a entrenar los modelos. La siguiente etapa es la **preparación de los datos**, que implica limpiar y transformar los datos para que sean aptos para el análisis. Esto puede incluir la normalización del texto, la eliminación de duplicados y el etiquetado de datos según su veracidad. Posteriormente, se avanza a la **modelización**, donde se seleccionan y aplican algoritmos de aprendizaje automático adecuados para detectar patrones en los datos. Esta fase es crucial, ya que diferentes enfoques (como análisis basado en contenido, redes sociales o verificación automática) pueden ser implementados y evaluados. Una vez que se han desarrollado los modelos, se procede a la **evaluación**, donde se analizan los resultados obtenidos para determinar si cumplen con los objetivos establecidos al inicio del proyecto. Esto incluye métricas como precisión, recall y F1-score, que son esenciales para medir la efectividad del sistema en la detección de desinformación. Finalmente, en la etapa de **despliegue**, se implementa el sistema en un entorno real donde los usuarios pueden interactuar con él. Esta fase también contempla el monitoreo continuo del rendimiento del modelo y su ajuste según sea necesario, lo que permite una mejora constante basada en nuevos datos y feedback. La aplicación de CRISP-DM no solo proporciona una estructura clara y organizada para el desarrollo del proyecto, sino que también fomenta un enfoque iterativo que permite ajustes continuos a medida que se avanza en el proceso. Este enfoque es especialmente valioso en un campo tan dinámico como el análisis de desinformación, donde las tendencias y técnicas pueden evolucionar rápidamente. La combinación de metodologías ágiles como Scrum con CRISP-DM asegura que el equipo esté alineado y adaptándose constantemente a las necesidades cambiantes del proyecto y del entorno informativo.

4. *Habiendo identificado el TRL con el que se categoriza el proyecto, y de acuerdo con la categoría que se autoasignó al inscribirte en la convocatoria, **llena la tabla del punto 2. Escenario de Pruebas Según TRL***

5. *Expón los 3 siguientes pasos que requerirías realizar para lograr avanzar tu proyecto al siguiente nivel de TRL.*

Para avanzar el proyecto al siguiente nivel de TRL (Technology Readiness Level), se requieren tres pasos clave que son fundamentales para la validación y escalabilidad del sistema de detección de desinformación. El primer paso es la **realización de pruebas piloto en entornos controlados**. Esto implica implementar el sistema en un entorno de prueba donde se puedan recopilar datos reales sobre su desempeño. Durante esta fase, se evaluará la precisión del sistema en la identificación de noticias falsas en comparación con fuentes verificadas. Se realizarán ajustes basados en los resultados obtenidos, lo que permitirá optimizar los algoritmos de detección y mejorar la efectividad del modelo. La retroalimentación de los usuarios durante estas pruebas también será crucial para identificar áreas de mejora y asegurar que el sistema sea comprensible y accesible. El segundo paso consiste en la **integración de múltiples fuentes de datos y colaboración con plataformas sociales y medios de comunicación**. Esto permitirá enriquecer el conjunto de datos utilizado





# Acelerando México con IA



para entrenar el modelo, asegurando que se incluya una variedad representativa de noticias verdaderas y falsas. Además, establecer alianzas con plataformas sociales facilitará el acceso a datos sobre cómo se comparte y propaga la información, lo que es vital para mejorar la detección basada en redes sociales. Esta colaboración también puede ayudar a implementar medidas preventivas contra la difusión de desinformación en tiempo real. El tercer paso es **el desarrollo de un plan de escalabilidad y comercialización**. Una vez que el sistema haya sido validado a través de pruebas piloto, es esencial diseñar una estrategia para llevarlo al mercado. Esto incluye definir el modelo de negocio, identificar a los usuarios finales (como periodistas, plataformas sociales y educadores) y establecer canales de distribución. También se debe considerar la creación de materiales educativos y recursos que faciliten la adopción del sistema por parte de los usuarios, asegurando así que las herramientas desarrolladas sean utilizadas efectivamente para combatir la desinformación. Estos pasos son fundamentales para avanzar el proyecto al siguiente nivel de TRL, asegurando no solo su viabilidad técnica, sino también su relevancia y aplicación práctica en un entorno real.

## 1.4 Innovación con IA

### 1. ¿Cuál es el impacto de usar AI para este proyecto?

El uso de inteligencia artificial (IA) en el proyecto de detección de desinformación tiene un impacto significativo y multifacético que puede transformar la manera en que se aborda este problema crítico en la sociedad actual. En primer lugar, la IA permite un **análisis automatizado y a gran escala** de grandes volúmenes de datos, lo que es esencial dado el ritmo acelerado con el que se propaga la desinformación, especialmente en plataformas digitales y redes sociales. La capacidad de procesar y analizar datos en tiempo real facilita la identificación rápida de noticias falsas, lo que puede ayudar a mitigar su difusión antes de que causen un daño significativo. Además, la IA puede mejorar la **precisión en la detección** de patrones lingüísticos y psicológicos asociados con la desinformación. Utilizando técnicas avanzadas de procesamiento del lenguaje natural (NLP), los modelos de IA pueden identificar características específicas que diferencian las noticias falsas de las verdaderas, como el uso de un lenguaje más subjetivo o exagerado. Esto se complementa con enfoques multimodales que integran texto e imágenes, permitiendo una evaluación más completa del contenido noticioso. Por ejemplo, estudios han demostrado que las noticias falsas a menudo utilizan elementos visuales engañosos para atraer a los lectores, y la IA puede ayudar a detectar estas inconsistencias. Otro aspecto crucial es la **explicabilidad** que se busca en el uso de IA. Aunque muchos modelos actuales son altamente efectivos, a menudo carecen de transparencia en cómo toman decisiones. El desarrollo de sistemas basados en IA que sean comprensibles para los usuarios es vital para fomentar la confianza en la tecnología. Esto implica no solo identificar desinformación, sino también proporcionar explicaciones claras sobre cómo se llegó a esa conclusión, lo cual es esencial para su aceptación y uso por parte del público. La inteligencia artificial también permite un enfoque **interdisciplinario**, integrando conocimientos de diversas áreas como psicología, lingüística y comunicación. Esta colaboración puede enriquecer el análisis y ofrecer soluciones más robustas. Por ejemplo, al combinar datos sobre el comportamiento del usuario en redes sociales con análisis lingüísticos, se pueden identificar patrones más complejos que contribuyen a la propagación de desinformación. Finalmente, el uso de IA en este contexto no solo aborda el problema inmediato de la desinformación, sino que también tiene implicaciones más amplias para la **estabilidad social y**



# Acelerando México con IA



intel

**económica.** La desinformación puede generar divisiones sociales y decisiones erróneas que afectan el funcionamiento democrático y el desarrollo sostenible. Al implementar tecnologías avanzadas para su detección, se busca restaurar la confianza pública en las fuentes de información y fomentar un entorno informativo más saludable.

2. *Proporciona datos específicos de qué elementos de IA se utilizan: Machine Learning, Deep Learning, Generative AI, etc. Indica con qué profundidad / amplitud (básico, avanzado, intermedio).*

El proyecto de detección de desinformación utiliza varios elementos de inteligencia artificial, cada uno con un nivel de profundidad y amplitud específico. En primer lugar, se emplea **Machine Learning** (aprendizaje automático) en un nivel **intermedio a avanzado**. Este enfoque es fundamental para desarrollar modelos que puedan identificar patrones en los datos, permitiendo la clasificación de noticias como verdaderas o falsas. Los algoritmos de Machine Learning se utilizan para analizar características lingüísticas y semánticas del contenido, así como para evaluar el comportamiento en redes sociales. Dentro del ámbito del Machine Learning, se implementan técnicas de **Deep Learning** (aprendizaje profundo), que son más avanzadas y se utilizan para tareas complejas como el procesamiento del lenguaje natural (NLP). Estos modelos, como las redes neuronales recurrentes (RNN) y las arquitecturas Transformer, permiten una comprensión más profunda del texto y son capaces de captar matices que los métodos más simples no pueden. La profundidad de este enfoque es **avanzada**, ya que implica la creación y entrenamiento de modelos complejos que requieren grandes volúmenes de datos para su efectividad. Además, el proyecto también considera el uso de **Generative AI** (IA generativa) en un contexto específico. Aunque no es el enfoque principal, esta tecnología puede ser utilizada para crear ejemplos sintéticos de noticias falsas que ayuden a entrenar los modelos de detección. Esto es especialmente útil en situaciones donde los datos reales son limitados o sesgados. El uso de IA generativa se puede clasificar como **básico a intermedio**, dependiendo del nivel de sofisticación requerido en la generación de contenido. Por último, se integra un enfoque **multimodal**, que combina diferentes tipos de datos (texto e imágenes) para mejorar la detección. Este enfoque permite una evaluación más completa del contenido noticioso y se considera **intermedio**, ya que requiere la fusión de diferentes fuentes de datos y técnicas analíticas.

