FUNDAMENTOS Y APLICACIONES DE INTELIGENCIA ARTIFICIAL

Derecho de la era digital:
Del ciudadano (y sus derechos fundamentales) a la Inteligencia Artificial

Summer School Derecho Digital

Fecha de actualización: 29th May 2025

Prof. Dr. José Luis Salmerón

BY NC SA

Escuela Politécnica Superior
CUNEF Universidad

1. Introducción a la Inteligencia Artificial

2. Tipos de Inteligencia Artificial

3. Fundamentos de IA

4. Aplicaciones de la IA

5. Limitaciones y Desafíos (Funcionales)

6. El Futuro de la IA

¿QUÉ ES LA INTELIGENCIA ARTIFICIAL (IA)?

Definición Conceptual

La Inteligencia Artificial (IA) es un campo de la informática dedicado a la creación de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana. Puntos clave

- Simulación de la inteligencia humana: No se trata de replicar la conciencia, sino las capacidades cognitivas como aprender, razonar, resolver problemas, percibir y comprender el lenguaje.
- Sistemas y agentes: La IA se manifiesta en software (programas) y hardware (robots) que interactúan con su entorno.
- Tareas específicas vs. generales: Hay IA diseñada para tareas muy concretas y la idea (aún lejana) de una IA con capacidades humanas generales.

Enfoque de esta sesión

Nos centraremos en el cómo funciona y qué puede hacer la IA, no en las implicaciones legales (ya que entender su funcionamiento es clave para análisis jurídicos posteriores).

UN POCO DE HISTORIA: HITOS CLAVE

- Años 40-50: Los Inicios
  - Alan Turing y la Máquina de Turing (concepto de computación universal).
  - Test de Turing (¿Puede una máquina exhibir un comportamiento inteligente indistinguible del de un humano?).
  - Conferencia de Dartmouth (1956): Acuñación del término Inteligencia Artificial. Primeros programas de IA (Logic Theorist, General Problem Solver).
- Años 60-70: Entusiasmo y primeros inviernos
  - Desarrollo de LISP (lenguaje de programación para IA).
  - Primeros chatbots (ELIZA).
  - Limitaciones computacionales y teóricas llevan a recortes de financiación (invierno de la IA).
- Años 80: Sistemas expertos
  - Auge de los sistemas basados en conocimiento, capaces de emular la toma de decisiones de un experto humano en un dominio específico.
- Años 90-2000: Resurgimiento y Machine Learning
  - Aumento de la capacidad computacional y disponibilidad de datos.
  - Enfoques estadísticos y el Machine Learning (Aprendizaje Automático) ganan tracción.
  - Hitos como Deep Blue venciendo a Kasparov en ajedrez (1997).
- 2010-Presente: Deep Learning y la Edad de Oro
  - Grandes avances gracias al Deep Learning (Aprendizaje Profundo), redes neuronales con muchas capas.
  - Explosión de aplicaciones: reconocimiento de voz, visión por computadora, traducción automática, vehículos autónomos, IA generativa.

OBJETIVOS DE LA IA

Objetivos a Corto/Medio Plazo (Prácticos):

- Automatizar tareas repetitivas: Liberar a los humanos de trabajos monótonos o peligrosos.
- Resolver problemas complejos: Analizar grandes cantidades de datos para encontrar patrones o soluciones que un humano no podría.
- Mejorar la toma de decisiones: Proporcionar información y recomendaciones basadas en datos.
- Crear nuevas herramientas y servicios: Desde asistentes virtuales hasta diagnósticos médicos asistidos por IA.
- Personalización: Adaptar productos y servicios a las necesidades individuales.

Objetivos a Largo Plazo (Más Teóricos/Ambiciosos):

- Comprender la inteligencia: Usar la IA como herramienta para entender mejor cómo funciona la inteligencia (humana y artificial).
- Alcanzar la Inteligencia Artificial General (AGI): Crear máquinas con la misma capacidad intelectual que un ser humano en cualquier tarea. (¡Aún muy lejos!)
- Superinteligencia (ASI): Una IA que supere significativamente la inteligencia humana en todos los aspectos. (Tema de mucho debate y especulación).

Importante

La mayoría de la IA que vemos hoy en día es IA Estrecha o Débil, diseñada para tareas específicas.