3. *Consideraciones éticas del uso de IA (si aplica) a lo largo de las diferentes etapas del proyecto: Privacidad, Sesgo, Discriminación, Injusticia, Sostenibilidad Ambiental, Transparencia, Seguridad, Prevención que asegure el bienestar de los usuarios (Safety)*

El uso de inteligencia artificial en el proyecto de detección de desinformación plantea varias consideraciones éticas que deben ser abordadas a lo largo de sus diferentes etapas. Estas consideraciones incluyen aspectos cruciales como la privacidad, el sesgo, la discriminación, la transparencia y la seguridad. En primer lugar, la **privacidad** es un tema fundamental, ya que el sistema utilizará datos de usuarios y contenido de redes sociales. Es esencial garantizar que se cumplan las normativas de protección de datos, como el GDPR en Europa, para asegurar que la información personal no sea recopilada ni utilizada sin el consentimiento explícito de los usuarios. Esto implica implementar medidas de anonimización y encriptación para proteger la identidad de los individuos. El **sesgo** es otra preocupación significativa. Los modelos de IA pueden heredar sesgos presentes en los datos con los que son entrenados, lo que podría llevar a



resultados discriminatorios o inexactos en la detección de desinformación. Para mitigar este riesgo, es fundamental utilizar conjuntos de datos diversos y representativos, así como realizar auditorías regulares del modelo para identificar y corregir cualquier sesgo que pueda surgir. La **discriminación** puede manifestarse si el sistema no considera adecuadamente las diferencias culturales y contextuales en la comunicación. La desinformación puede variar significativamente entre diferentes grupos demográficos y contextos sociales. Por lo tanto, el diseño del sistema debe incluir un enfoque inclusivo que contemple estas variaciones para evitar perpetuar estigmas o injusticias. La **transparencia** es crucial en el uso de IA, especialmente en un sistema destinado a detectar desinformación. Los usuarios deben entender cómo funciona el sistema y cómo se toman las decisiones sobre qué contenido se clasifica como falso o verdadero. Esto implica proporcionar explicaciones claras sobre los algoritmos utilizados y los criterios empleados para evaluar la veracidad del contenido. La **seguridad** también es un aspecto vital, ya que el sistema puede ser objeto de ataques maliciosos que intenten manipular su funcionamiento o comprometer su integridad. Es esencial implementar protocolos robustos de ciberseguridad para proteger tanto los datos del usuario como la infraestructura del sistema. Por último, la **prevención** del daño es un principio ético fundamental. El objetivo del proyecto es promover el bienestar social y reducir la propagación de desinformación. Esto implica no solo detectar contenido engañoso, sino también educar a los usuarios sobre cómo identificar noticias falsas por sí mismos y fomentar una cultura informativa más crítica y responsable.

## 1.5 Equipo de Desarrollo de la PoC

*Describe las disciplinas que cubre tu equipo y que se requirieron para desarrollar esta PoC. (Áreas de Electrónica, Software, Leyes, Salud, Administración, etc.) y cómo estos conocimientos o perspectivas aportaron para generar esta PoC o por qué fueron necesarios.*

El desarrollo de la prueba de concepto (PoC) para el proyecto de detección de desinformación ha requerido la colaboración de un equipo multidisciplinario que abarca diversas áreas del conocimiento. Estas disciplinas son fundamentales para abordar los complejos desafíos que presenta la desinformación y para crear un sistema efectivo y comprensible. En primer lugar, el área de **informática** es esencial, ya que proporciona las bases tecnológicas para el desarrollo del sistema. Los expertos en esta disciplina han contribuido con conocimientos en algoritmos de **Machine Learning** y **Deep Learning**, lo que permite la creación de modelos capaces de analizar grandes volúmenes de datos y detectar patrones en el contenido noticioso. La experiencia en **procesamiento del lenguaje natural (NLP)** es particularmente relevante, ya que se utiliza para entender y clasificar el lenguaje en las noticias, identificando características lingüísticas que diferencian las noticias falsas de las verdaderas. La **psicología** también juega un papel crucial en el proyecto, ya que ayuda a comprender los mecanismos cognitivos detrás de la propagación de desinformación. Investigaciones en esta área han mostrado cómo factores como la incertidumbre y la ansiedad pueden influir en la credulidad de las personas hacia las noticias falsas. Este conocimiento es vital para diseñar estrategias que no solo detecten desinformación, sino que también eduquen a los usuarios sobre cómo reconocerla. Desde la perspectiva de la **lingüística**, los expertos han aportado técnicas para analizar el contenido textual, enfocándose en las estructuras lingüísticas y pragmáticas que suelen estar presentes en las noticias falsas. Esto incluye el uso de indicadores lingüísticos para detectar diferencias sistemáticas entre textos verdaderos y falsos,

lo cual es esencial para entrenar los modelos de detección.El campo de la **comunicación** también es relevante, ya que permite entender cómo se difunde la desinformación a través de diferentes canales y plataformas. Los investigadores en esta área analizan los mecanismos de diseminación y cómo las redes sociales influyen en la propagación del contenido engañoso. Este conocimiento ayuda a diseñar un sistema que no solo detecte desinformación, sino que también considere el contexto social en el que se produce.Finalmente, la **neurociencia cognitiva** aporta una dimensión adicional al proyecto al estudiar cómo los usuarios responden a la desinformación a nivel neurológico. Esto puede incluir investigaciones sobre cómo se procesan las noticias falsas en el cerebro y qué características hacen que sean más persuasivas.La combinación de estos conocimientos interdisciplinarios ha sido fundamental para desarrollar una PoC robusta y efectiva. Cada disciplina aporta perspectivas únicas que enriquecen el análisis y mejoran la capacidad del sistema para detectar desinformación de manera precisa y comprensible. Esta colaboración no solo fortalece el enfoque técnico del proyecto, sino que también asegura que se consideren aspectos sociales y psicológicos en su diseño e implementación.

2. Escenario de Pruebas según TRL:

1

TRL 2 – 3: Prueba de Concepto: Proceso de Validación a Nivel de Laboratorio

Describe las pruebas preliminares de tu proyecto y los resultados de éstas:

NOMBRE DE LA PRUEBA	TIPO DE PRUEBA	RESULTADOS / CONCLUSIONES
Prueba de Verificación Automática de Hechos	Prueba de Verificación Automática	Se implementó un sistema de verificación automática utilizando grafos de conocimiento. Los resultados mostraron una precisión superior al 80% en la identificación de hechos falsos, validando la efectividad del enfoque automatizado en comparación con métodos manuales.
Evaluación de Algoritmos Multimodales	Prueba Multimodal	Se realizaron pruebas utilizando conjuntos de datos multimodales (texto e imágenes). Los resultados indicaron que el uso de características visuales mejoró la precisión en un 15% en comparación con los modelos que solo usaban texto, destacando la importancia de un enfoque integral.
Análisis de Sensibilidad a Sesgos	Prueba de Sesgo	Se evaluó el modelo para detectar sesgos en la clasificación de noticias según diferentes grupos demográficos. Los resultados revelaron que el modelo presentaba sesgos menores, lo que sugiere que se requieren ajustes adicionales para garantizar equidad en la detección.



# Acelerando México con IA

intel

Simulación de Propagación  
en Redes Sociales

Prueba  
Comportamental

Se simuló la propagación de noticias falsas y verdaderas en redes sociales utilizando datos históricos. Los resultados confirmaron que las noticias falsas se propagan más rápidamente, lo que refuerza la necesidad de detección temprana y proactiva.

Test de Usabilidad del  
Sistema

Prueba de Usabilidad

Se realizó una evaluación con usuarios finales para medir la usabilidad del sistema. Los resultados mostraron una alta satisfacción del usuario con una puntuación promedio de 4.5 sobre 5, indicando que el sistema es accesible y fácil de usar.



# PROTOCOLO DE INVESTIGACIÓN

Ciudad de México a 26 de noviembre de 2024.

## 1. DATOS GENERALES DEL PROYECTO DE INVESTIGACIÓN

1.1 Nombre del proyecto	Identificación de patrones de desinformación en medios de comunicación.
1.2 Departamento involucrado	Departamento de Estudios en Ingeniería para la Innovación
1.3 Línea de Investigación en la que participa el proyecto	<ul style="list-style-type: none"><li>LGAC: “Sistemas dinámicos y control”, Doctorado y Maestría en Ciencias de la Ingeniería.</li><li>INIAT: “Tecnología para cómputo y ciencia de datos.”</li></ul>
1.4 Palabras clave	Noticias Falsas, Desinformación, Detección automática, Procesamiento del Lenguaje Natural, Aprendizaje Automatizado, Principios de Persuasión

## 2. DATOS GENERALES DE LOS INVESTIGADORES

### 2.1. Investigador Responsable:

Nombre	Dr. Lázaro Bustio Martínez
Puesto	ATC, DEII, IBERO CDMX
Nombramiento SNI	SNI – Nivel 1

### 2.2. Listado de los participantes

#### Coinvestigadores:

Nombre	Puesto	Institución	Nombramiento
Dr. Carlos Francisco Betancourt	ATC, DEII	IBERO CDMX	
MSc. Felipe Antonio Trujillo	ATC, DEII	IBERO CDMX	
Dr. Vitali Herrera Semenets	Investigador Adjunto	CENATAV Cuba	
Dr. Leovildo Diago	PSPD, FisMat	IBERO CDMX	SNI – Nivel 1
Dr. Miguel Ángel Álvarez Carmona	Investigador	CIMAT	SNI – Nivel 1
Un Lingüista??			

\*ATC= Académico de Tiempo Completo

\*PSPD = Prestador de Servicios Profesionales Docentes

#### Colaboradores:

Nombre	Puesto	Institución	Nombramiento
Dr. Juan Luis Mendoza García	Investigador	Pangeanic - España	
Dr. Pedro Santander Molina	Investigador	PUCV - Chile	
Dr. Jan van den Berg	Profesor Emérito	TU-DELFT – Países Bajos	
Viviana Fuentes Fuentes	Psicóloga	Autónomo	
Dr. Jorge Ángel González Ordiano	Investigador	IBERO CDMX	SNI – Nivel 1
Ing.Fis. Alejandro Gómez Salcido	Investigador	Técnica Salgar	

### 3. RESUMEN EJECUTIVO

Con la proliferación de la tecnología digital, el fenómeno de la propagación de desinformación<sup>1</sup> se ha convertido en un problema cada vez más frecuente en el entorno informativo. Esta difusión maliciosa de contenidos perjudiciales, ofensivos o ilegítimos puede resultar en engaños, malentendidos y tensiones sociales, lo que impacta negativamente en la estabilidad y el desarrollo económico sostenible de las comunidades. Aunque se ha abordado la detección de noticias falsas desde diversas perspectivas, con el constante avance de la inteligencia artificial se hace necesario explorar nuevas formas de extracción y análisis automático e inteligente de datos noticiosos (como la comunicación, la lingüística y la psicología) para identificar eficazmente las noticias falsas.

Si bien existen propuestas basadas en inteligencia artificial para la detección de noticias falsas, es fundamental desarrollar soluciones comprensibles desde la perspectiva de la explicabilidad. En la actualidad, tanto el análisis de la estructura del modelo como el análisis del comportamiento del modelo en los métodos de inteligencia artificial para la detección de noticias falsas aún no están completamente establecidos. Por lo tanto, el diseño de un sistema de detección de noticias falsas se ha vuelto crucial en el complejo entorno de difusión de información actual. Es por ello por lo que este protocolo propone un proyecto cuyo objetivo es desarrollar métodos capaces de identificar desinformación en mensajes electrónicos mediante el uso de técnicas de inteligencia artificial comprensibles basados en la identificación de patrones lingüísticos y psicológicos.

### 4. ANTECEDENTES GENERALES

Con el rápido desarrollo de la tecnología de Big Data y la difusión de información, la desinformación se propaga a través de las redes sociales, representando una amenaza destructiva para el desarrollo sostenible de la sociedad. La desinformación no solo mina la confianza pública y perturba el orden social, sino que también genera decisiones erróneas, divisiones sociales y conflictos que obstaculizan el funcionamiento adecuado y el avance progresivo de la sociedad. La desinformación se refiere específicamente a informes que son falsos o exagerados, como se ilustra en la Figura 1. Estos informes pueden ser creados con la intención deliberada de engañar al público o promover una agenda determinada. Se ha llegado a un acuerdo en la comunidad científica en torno al contexto de intensificación de la desinformación, es decir, a los momentos en los cuales se manifiesta con mayor fuerza e intensidad este “desorden informativo”. Se determinó que las crisis (sanitarias, por ejemplo, como la del Covid) y los momentos de campañas electorales son buenos contextos para explorar desinformación (y recolectar datos). En comparación con la verdad, los rumores tienden a propagarse más lejos, más rápido y de manera más amplia, lo que sugiere que la desinformación es más atractiva que la información. Sin embargo, la desinformación suele generar ansiedad en las personas, afecta el funcionamiento normal de la sociedad y representa una amenaza para su desarrollo sostenible.

---

<sup>1</sup> El parlamento británico, muy influido por lo que ocurrió un año antes durante la campaña del Brexit, elaboró en 2017 un Informe en el cual se hacen precisiones definicionales sobre el concepto; se trata de uno de los documentos pioneros que propone no trabajar con la noción de ‘fake news’ o noticias falsas, sino con los de ‘misinformation’ (información errónea) y ‘disinformation’ (desinformación). Ese mismo año, el Consejo de Europa, con sede en Estrasburgo, en otro informe sobre el tema, agregó un tercer concepto, el de ‘mal-information’ o información malintencionada (Wardle y Derakshan, 2017).

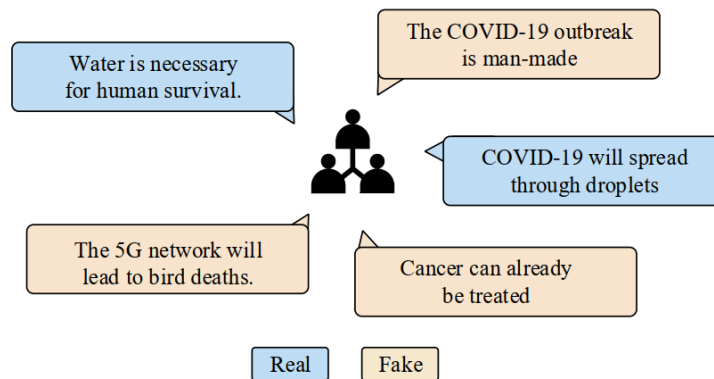


Figura 1 Diseminación de noticias falsas en medios de comunicación

El progreso de la investigación multidisciplinaria sobre la desinformación ha involucrado a varias disciplinas como la informática, sociología, psicología, lingüística, comunicación y neurociencia cognitiva<sup>2</sup>. Cada una aporta sus propios métodos y contenidos de investigación (ver Figura 2). Algunos estudios han adoptado enfoques interdisciplinarios para mejorar la detección de desinformación (noticias falsas en este caso<sup>3</sup>), combinando conocimientos de estas áreas. Se espera que la investigación teórica multidisciplinaria sobre la desinformación contribuya a la detección de estas en las ciencias naturales y que la investigación futura integre aún más conocimientos interdisciplinarios.

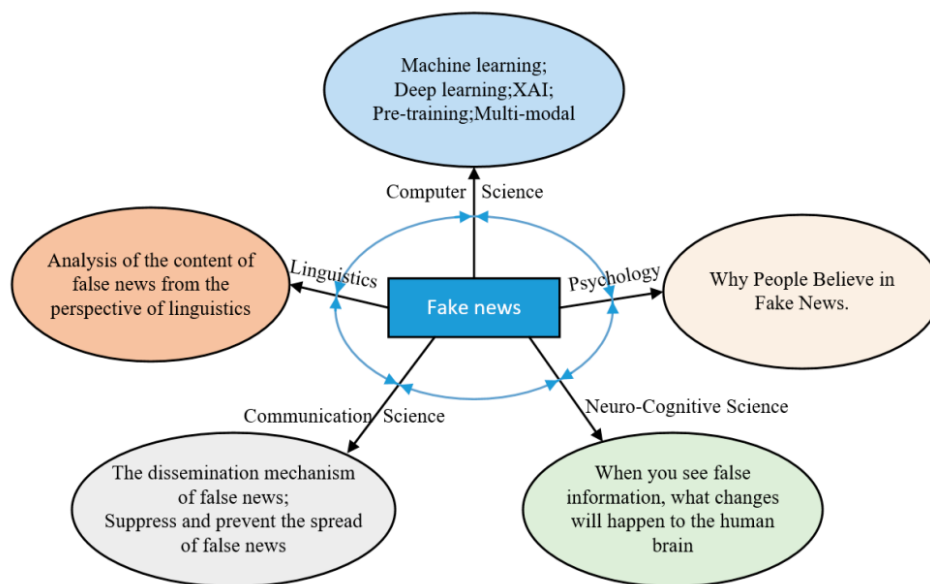


Figura 2 Investigaciones en campos multidisciplinarios

**En psicología**, los investigadores se centran en el comportamiento cognitivo relacionado con la desinformación y los mecanismos psicológicos que impulsan su difusión<sup>4</sup>. Bordia et al.<sup>5</sup> descubrieron que la

<sup>2</sup> Guo, B.; Ding, Y.; Sun, Y.; Ma, S.; Li, K.; Yu, Z. The mass, fake news, and cognition security. *Front. Comput. Sci.* **2021**, *15*, 153806

<sup>3</sup> Zhou, X.; Jain, A.; Phoha, V.V.; Zafarani, R. Fake news early detection: An interdisciplinary study. *arXiv* **2019**, arXiv:1904.11679

<sup>4</sup> Greifeneder, R.; Jaffe, M.; Newman, E.; Schwarz, N. *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation*; Routledge: London, UK, 2021.

<sup>5</sup> Bordia, P.; DiFonzo, N. 10 Rumors during organizational change: A motivational analysis. In *The Psychology of Organizational Change: Viewing Change from the Employee's Perspective*; Cambridge University Press: Cambridge, UK, 2013; p. 232.

interacción con la desinformación se basa en gran medida en la necesidad psicológica de las personas por la verdad de los hechos. Factores como la incertidumbre y la ansiedad promueven la propagación de la desinformación. Pennycook et al.<sup>6</sup> estudiaron por qué las personas creen y comparten noticias falsas o engañosas en línea, concluyendo que la motivación política tiene un fuerte efecto causal en las creencias que lleva a las personas a creer en la desinformación y noticias falsas.

**En neurociencia cognitiva**, la investigación de Arisoy et al.<sup>7</sup> se centró en la sensibilidad de los usuarios a la desinformación mediante métodos neurocognitivos. Utilizaron electroencefalogramas (EEG) para estudiar la base neural que diferencia las noticias falsas de las verdaderas. Sus experimentos revelaron que la detección humana de noticias falsas puede ser ineficaz y tener efectos adversos potenciales.

**Los investigadores lingüísticos** utilizan la tecnología informática para analizar el contenido lingüístico de la desinformación y resumir sus características pragmáticas y estructuras lingüísticas que promueven su difusión<sup>8</sup>. Choudhary et al.<sup>9</sup> consideran prometedor el uso de indicadores lingüísticos, aplicando análisis de datos cualitativos y cuantitativos para detectar y comparar atributos específicos del vocabulario, gramática y características sintácticas, con el fin de identificar las diferencias sistemáticas entre las noticias falsas y las verdaderas.