1. Introducción a la Inteligencia Artificial

2. Tipos de Inteligencia Artificial

3. Fundamentos de IA

4. Aplicaciones de la IA

5. Limitaciones y Desafíos (Funcionales)

6. El Futuro de la IA

CLASIFICACIÓN DE LA IA: POR CAPACIDAD

IA Estrecha o Débil (Artificial Narrow Intelligence - ANI)

- Definición: IA especializada en una tarea específica o un conjunto limitado de tareas.
- Características:
  - Opera bajo un rango predefinido y limitado.
  - No posee conciencia ni autoconciencia.
  - Es la forma de IA que existe y se utiliza ampliamente hoy en día.
- Ejemplos Funcionales:
  - Sistemas de recomendación (Netflix, Amazon).
  - Software de reconocimiento facial.
  - Asistentes virtuales (Siri, Alexa) para tareas específicas.
  - Software de ajedrez o Go.
  - Filtros de spam en correos electrónicos.
  - Algoritmos de trading en bolsa.

CLASIFICACIÓN DE LA IA: POR CAPACIDAD (CONTINUACIÓN)

IA General (Artificial General Intelligence - AGI)

- Definición: Tipo de IA hipotético que poseería la capacidad intelectual de un ser humano, pudiendo aprender y realizar cualquier tarea intelectual que un humano pueda.
- Características:
  - Capacidad de razonar, planificar, resolver problemas de forma abstracta, comprender ideas complejas y aprender de la experiencia.
  - Podría transferir conocimiento de un dominio a otro.
  - Aún no existe; es un objetivo de investigación a largo plazo.
- Implicaciones (Conceptuales):
  - Si se lograra, tendría un impacto transformador en la sociedad.
  - Plantea profundas cuestiones filosóficas y éticas.

Superinteligencia Artificial (Artificial Superintelligence - ASI)

- Definición: IA hipotética que superaría significativamente la inteligencia humana en prácticamente todos los dominios relevantes.
- Características: Sería capaz de un rendimiento cognitivo muy superior al de los humanos más brillantes.
- Estado: Puramente especulativo y tema de debate entre expertos sobre su viabilidad y posibles consecuencias.

CLASIFICACIÓN DE LA IA: POR FUNCIONALIDAD

Máquinas Reactivas

- Definición: El tipo más básico de IA. No tienen memoria ni capacidad para usar experiencias pasadas para informar decisiones futuras.
- Funcionamiento: Reaccionan a estímulos actuales de forma idéntica cada vez. No aprenden.
- Ejemplo Clásico: Deep Blue, el programa de ajedrez de IBM que venció a Garry Kasparov. Podía identificar las piezas en el tablero y conocer las reglas, predecir los movimientos del oponente y elegir el movimiento óptimo, pero no recordaba partidas anteriores más allá del juego actual.

Memoria Limitada

- Estas máquinas usan memoria para tomar decisiones, pero esta memoria es temporal o de corto plazo.
- La información pasada se utiliza para tomar decisiones en el presente inmediato, pero no se almacena como parte de una biblioteca de experiencias a largo plazo. La mayoría de la IA actual cae en esta categoría.
- Ejemplos Funcionales:
  - Vehículos autónomos: Observan la velocidad y dirección de otros coches, señales de tráfico recientes, etc., para tomar decisiones de conducción. Esta información no se guarda indefinidamente.
  - Muchos chatbots y asistentes virtuales recuerdan una conversación para mantener el contexto.

CLASIFICACIÓN DE LA IA: POR FUNCIONALIDAD (CONTINUACIÓN)

Teoría de la Mente (Aún no desarrollada)

- IA que tendría la capacidad de comprender que los seres humanos (y otras entidades) tienen pensamientos, sentimientos, creencias, intenciones y expectativas que afectan su comportamiento.
- Funcionamiento (hipotético): Sería capaz de inferir estados mentales y predecir cómo se comportarán los demás. Implica un nivel de comprensión social y emocional.
- Es un área activa de investigación, pero aún no se ha logrado una IA con verdadera Teoría de la Mente.

Autoconciencia (Aún no desarrollada)

- El máximo de la IA, donde las máquinas tendrían una conciencia de sí mismas, un sentido de ser. Serían conscientes de su propia existencia y de sus estados internos.
- Funcionamiento (hipotético): Similar a la conciencia humana.
- Es el tipo de IA más avanzado y actualmente solo existe en la ciencia ficción. No sabemos si es posible ni cómo se lograría.

Recordatorio

Las categorías Teoría de la Mente y Autoconciencia son teóricas y representan fronteras futuras de la investigación en IA, no capacidades actuales.