**En el campo de la ciencia de la comunicación**, los investigadores se enfocan en analizar el concepto de noticias falsas<sup>10</sup> y en descubrir los mecanismos de diseminación de información, así como modelos de prevención y gobernanza de la desinformación en el contexto del uso continuo de las redes sociales<sup>11</sup>. Jana et al.<sup>12</sup> sugieren que el concepto actual de desinformación es amplio y constituye un fenómeno bidimensional de la comunicación pública, estableciendo así un marco teórico para la investigación de las noticias falsas. Di et al.<sup>13</sup> identificaron que la motivación para compartir noticias falsas no siempre es por objetivos financieros o políticos/ideológicos, sino que también puede ser por la búsqueda de reconocimiento social dentro de un grupo deseado, informando a otros miembros sobre temas específicos relacionados, lo cual fortalece la unidad del grupo.

La categorización de la desinformación y específicamente de las noticias falsas es multifacética y diversa, abarcando desde rumores no verificados que circulan en redes sociales hasta propaganda engañosa

---

<sup>6</sup> Pennycook, G.; Rand, D.G. The psychology of fake news. *Trends Cogn. Sci.* **2021**, *25*, 388–402.

<sup>7</sup> Arisoy, C.; Mandal, A.; Saxena, N. Human Brains Can't Detect Fake News: A Neuro-Cognitive Study of Textual Disinformation Susceptibility. In Proceedings of the 2022 19th Annual International Conference on Privacy, Security & Trust (PST), IEEE, Fredericton, NB, Canada, 22–24 August 2022; pp. 1–12.

<sup>8</sup> Giachanou, A.; Rissola, E.A.; Ghanem, B.; Crestani, F.; Rosso, P. The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In Proceedings of the Natural Language Processing and Information Systems: 25th International Conference on Applications of Natural Language to Information Systems, NLDB 2020, Saarbrücken, Germany, 24–26 June 2020; Proceedings 25. Springer: Cham, Switzerland, 2020; pp. 181–192.

<sup>9</sup> Choudhary, A.; Arora, A. Linguistic feature-based learning model for fake news detection and classification. *Expert Syst. Appl.* **May 2021**, *169*, 114171.

<sup>10</sup> Shu, K.; Cui, L.; Wang, S.; Lee, D.; Liu, H. defend: Explainable fake news detection. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 395–405.

<sup>11</sup> Rubin, V.L. Disinformation and misinformation triangle: A conceptual model for “fake news” epidemic, causal factors and interventions. *J. Doc.* **2019**, *75*, 1013–1034

<sup>12</sup> Egelhofer, J.L.; Lecheler, S. Fake news as a two-dimensional phenomenon: A framework and research agenda. *Ann. Int. Commun. Assoc.* **2019**, *43*, 97–116.

<sup>13</sup> Di Domenico, G.; Sit, J.; Ishizaka, A.; Nunan, D. Fake news, social media and marketing: A systematic review. *J. Bus. Res.* **2021**, *124*, 329–341.

deliberadamente difundida por sus creadores. Según las referencias<sup>14 15</sup>, las noticias falsas se clasifican en cinco categorías (ver Tabla 1): (1) noticias falsas engañosas; (2) información falsa de naturaleza rumorística; (3) información falsa de comentarios; (4) noticias falsas del tipo sensacionalista en los titulares; (5) recombinación basada en hechos de información falsa.

Adicionalmente, el fenómeno de la desinformación puede explicarse mediante un concepto ampliamente utilizado en la actualidad: la clasificación tripartita. Esta clasificación engloba tres nociones clave que, en conjunto, conforman lo que se denomina un ambiente de desorden informativo. Estas nociones son las más empleadas en el debate especializado sobre el tema, ya que permiten delimitar, tanto definicional como conceptualmente, este término paraguas de la siguiente manera:

- a) Disinformation (desinformación): información falsa creada deliberadamente para dañar a una persona, grupo social u organización.
- b) Mal-information, traducido como “información maliciosa” o “malintencionada”: se refiere a información basada en hechos reales, es decir, se deriva de la verdad, pero que contiene medias verdades y contenido descontextualizado y que se exagera de tal manera que induce a error y es usada para infligir daño a una persona, grupo social u organización
- c) Mis-information (información errónea): información que es falsa y se comparte, pero no es creada con la intención de causar daño.

La difusión de información exhibe diversas características que incluyen la rapidez con la que se propaga, la sobrecarga de información, la universalidad del contenido, la dificultad para distinguir la autenticidad, el impacto perjudicial, el alcance transregional, la discriminación de estigmas, la interacción social de los medios, entre otros aspectos<sup>16</sup>. Una gran cantidad de desinformación se difunde en las redes sociales y en las principales plataformas de medios de comunicación, dando lugar al fenómeno conocido como ambiente de desorden informativo. La esencia de esta epidemia radica en la presentación de un patrón complejo que integra diversos mecanismos de comunicación, como la comunicación de masas, la comunicación en red y la comunicación inteligente, en el contexto de la nueva tecnología. Por consiguiente, la detección de tecnologías para identificar noticias falsas se convierte en una necesidad apremiante en la sociedad contemporánea.

Tabla 1 Clasificación y definición de las noticias falsas.

CLASIFICACIÓN DE NOTICIAS FALSAS	DEFINICIÓN
<b>NOTICIAS FALSAS ENGAÑOSAS</b>	Información falsa destinada a engañar y confundir al lector. Este tipo de desinformación es especialmente insidioso, ya que su objetivo es deliberadamente engañar a los lectores o causar efectos adversos.
<b>INFORMACIÓN FALSA DE NATURALEZA RUMORAL</b>	Rumores no confirmados, rumores o mensajes anónimos, etc.
<b>INFORMACIÓN FALSA DE COMENTARIOS</b>	Un comentario falso o engañoso publicado en una plataforma en línea, redes sociales u otra plataforma interactiva.
<b>NOTICIAS FALSAS DEL TIPO SENSACIONALISTA EN LOS TITULARES</b>	Editar titulares falsos llamativos, que presentan contenido real, pero carecen de valor referencial para la noticia.

<sup>14</sup> Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; Liu, H. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* **2020**, *8*, 171–188

<sup>15</sup> Zellers, R.; Holtzman, A.; Rashkin, H.; Bisk, Y.; Farhadi, A.; Roesner, F.; Choi, Y. Defending Against Neural Fake News. *arXiv* **2020**, arXiv:1905.12616.

<sup>16</sup> Apuke, O.D.; Omar, B. Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telemat. Inform.* **2021**, *56*, 101475.



La mayoría de las investigaciones actuales sobre la desinformación consideran esta tarea como una de clasificación. Los métodos se dividen en cuatro categorías según las características principales utilizadas por el modelo de clasificación: métodos basados en contenido, métodos basados en redes sociales, métodos basados en conocimiento y multimodal.

El **enfoque basado en contenido** busca extraer características semánticas del contenido de los mensajes para determinar su autenticidad. Los textos falsos suelen ser más subjetivos y utilizan un lenguaje diferente al de los textos verdaderos, con un mayor uso de palabras exageradas y pronombres en primera y segunda persona. Los autores de textos falsos tienden a ser más extremos, mostrando tendencias políticas y odio. No todas las características lingüísticas tienen el mismo peso, y la importancia de cada una varía. En algunos estudios se analizan los estilos de escritura de los textos de izquierda y derecha<sup>17 18</sup> y se ha identificado que todos tienen tendencias extremistas, políticas y de odio. Song et al.<sup>19</sup> extrajeron un conjunto completo de características de contenido de noticias reales y falsas, y mediante experimentos, clasificaron su importancia, encontrando que el número total de palabras y la longitud del contenido son significativos.

El enfoque basado en contenidos puede descubrir las características lingüísticas de los textos verdaderos y falsos. Sin embargo, en ocasiones, la desinformación logra engañar a los lectores al imitar deliberadamente las técnicas de redacción utilizadas en los textos reales. A pesar de su eficacia, el enfoque basado en contenidos no puede distinguir de manera confiable entre textos falsos y verdaderos.

Para abordar este desafío, el **enfoque basado en redes sociales** aprovecha al máximo la información oculta, como datos auxiliares, antecedentes sociales y rutas de propagación en las redes. Por ejemplo, Shu et al.<sup>20</sup> exploraron la relación entre los datos de los usuarios y la desinformación en las redes sociales, utilizando la participación social del usuario como información auxiliar para la detección. Además, propusieron un marco que simula la relación triádica entre editores de noticias, artículos de noticias y usuarios<sup>21</sup>. Mediante la extracción de características efectivas del comportamiento de participación de editores y lectores de noticias, se logra capturar la interacción entre estos actores. Estudios han demostrado que el uso de información sobre antecedentes sociales no solo mejora la detección de desinformación, sino que también permite predecirla de manera temprana y efectiva.

Otra línea de investigación basada en este enfoque se centra en detectar noticias falsas al simular la ruta de propagación de las noticias en la red. A través de experimentos, Monti et al.<sup>22</sup> descubrieron que el modo

---

<sup>17</sup> Potthast, M.; Kiesel, J.; Reinartz, K.; Bevendorff, J.; Stein, B. A stylometric inquiry into hyperpartisan and fake news. *arXiv* **2017**, arXiv:1702.05638.

<sup>18</sup> Rashkin, H.; Choi, E.; Jang, J.Y.; Volkova, S.; Choi, Y. Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 2931–2937.

<sup>19</sup> Sheikhi, S. An effective fake news detection method using WOA-xgbTree algorithm and content-based features. *Appl. Soft Comput.* **2021**, *109*, 107559.

<sup>20</sup> Shu, K.; Wang, S.; Liu, H. Understanding user profiles on social media for fake news detection. In Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE, Miami, FL, USA, 10–12 April 2018; pp. 430–435.

<sup>21</sup> Shu, K.; Wang, S.; Liu, H. Beyond news contents: The role of social context for fake news detection. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, VIC, Australia, 11–15 February 2019; pp. 312–320.