1. Introducción a la Inteligencia Artificial

2. Tipos de Inteligencia Artificial

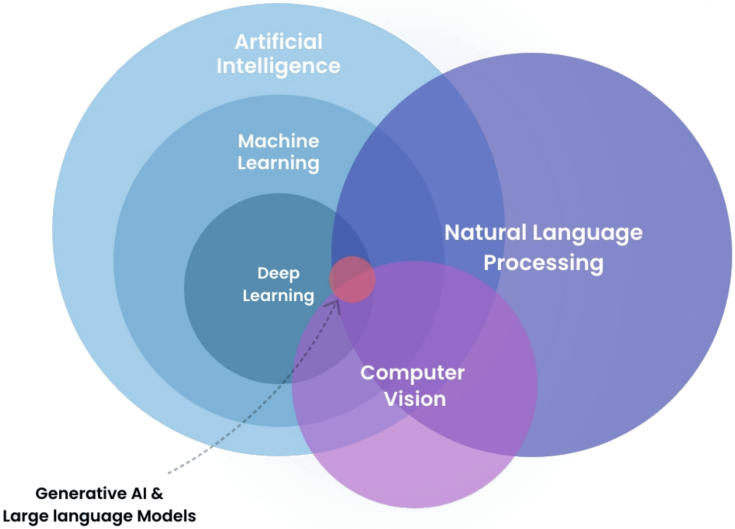
3. Fundamentos de IA

4. Aplicaciones de la IA

5. Limitaciones y Desafíos (Funcionales)

6. El Futuro de la IA

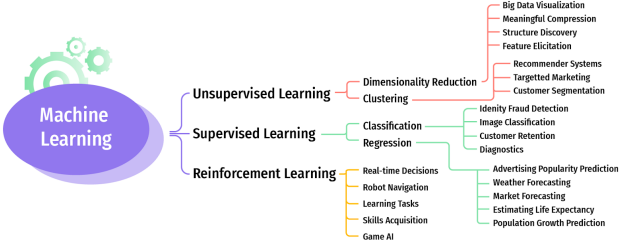
AI LANDSCAPE

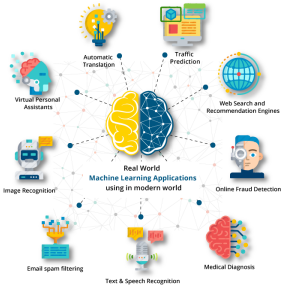


APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

El Aprendizaje Automático (Machine Learning) se ocupa de algoritmos y modelos que permiten a las máquinas aprender y mejorar automáticamente a partir de datos sin ser programadas explícitamente.

- Gran capacidad para procesar grandes volúmenes de datos y extraer información valiosa.
- Aplicaciones: Medicina y sanidad, automatización industrial, transporte inteligente, logística, finanzas, consultoría legal, análisis de mercado y potencialmente cualquier actividad humana.

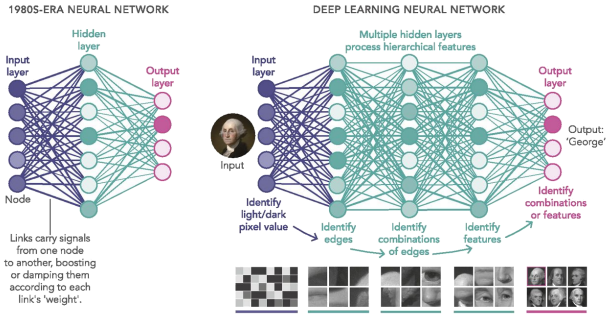




REDES NEURONALES SUPERFICIALES Y PROFUNDAS (NEURAL NETWORKS)

Son modelos computacionales inspirados en el cerebro humano. Se componen nodos interconectados (neuronas artificiales) organizados en capas, que procesan información y aprenden ajustando sus conexiones a partir de datos.

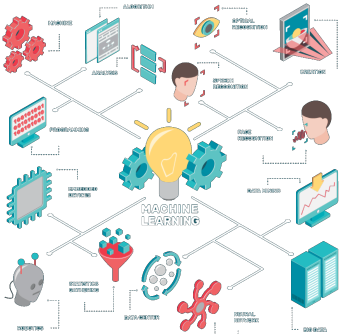
- Gran capacidad para aprender patrones intrincados y relaciones no lineales directamente de los datos, superando a menudo a los modelos tradicionales en tareas complejas.
- El Aprendizaje Profundo permite arquitecturas con múltiples capas que logran representaciones jerárquicas de la información.
- Versatiles para una amplia gama de tipos de datos, incluyendo no estructurados como imágenes, audio y texto.



REDES NEURONALES (APLICACIONES)

Aplicaciones destacadas:

- Visión Artificial: Reconocimiento y clasificación de imágenes, detección de objetos (ej. en análisis de documentos legales escaneados, control de calidad en la industria).
- Procesamiento del Lenguaje Natural (PLN): Traducción automática, análisis de sentimientos en opiniones de clientes, chatbots para asistencia legal o de negocios, generación de texto.
- Reconocimiento de Voz: Asistentes virtuales, transcripción de audio a texto (ej. de declaraciones o reuniones).
- Finanzas y Negocios: Detección de fraude en transacciones, predicción de mercados, scoring crediticio avanzado, sistemas de recomendación personalizados.
- Medicina y salud: Diagnóstico asistido por imágenes médicas, descubrimiento de fármacos.