<sup>22</sup> Monti, F.; Frasca, F.; Eynard, D.; Mannion, D.; Bronstein, M.M. Fake news detection on social media using geometric deep learning. *arXiv* **2019**, arXiv:1902.06673.

de transmisión es una característica crucial de las noticias falsas, superando incluso otros aspectos como el contenido de las noticias, los datos del usuario y el comportamiento social. En el trabajo de Raza et al.<sup>23</sup> propusieron un marco de detección de noticias falsas basado en la arquitectura Transformer, que incluye partes de codificador y decodificador. La parte del codificador se utiliza para aprender la representación de los datos de noticias falsas, y la parte del decodificador se utiliza para predecir el comportamiento futuro basado en observaciones pasadas. El modelo utiliza las características del contenido de las noticias y el contexto social para mejorar la precisión de la clasificación.

La **detección de desinformación basada en conocimiento** verifica la autenticidad de las información al verificar las noticias falsas y los hechos, por lo que también se le llama verificación de hechos. Existen dos categorías de verificación de hechos: verificación manual y verificación automática. El método manual utiliza el conocimiento de expertos en el dominio o el método de *crowdsourcing*, siendo preciso, pero poco eficiente para la era de grandes volúmenes de datos.

Por otro lado, el método de verificación automática emplea procesamiento de lenguaje natural y tecnología de aprendizaje automático, y se ha convertido en un campo de investigación candente. La verificación de hechos comienza construyendo una base de conocimientos o un grafo de conocimientos a partir de la red mediante la extracción de conocimientos. Luego, compara y verifica las noticias falsas con la base de conocimientos para juzgar la autenticidad de las noticias. Por ejemplo, Pan et al. utilizaron grafos de conocimiento para detectar noticias falsas basadas en el contenido de las noticias, resolviendo el problema de que la verificación computacional de hechos no es lo suficientemente completa. Su método, que extrae tripletas de los artículos de noticias, logra un puntaje F0 superior al 80.1. Además, Hu et al. desarrollaron una red de atención de gráficos heterogéneos para aprender el contexto de la representación de las noticias y codificar la semántica del contenido de las noticias. Utilizando una red de comparación de entidades, comparan la representación de la entidad de contexto con la representación derivada de la base de conocimientos (KB), con el objetivo de capturar la coherencia entre el contenido de las noticias y la KB .

El **enfoque multimodal** se refiere al uso de múltiples fuentes de datos y metadatos, como texto e imágenes, para identificar contenido engañoso o inexacto en las noticias. Recientemente, se ha puesto más énfasis en las características visuales, ya que las noticias falsas a menudo utilizan contenido visual no verificado para engañar a los lectores. Afortunadamente, existen conjuntos de datos multimodales, como FakeNewsNet, que proporcionan una amplia gama de información para mejorar la detección de noticias falsas mediante la fusión de múltiples características.

La investigación específica sobre la detección de desinformación abarca la verificación de hechos, la identificación de posiciones, la detección de temas, así como tareas que implican la clasificación y agrupación de texto, la comprensión de imágenes, el reconocimiento de voz, y otras áreas de investigación. Utilizando tecnologías como la minería de texto, el aprendizaje automático, el aprendizaje profundo, el procesamiento del lenguaje natural, la visión por computadora, entre otras, se extraen e identifican información clave a partir de percepciones visuales subjetivas de textos o páginas de noticias.

A partir del análisis y clasificación exhaustiva de diversas fuentes, se identifican tres enfoques fundamentales orientados a la detección de desinformación:

**Detección basada en aprendizaje automático convencional:** Este enfoque utiliza modelos de clasificación estándar como Máquinas de Soporte Vectorial<sup>24</sup>, Naive Bayes<sup>25</sup>, Regresión Logística<sup>26</sup> y

---

<sup>23</sup> Raza, S.; Ding, C. Fake news detection based on news content and social contexts: A transformer-based approach. *Int. J. Data Sci. Anal.* **2022**, *13*, 335–362.

<sup>24</sup> Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28.

<sup>25</sup> Duda, R.O.; Hart, P.E. *Pattern Classification and Scene Analysis*; Wiley: New York, NY, USA, 1973; Volume 3, Available online: <https://api.semanticscholar.org/CorpusID:12946615> (accessed on 10 April 2024).

Árboles de Decisión<sup>27</sup>, incluyendo clasificadores de Bosques Aleatorios<sup>28</sup>. Estos modelos operan identificando patrones en el texto a través de características manuales derivadas del conocimiento experto, tales como aspectos lingüísticos, temáticos, de usuarios y de comunicación. Por ejemplo, Eldesoky et al.<sup>29</sup> presentaron un modelo de clasificación que emplea técnicas de extracción de características como Doc2vec y Word2vec, obteniendo una precisión del 95.5% en datos del mundo real.

**Detección basada en aprendizaje profundo:** Esta fase supera a los enfoques tradicionales de aprendizaje automático, los cuales dependen de la extracción manual de características y pueden ser propensos a errores y lentos en su procesamiento. Investigaciones como la de Lai et al.<sup>30</sup> han demostrado que los modelos de redes neuronales ofrecen un rendimiento superior, con una precisión aproximadamente un 6% mayor. Estos modelos emplean técnicas de incrustación de palabras para fusionar modelos de lenguaje y características de aprendizaje, mostrando eficacia en la detección de noticias falsas en diferentes idiomas<sup>31</sup>. Por ejemplo, Ilie et al.<sup>32</sup> utilizaron diversas incrustaciones de palabras como Word2Vec, FastText y GloVe para entrenar arquitecturas de aprendizaje profundo, obteniendo los mejores resultados con una arquitectura basada en redes neuronales convolucionales. Asimismo, Ma et al.<sup>33</sup> emplearon la capa oculta de una red neuronal recurrente para representar información sobre noticias falsas, demostrando la superioridad del modelo. Desde entonces, se propuso un modelo llamado FNDNet (Deep CNN)<sup>34</sup> para aprender las características discriminativas de la detección de noticias falsas utilizando múltiples capas ocultas. Además, Huang et al.<sup>35</sup> utilizaron una red neuronal convolucional gráfica (GCN)<sup>36</sup> para aprender representaciones de usuarios a partir de grafos creados por información de comportamiento de usuarios.

**Detección basada en modelos preentrenados:** Los *embeddings* tradicionales de palabras pueden no capturar relaciones contextuales complejas, tratando las palabras como entidades independientes sin considerar la estructura completa de la oración. Sin embargo, el *Transformer*, presentado por Vaswani et al.<sup>37</sup> en 2017, ofrece resultados sobresalientes en tareas de procesamiento de lenguaje natural al permitir una mejor captura del contexto y la semántica del texto. Este modelo utiliza codificación de posiciones para preservar la información de ubicación y su estructura codificador-decodificador facilita la comprensión del texto en diferentes niveles, demostrando efectividad en la detección de comportamientos maliciosos,

---

<sup>26</sup> Hosmer Jr, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398.

<sup>27</sup> Hunt, E.B.; Marin, J.; Stone, P.J. *Experiments in Induction*; Academic Press: Cambridge, MA, USA, 1966.

<sup>28</sup> Antony Vijay, J.; Anwar Basha, H.; Arun Nehru, J. A dynamic approach for detecting the fake news using random forest classifier and NLP. In *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 2*; Springer: Singapore, 2020; pp. 331–341.

<sup>29</sup> Eldesoky, I.; Moussa, F. Fake news detection based on word and document embedding using machine learning classifiers. *J. Theor. Appl. Inf. Technol.* **2021**, *99*, 1891–1901.

<sup>30</sup> Lai, C.-M.; Chen, M.-H.; Kristiani, E.; Verma, V.K.; Yang, C.-T. Fake News Classification Based on Content Level Features. *Appl. Sci.* **2022**, *12*, 1116.

<sup>31</sup> Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.

<sup>32</sup> Ilie, V.-I.; Truică, C.-O.; Apostol, E.-S.; Paschke, A. Context-Aware Misinformation Detection: A Benchmark of Deep Learning Architectures Using Word Embeddings. *IEEE Access* **2021**, *9*, 162122–162146.

<sup>33</sup> Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.-F.; Cha, M. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, New York, NY, USA, 9–15 July **2016**.

<sup>34</sup> Kaliyar, R.K.; Goswami, A.; Narang, P.; Sinha, S. FNDNet—a deep convolutional neural network for fake news detection. *Cogn. Syst. Res.* **2020**, *61*, 32–44.

<sup>35</sup> Huang, Q.; Zhou, C.; Wu, J.; Wang, M.; Wang, B. Deep structure learning for rumor detection on twitter. In *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Budapest, Hungary, 14–19 July **2019**; pp. 1–8.

<sup>36</sup> Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R. Gated graph sequence neural networks. *arXiv* **2015**, arXiv:1511.05493.

<sup>37</sup> Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Long Beach, CA, USA, 4–9 December 2017; Volume 30

incluyendo la identificación de noticias falsas. Las investigaciones realizadas en<sup>38 39 40 41</sup> han logrado un buen rendimiento en la detección de noticias falsas utilizando la arquitectura de *Transformer*. En la práctica, la elección del método de incrustación de palabras depende de factores como el tamaño del conjunto de datos, los recursos informáticos y la complejidad de las tareas de detección de noticias falsas.

Por otro lado, el modelo preentrenado BERT<sup>42</sup> ha revolucionado las tareas de procesamiento de lenguaje natural desde 2018, mejorando significativamente el rendimiento en diversas aplicaciones. Al combinar el preentrenamiento con el ajuste fino, BERT logra capturar de manera única el contexto semántico de las oraciones o textos, sin necesidad de profundizar en los detalles específicos de las noticias falsas. A pesar de los avances, aún persisten desafíos en la detección práctica de noticias falsas, y la investigación se centra en cómo extraer características semánticas más complejas para establecer modelos de detección más efectivos, y comprensibles, dentro del paradigma de "preentrenamiento + ajuste fino". La investigación en este ámbito continúa evolucionando, en busca de soluciones cada vez más efectivas para abordar el desafío de la detección de noticias falsas.

## 5. PLANTEAMIENTO DEL PROBLEMA

La desinformación se propaga y es aceptada con facilidad porque explotan el eslabón más débil en la cadena de información: las personas. Los textos de desinformación son particularmente efectivos porque se apoyan en creencias y sesgos cognitivos, y explotan las emociones humanas. Están diseñados para generar reacciones emocionales intensas, como ira, miedo o indignación, lo que impulsa a las personas a compartirlos sin verificar su autenticidad. Esta combinación de factores psicológicos, emocionales, lingüísticos y sociales convierte a la desinformación en una amenaza significativa para la sociedad, ya que puede influir en la opinión pública, polarizar a la sociedad y erosionar la confianza en las instituciones y los medios de comunicación.

Otros aspectos relevantes a considerar incluyen:

- **La velocidad y el alcance de la difusión en línea:** Las redes sociales y otras plataformas digitales permiten que la desinformación se propaguen a una velocidad y escala sin precedentes, alcanzando a un público masivo en cuestión de minutos.
- **La falta de alfabetización mediática:** Muchas personas carecen de las habilidades necesarias para evaluar críticamente la información que encuentran en línea, lo que las hace más vulnerables a la desinformación.
- **La sofisticación de las técnicas de desinformación:** Los creadores de desinformación emplean técnicas cada vez más avanzadas, como la manipulación de imágenes, voz y videos, para hacer que su contenido parezca auténtico.

Para lograr su difusión y aceptación, los creadores de textos de desinformación utilizan enfoques multidisciplinarios que explotan diversos aspectos, tales como:

---

<sup>38</sup> Raza, S.; Ding, C. Fake news detection based on news content and social contexts: A transformer-based approach. *Int. J. Data Sci. Anal.* **2022**, *13*, 335–362

<sup>39</sup> Low, J.F.; Fung, B.C.M.; Iqbal, F.; Huang, S.-C. Distinguishing between fake news and satire with transformers. *Expert Syst. Appl.* **2022**, *187*, 115824.

<sup>40</sup> Jwa, H.; Oh, D.; Park, K.; Kang, J.M.; Lim, H. exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). *Appl. Sci.* **2019**, *9*, 4062

<sup>41</sup> Gundapu, S.; Mamidi, R. Transformer based Automatic COVID-19 Fake News Detection System. *arXiv* **2021**, arXiv:2101.00180

<sup>42</sup> Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805

- **Ingeniería social:** Emplean lenguaje diseñado para evocar emociones en las víctimas, como curiosidad, miedo o duda, y, por otro lado, confianza, autoridad o pertenencia. Aplican diversas técnicas sociales y psicológicas, incluyendo sesgos cognitivos y heurísticas que influyen en el juicio humano.
- **Desconocimiento y sobrecarga informativa:** Usan un exceso de información técnica o términos especializados que pueden confundir al lector promedio. Esto también se apoya en elementos de ingeniería social y en la sobrecarga de información, lo que dificulta el análisis crítico.
- **Medios digitales y algoritmos de recomendación:** Se aprovechan de las características de las plataformas digitales y de los algoritmos de recomendación que priorizan contenido viral y emocionalmente cargado, facilitando la difusión rápida y masiva de información falsa.

Aunque se han propuesto métodos para la detección de desinformación que abordan aspectos específicos, como el análisis de texto o la verificación de hechos, no existen modelos que realicen un análisis multifactorial de la información compartida. Dado que la desinformación es un fenómeno multidisciplinario, las técnicas de detección y mitigación deben ser abordadas de la misma manera. La literatura revisada no muestra un esfuerzo efectivo en esta dirección que esté a la par del desarrollo de las tácticas de desinformación.

Teniendo en cuenta estas consideraciones, este proyecto plantea **la creación de un modelo computacional que combina técnicas de Aprendizaje Automático, análisis del discurso, análisis lingüístico, identificación de patrones psico-sociales entre otros, para la identificación y detección de desinformación.** La hipótesis general que se postula con este proyecto es **se pueden detectar la desinformación, de una manera efectiva y eficaz, si se logra identificar los patrones de lenguaje, psicológicos y sociales empleados en su elaboración y propagación empleando técnicas de aprendizaje automatizado.** Para lograr la identificación de estos patrones, se propone el uso de técnicas de Aprendizaje Automatizado, Procesamiento del Lenguaje Natural, Análisis del Discurso y análisis psicológicos de la comunicación entre otros.

El modelo computacional propuesto deberá tener las siguientes características:

- **Ligero:** Deberá ser lo suficientemente ligero para ser utilizado en sistemas computacionales de uso cotidiano, considerando las diversas características de los equipos de cómputo actuales.
- **Compatibilidad:** El modelo debe funcionar en las distintas plataformas digitales.
- **Alto desempeño:** El modelo desarrollado deberá mostrar resultados de desempeño (medido en términos de eficiencia y eficacia) competitivos o incluso mejores que los reportados en el estado del arte. Los resultados obtenidos serán validados mediante la comparación con otros trabajos reportados en el estado del arte del tema.
- **Independencia del conocimiento técnico del usuario:** Deberá ser fácil de usar y no requerir conocimientos técnicos avanzados por parte del usuario.
- **No invasivo y respetuoso de la privacidad:** El modelo deberá ser no invasivo y respetuoso de la información personal de los usuarios, garantizando la privacidad y seguridad de los datos.
- **Explicabilidad:** El modelo desarrollado deberá ser capaz de explicar su comportamiento y decisiones a los usuarios de manera clara y efectiva.

## 6. PREGUNTAS DE INVESTIGACIÓN E HIPÓTESIS DE TRABAJO

Partiendo del problema planteado, se considera para su solución las siguientes hipótesis:

- H1.** Las noticias falsas presentan patrones lingüísticos, sociales y psicológicos específicos que pueden ser identificados mediante técnicas de Procesamiento de Lenguaje Natural y Aprendizaje Automático.
- H2.** Los modelos de detección de noticias falsas basados en Aprendizaje Automático pueden ofrecer resultados de detección menos precisos que los modelos basados en Aprendizaje Profundo, pero



requieren menos poder de procesamiento y son más comprensibles para el usuario final. Esto es preferible a tener modelos más precisos pero que sean muy demandantes computacionalmente y difíciles de interpretar.

**H3.** La estructura de las redes sociales y el comportamiento de los usuarios son factores críticos en la propagación de noticias falsas, y su análisis puede mejorar la detección temprana.

**H4.** La implementación de técnicas de explicabilidad en modelos de inteligencia artificial mejora la confianza de los usuarios en los sistemas de detección de noticias falsas.

Derivado de las hipótesis planteadas surgen las siguientes preguntas de investigación que guiarán el desarrollo de este proyecto:

**P1.** ¿Cuáles son las bases de datos existentes de noticias falsas y qué características tienen en términos de tamaño, diversidad de fuentes y calidad de anotación?

- a) ¿Cuáles son las fuentes principales de estas bases de datos?
- b) ¿Qué criterios se utilizan para la clasificación y anotación de las noticias dentro de estas bases de datos?
- c) ¿Cómo se comparan en términos de tamaño y diversidad con las bases de datos de noticias verdaderas?

**P2.** ¿Cuáles son los patrones lingüísticos específicos que distinguen las noticias falsas de las verdaderas?

- a) ¿Qué técnicas de Procesamiento de Lenguaje Natural son más efectivas para identificar estos patrones?
- b) ¿Qué características semánticas y sintácticas son más prominentes en las noticias falsas?

**P3.** ¿Qué emociones específicas intentan provocar los atacantes con sus mensajes de noticias falsas y cómo estas emociones influyen en la propagación y credibilidad de la información falsa?

- a) ¿Qué técnicas retóricas y psicológicas se utilizan en la redacción de noticias falsas para suscitar emociones como el miedo, la ira o la ansiedad?
- b) ¿Cómo varían estas técnicas según el contexto sociocultural y político?
- c) ¿Qué impacto tienen estas emociones en el comportamiento de los usuarios en línea, incluyendo la compartición y la creencia en las noticias falsas?
- d) ¿Cómo pueden las técnicas de análisis de sentimientos y emociones en el procesamiento de lenguaje natural ayudar a identificar la intención emocional detrás de las noticias falsas?

**P4.** ¿Qué representación de datos resulta adecuada para la identificación de noticias falsas?

**a.** ¿Cómo se comparan las representaciones basadas en características textuales (como n-gramas o bag-of-words) con las representaciones basadas en embeddings de palabras (como Word2Vec o GloVe) en términos de efectividad para detectar noticias falsas?

**b.** ¿Qué impacto tienen las representaciones semánticas contextuales (como las generadas por BERT o GPT) en la precisión de los modelos de identificación de noticias falsas en comparación con las representaciones estáticas?

**c.** ¿Qué rol juegan las características de los datos contextuales (como la fuente de la noticia, el autor o el medio de comunicación) en la efectividad de la representación de datos para la identificación de noticias falsas?

**d.** ¿Qué combinaciones de técnicas de representación de datos (por ejemplo, integración de representaciones textuales y semánticas) resultan más efectivas para identificar noticias falsas?

- e. ¿Cómo afectan las características lingüísticas y discursivas (como el uso de lenguaje emocional, sesgado o manipulativo) a la efectividad de las representaciones de datos para identificar noticias falsas?
- f. ¿Qué técnicas de análisis del discurso pueden mejorar la identificación de patrones lingüísticos que son comunes en noticias falsas, y cómo se integran estas técnicas en las representaciones de datos?
- g. ¿Qué papel juega la psicología del lector en la percepción de noticias falsas y cómo puede esta información ser reflejada en las representaciones de datos para mejorar la detección?
- h. ¿Cómo influyen las técnicas de análisis de sentimientos y emociones en la representación de datos para la identificación de noticias falsas, y qué características emocionales son más indicativas de desinformación?

**P5.** ¿Pueden los modelos de detección de noticias falsas basados en Aprendizaje Automático ofrecer un balance adecuado entre precisión, demanda computacional y comprensibilidad?

- a) ¿Qué algoritmos de aprendizaje automático ~~de bajo costo computacional~~ son más efectivos para la detección de noticias falsas empleando la mejor representación de datos?
- b) ¿Cómo se compara la precisión de estos modelos con los basados en Aprendizaje Profundo?
- c) ¿De qué manera la explicabilidad de los modelos influye en la aceptación y confianza del usuario final?

~~**P6.** ¿Cómo afecta la estructura y el comportamiento de los usuarios en las redes sociales a la propagación de noticias falsas?~~

- ~~a) ¿Qué roles juegan los influenciadores y nodos clave en la red en la difusión de desinformación?~~
- ~~b) ¿Cómo puede la visualización de las rutas de propagación ayudar en la identificación temprana de noticias falsas?~~

**P7.** ¿De qué manera pueden ser integradas las técnicas de Procesamiento de Lenguaje Natural, Análisis del Discurso y conocimientos de psicología en el diseño y desarrollo de un prototipo de aplicación destinado a distinguir entre noticias falsas y noticias legítimas con el fin de mejorar la precisión y la eficacia en la detección de desinformación?

**P8.** ¿Qué estrategias pueden implementarse para mejorar la explicabilidad y la confianza en los sistemas de detección de noticias falsas basados en Inteligencia Artificial?

- a) ¿Cuáles son los desafíos técnicos y éticos en la creación de modelos de inteligencia artificial explicables?
- b) ¿Cómo pueden los usuarios ser educados para interpretar y confiar en los resultados proporcionados por estos sistemas?
- c) ¿Cómo pueden los usuarios ser educados para detectar e identificar las noticias falsas?

## 7. OBJETIVO GENERAL Y OBJETIVOS ESPECÍFICOS

Para validar las hipótesis y responder a las preguntas de investigación formuladas, se presentan a continuación los objetivos que guiarán este estudio.

### 7.1 Objetivo General

Se propone como objetivo general de este proyecto:

Desarrollar un método computacional explicable para la detección de noticias falsas utilizando técnicas de Procesamiento de Lenguaje Natural y Aprendizaje Automático, que integre representaciones textuales,

semánticas, psicológicas y socioculturales de los textos de las noticias, y que mejore la precisión de detección respecto a los métodos reportados en el estado del arte.

## 7.2 Objetivos Específicos

Para lograr el objetivo general, se establecen los siguientes objetivos particulares:

- OE1.** Estudiar el estado del arte sobre:
  - a. Métodos de identificación de noticias falsas.
  - b. Uso y combinación de información multimodal y metadatos para la identificación de noticias falsas.
  - c. Identificación de patrones psicológicos, lingüísticos y de análisis del discurso en noticias falsas para determinar cuáles potencian su divulgación.
- OE2.** Generar u obtener una base de datos que contenga noticias falsas y legítimas.
- OE3.** Identificar la mejor forma de representar y combinar la información presente en los textos de las noticias con el objetivo de maximizar la detección de noticias falsas.
- OE4.** Proponer un método de detección de noticias falsas basado en algoritmos de Aprendizaje **Automatizado** que emplee la representación de datos propuesta.
- OE5.** Desarrollar una aplicación prototipo que integre la representación de datos y el método desarrollado propuestos en objetivos anteriores.
- OE6.** Desarrollar acciones de divulgación y difusión encaminadas a capacitar, concientizar e informar a las personas en la importancia de no ser víctimas camàñas de divulgación de noticias falsas y en técnicas, tecnologías y buenas prácticas para evitar estos fraudes. Esas acciones de divulgación y difusión pueden servir de plataforma para promocionar el prototipo desarrollado. También como parte de este objetivo se desarrollarán artículos que se someterán al criterio de expertos de renombre en el tema para su publicación en revistas de impacto y presentación en conferencias internacionales.

## 8. METODOLOGÍA

Para la realización de este proyecto se propone la siguiente metodología que se compone de 2 etapas anuales, desglosadas en actividades (A) ligadas a los objetivos específicos (OE) antes propuestos.

Etapa 1	<p><b>A1.</b> Revisar el estado del arte de <b>métodos de identificación de noticias falsas</b> y temas asociados, específicamente aquellos que usen técnicas de Inteligencia Artificial e información multimodal. (OE1)</p> <p><b>A2.</b> Revisar el estado del arte de técnicas para <b>representar la información en noticias falsas</b>. (OE1)</p> <p><b>A3.</b> Identificar <b>bases de datos</b> que puedan ser utilizadas en este proyecto. (OE2)</p> <p><b>A4.</b> <b>Recopilar datos</b> necesarios en caso de no encontrar bases de datos que puedan ser útiles. (OE2)</p> <p><b>A5.</b> <b>Representar la información contenida en las noticias falsas</b> utilizando distintas técnicas. (OE3)</p> <p><b>A6.</b> Realizar una primera evaluación de la capacidad de las distintas representaciones para diferenciar entre noticias falsas y legítimas. (OE3)</p>
Etapa 2	<p><b>A7.</b> <b>Seleccionar los algoritmos</b> de Aprendizaje Automático que mejores resultados han demostrado en la identificación de noticias falsas y que serán utilizados en este proyecto. (OE4)</p> <p><b>A8.</b> <b>Evaluar la combinación de distintas representaciones de información</b> y algoritmos de Aprendizaje <b>Automático</b> con el objetivo de determinar la combinación que mejor diferencie entre noticias falsas y legítimas. (OE3 y OE4)</p> <p><b>A9.</b> <b>Comparar el método desarrollado</b> con los resultados reportados en el estado del arte (OE3 y OE4)</p>

- |  |  |
|--|--|
|  | <p><b>A10.</b> Desarrollar una aplicación prototipo que integre el método desarrollado (OE5).</p> <p><b>A11.</b> Participar en eventos de divulgación/difusión donde se presenten los resultados alcanzados (OE6).</p> |
|--|--|

## 9. PRODUCTOS COMPROMETIDOS

De la investigación desarrollada se obtendrán los siguientes productos:

- P1.** Un evento de divulgación con el objetivo de concientizar acerca de los riesgos y las formas de evitar ser víctimas de la desinformación.
- P2.** Presentación en evento de difusión donde se presenten los métodos y técnicas desarrollados.
- P3.** El desarrollo de infografías de buenas prácticas para evitar ser víctima y técnicas de identificación de ataques de phishing basados en los hallazgos obtenidos en este proyecto.
- P4.** Repositorio en línea donde se encuentren el código fuente de los métodos desarrollados. Este repositorio se compartirá bajo la licencia GPL de GNU<sup>43</sup> para que cualquiera que esté interesado pueda utilizarlos, adaptarlos y mejorarlos siempre bajo la condición de que sus adaptaciones sean publicadas bajo la propia licencia GPL de GNU.
- P5.** El prototipo de una herramienta/plugin que integre los métodos desarrollados y que pueda ser utilizado para detectar noticias falsas. Este prototipo deberá ser diseñado e implementado para ser ejecutado en los diferentes sistemas computacionales existentes (dispositivos de escritorio, dispositivos móviles, sobre los diferentes navegadores, etc.).
- P6.** Una participación en una conferencia internacional.
- P7.** Una publicación en una revista de alto impacto indexada en JCR y/o Scopus.

## 10. PERTINENCIA E INCIDENCIA SOCIAL DEL PROYECTO

La proliferación de desinformación se ha convertido en una amenaza grave que compromete la integridad de la información y afecta profundamente diversos aspectos de la vida social y política. La diseminación de noticias falsas no solo distorsiona la opinión pública y manipula la percepción de eventos actuales, sino que también puede influir en la toma de decisiones políticas y sociales de manera adversa. En contextos más críticos, la desinformación puede fomentar conflictos, aumentar la polarización social e incluso incitar a la violencia. En este sentido, la capacidad de identificar y combatir las noticias falsas es esencial para preservar la confianza en los medios de comunicación y en las plataformas digitales, asegurando que el público reciba información precisa y confiable.

El desarrollo de un método computacional ligero y explicable para la detección de noticias falsas responde a esta necesidad urgente al ofrecer una solución que no solo detecta desinformación, sino que también proporciona una comprensión clara de los mecanismos que subyacen a la detección. Esta capacidad de explicación incrementa la transparencia y la confianza en los sistemas de información, al permitir a los usuarios entender cómo se llega a las conclusiones sobre la veracidad de las noticias.

Además, este proyecto representa un avance significativo en el campo del Procesamiento de Lenguaje Natural y el Aprendizaje Automático al integrar un enfoque multidimensional que abarca representaciones textuales, semánticas, psicológicas y socioculturales. Tal integración no solo promueve innovaciones tecnológicas, sino que también mejora las herramientas existentes al ofrecer una capacidad más profunda y matizada para identificar noticias falsas. Este enfoque integral también tiene un impacto positivo en la educación y la conciencia digital, ya que la herramienta desarrollada puede ser utilizada como recurso educativo para elevar el nivel de alfabetización mediática y fomentar un consumo de información más crítico entre los ciudadanos.

<sup>43</sup> <https://www.gnu.org/licenses/licenses.es.html>

Este proyecto también contribuye a un uso más ético y responsable de la inteligencia artificial al abordar la detección de noticias falsas desde una perspectiva que incluye tanto factores contextuales como emocionales. Esta consideración de múltiples dimensiones no solo mejora la precisión en la identificación de desinformación, sino que también refuerza un enfoque equitativo y consciente en el desarrollo de tecnologías de detección. El desarrollo de este proyecto subraya la importancia de mantener la integridad de la información en una era donde la veracidad y la confianza en las fuentes de noticias son más cruciales que nunca.

### 11. CRONOGRAMA DE ACTIVIDADES

Para la realización del proyecto de investigación que se propone en este protocolo, las actividades identificadas se desarrollarán durante 2 años, organizadas como se muestra en el siguiente cronograma. El cronograma también muestra las fechas tentativas para la entrega de los productos comprometidos:

ETAPA 1. AÑO 2025												
Actividad	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
A1												
A2												
A3												
A4												
A5												
A6												
Productos												
P1												
ETAPA 2. AÑO 2026												
Actividad	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
A7												
A8												
A9												
A10												
A11												
Productos												
P2												
P3												
P4												
P5												
P6												
P7												

### 12. INVOLUCRACIÓN DE ESTUDIANTES DE LICENCIATURA Y POSGRADO

En el entendido de que los miembros del equipo de proyecto son académicos de tiempo completo, es necesario solicitar el apoyo de estudiantes que deseen desarrollar sus temas de investigación en los temas que el proyecto aborda. En este sentido se propone involucrar a:

- a) Un estudiante de posgrado con tesis relacionada con el proyecto.
- b) Dos estudiantes de Licenciatura entre 2025 y 2026.

Estos estudiantes realizarán labores de investigación e implementación en conjunto con los demás miembros del equipo de proyecto y serán incluidos como autores de los artículos y demás productos que se



deriven. Cabe mencionar que la investigación y desarrollo realizado por los estudiantes podría ser considerado como tema de investigación de tesis de licenciatura y postgrado.

### 13. ESTRATEGIA PARA LA OBTENCIÓN DE FONDOS EXTERNOS

El proyecto que se propone tiene el objetivo de sentar la base de futuras investigaciones en el área de la Ciencia de Datos aplicada a la Ciberseguridad. Por tal motivo, este proyecto debe ser considerado como la primera de una serie de acciones enfocadas a la obtención de fondos externos que permitan el desarrollo de investigaciones en esta área en del Departameto y la Universidad. En este sentido, se buscará participar en futuras convocatorias de CONACyT así como convocatorias de financiamiento extranjero mediante la colaboración con instituciones de prestigio internacional que desarrollen o tengan el interés de desarrollar investigaciones en el tema que este proyecto aborda. Entre estas instituciones se encuentra el Instituto Tecnológico y de Estudios Superiores de Monterrey mediante su Hub de Ciberseguridad<sup>44</sup> y el Instituto Nacional de Astrofísica, Óptica y Electrónica a través de la Maestría en Ciencias en Tecnologías de Seguridad<sup>45</sup> (perteneciente al PNPC de CONACyT). De igual forma, se buscará un acercamiento a la iniciativa privada en eventos de ciberseguridad como Infosecurity México<sup>46</sup> y Expo Seguridad Industrial<sup>47</sup> entre otros, con el objetivo de definir proyectos conjuntos que permitan expandir y continuar la investigación. También se pretende participar en convocatorias de financiamiento patrocinadas por instituciones internacionales.

### 14. PRESUPUESTO

Para el desarrollo del proyecto propuesto se planea ejecutar los fondos que se entregarán (de resultar seleccionado como ganadores de la convocatoria) de acuerdo con la siguiente tabla:

Rubro	2025	2026	Total
E012. Otros servicios / Honorarios y servicios profesionales.	MX\$ 55,000.00	MX\$ 20,000.00	MX\$ 65.000,00
E016. Servicios, acreditaciones y aportaciones.	MX\$ 25,000.00	MX\$ 30,000.00	MX\$ 55.000,00
E061. Becas de investigación.	MX\$ 80,000.00	MX\$ 80,000.00	MX\$ 160.000,00
E109. Viajes y viáticos.	MX\$ 40,000.00	MX\$ 70,000.00	MX\$ 110.000,00
<b>Total</b>	MX\$ 200,000.00	MX\$ 200,000.00	MX\$ 400.000,00

- E012. Otros servicios / Honorarios y servicios profesionales: Esta subcuenta considera el pago de servicios relacionados con el proceso de recabar los datos necesarios para el proyecto.
- E016. Servicios, acreditaciones y aportaciones: Esta subcuenta considera el pago de inscripción en a una conferencia.
- E061. Becas de investigación: El presupuesto asignado a esta subcuenta se destinará al pago de becas para estudiantes asociados con el proyecto.
- E109. Viajes y viáticos: Este rubro considera el pago de viáticos para participación en congresos nacionales o internacionales.

### 15. ACTIVIDADES DE DIVULGACIÓN

La proliferación de noticias falsas representa un desafío significativo que compromete la integridad de la información y la confianza del público en los medios digitales. La falta de herramientas adecuadas para

<sup>44</sup> <https://tec.mx/es/noticias/nacional/institucion/tec-de-monterrey-crea-hub-para-combatir-la-ciberdelincuencia>

<sup>45</sup> <https://posgrados.inaoep.mx/ts/maestria.php>

<sup>46</sup> <https://www.infosecuritymexico.com/es.html>

<sup>47</sup> <https://www.exposeguridadindustrial.com/es-mx.html>

identificar noticias falsas puede dejar a las personas vulnerables a la desinformación. Por lo tanto, uno de los objetivos clave de este proyecto es proporcionar al público información que le permita identificar y manejar noticias falsas de manera efectiva.

La estrategia de divulgación del proyecto se desarrollará a través de varias actividades clave. En primer lugar, se organizará un evento de divulgación para presentar los hallazgos y la tecnología desarrollada durante el proyecto de manera accesible para el público general. Este evento tiene como objetivo sensibilizar a los participantes sobre los riesgos asociados con las noticias falsas y capacitarlos en el uso de herramientas y técnicas para identificar y gestionar la desinformación. Se incluirán demostraciones interactivas para mostrar cómo funciona el método computacional desarrollado.

En segundo lugar, se elaborarán infografías y materiales de difusión basados en los resultados del proyecto, que ofrecerán una serie de buenas prácticas y recomendaciones para ayudar a los usuarios a identificar y manejar noticias falsas. Estos materiales serán diseñados para ser comprensibles y accesibles, y se distribuirán a través de sitios web, redes sociales y plataformas de noticias.

Además, el proyecto participará en conferencias nacionales e internacionales para presentar los resultados alcanzados, compartir conocimientos y fomentar la discusión sobre la detección de noticias falsas. Esta participación permitirá divulgar los avances y contribuir al debate académico y profesional en el campo.

Finalmente, se escribirán artículos de divulgación sobre los resultados obtenidos, que serán publicados en revistas especializadas y otros medios relevantes. Estos artículos estarán dirigidos a una audiencia más amplia, incluyendo investigadores, profesionales del área y el público en general, con el fin de maximizar el impacto y la utilidad de los hallazgos del proyecto.

Estas actividades de divulgación no solo promoverán una mayor comprensión de la tecnología y los métodos desarrollados, sino que también contribuirán a mejorar la alfabetización mediática del público, permitiendo a los usuarios desarrollar habilidades críticas para evaluar la veracidad de la información en línea.

## **16. CONCLUSIONES**

La creciente preocupación por la desinformación y las noticias falsas subraya la necesidad urgente de desarrollar herramientas efectivas para su detección. Esta preocupación cobra aún más relevancia en la época de la post-verdad que se está viviendo, cuyo principal exponente es Donald Trump, quién formula indiscriminadamente “verdades” según su punto de vista y conveniencia. Esta propuesta de proyecto busca abordar el desafío de la desinformación mediante el desarrollo de un método computacional explicable que integre técnicas avanzadas de Procesamiento de Lenguaje Natural y Aprendizaje Automático, considerando representaciones textuales, semánticas, psicológicas y socioculturales de los textos.

Primero, el enfoque integral del proyecto permitirá una mejora significativa en la precisión de la detección de noticias falsas, superando las limitaciones de los métodos actuales. Al combinar técnicas de representación textual y semántica con análisis psicológicos y socioculturales, el proyecto busca ofrecer una solución robusta y versátil que pueda adaptarse a la evolución de las estrategias de desinformación.

Además, la inclusión de técnicas de explicabilidad en el modelo desarrollado es un componente crucial que facilitará la transparencia y la comprensión de los resultados por parte de los usuarios finales. Esta característica no solo fortalecerá la confianza en el sistema de detección, sino que también contribuirá a la aceptación y adopción generalizada de la herramienta.

Las actividades de divulgación propuestas, que incluyen eventos educativos, materiales informativos y la participación en conferencias, desempeñarán un papel fundamental en la sensibilización del público y en la promoción de buenas prácticas para la identificación de noticias falsas. La publicación de artículos y la

presentación en foros académicos y profesionales ampliarán el alcance del proyecto y fomentarán el intercambio de conocimientos en la comunidad científica y más allá.

Este proyecto no solo representa un avance significativo en la detección de noticias falsas, sino que también aborda aspectos críticos de transparencia, educación y concienciación. Al implementar un enfoque multidimensional y explicable, se espera que la herramienta desarrollada tenga un impacto positivo en la integridad de la información, la alfabetización mediática y la confianza pública en los sistemas digitales.