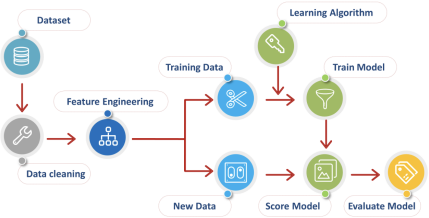


APRENDIZAJE SUPERVISADO

El Aprendizaje Supervisado usa datos etiquetados (pares entrada-salida) para aprender. El modelo intenta encontrar patrones en los datos para realizar predicciones precisas sobre datos nuevos y no vistos.

Regresión: Predecir un valor numérico continuo.

- Ejemplo: Predecir la probabilidad de éxito de un caso legal basado en sus características, el precio futuro de una acción, o la demanda esperada de un nuevo producto.
- Por qué es útil: Optimiza la toma de decisiones estratégicas, la asignación de recursos y la gestión de riesgos.



Clasificación: Asignar puntos de datos a categorías o clases discretas.

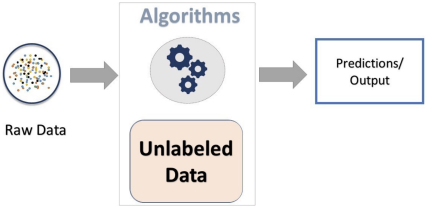
- Ejemplo: Clasificar documentos legales por tipo (contrato, demanda, sentencia) o identificar transacciones financieras como fraudulentas o legítimas.
- Por qué es útil: Automatiza tareas y optimiza las estrategias de retención de clientes o cumplimiento normativo.

APRENDIZAJE NO SUPERVISADO: DESCUBRIENDO ESTRUCTURAS OCULTAS

Los algoritmos trabajan con datos no etiquetados para descubrir estructuras ocultas, patrones o relaciones sin conocimiento previo de la salida. Se trata de encontrar un orden inherente en los datos.

Clustering (Agrupamiento): Agrupar puntos de datos similares en clústeres distintos.

- Ejemplo: Segmentar automáticamente a los clientes basándose en su comportamiento de compra o navegación para campañas de marketing personalizadas. En el ámbito legal, agrupar casos judiciales con características similares para análisis de precedentes o tendencias jurisprudenciales.
- Por qué es útil: Revela relaciones desconocidas, simplifica la exploración de datos y ayuda a categorizar conjuntos de datos complejos provenientes de interacciones con clientes o grandes volúmenes de textos legales.

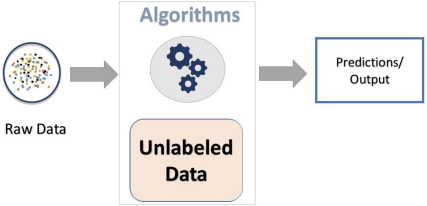


APRENDIZAJE NO SUPERVISADO: DESCUBRIENDO ESTRUCTURAS OCULTAS

Los algoritmos trabajan con datos no etiquetados para descubrir estructuras ocultas, patrones o relaciones sin conocimiento previo de la salida. Se trata de encontrar un orden inherente en los datos.

Reducción de Dimensionalidad: Simplificar datos de alta dimensión preservando la información esencial.

- Ejemplo: Analizar grandes conjuntos de datos de encuestas de mercado con muchas preguntas para identificar los pocos factores clave que realmente impulsan la preferencia del consumidor. En el ámbito legal, reducir el número de variables en el análisis de grandes carteras de litigios para una mejor visualización de riesgos.
- Por qué es útil: Hace que los datos complejos sean más manejables, reduce la carga computacional para análisis posteriores y ayuda a visualizar información de alta dimensión.



APRENDIZAJE POR REFUERZO

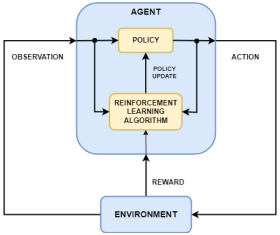
Un agente aprende a tomar decisiones secuenciales en un entorno para maximizar una recompensa acumulada. Opera mediante prueba y error, obteniendo retroalimentación de sus acciones para optimizar su estrategia.

Optimización de estrategias dinámicas:

- Ejemplo: En negociación de contratos, un sistema que aprende la secuencia óptima de ofertas y contraofertas.
- Por qué es útil: Permite la adaptación dinámica a las condiciones cambiantes del mercado o de la contraparte y maximiza la eficiencia de procesos automatizados.

Optimización de procesos de negocio:

- Ejemplo: Optimizar las políticas de precios dinámicos en comercio electrónico para maximizar ingresos.
- Por qué es útil: Reduce la intervención humana en tareas repetitivas de optimización.



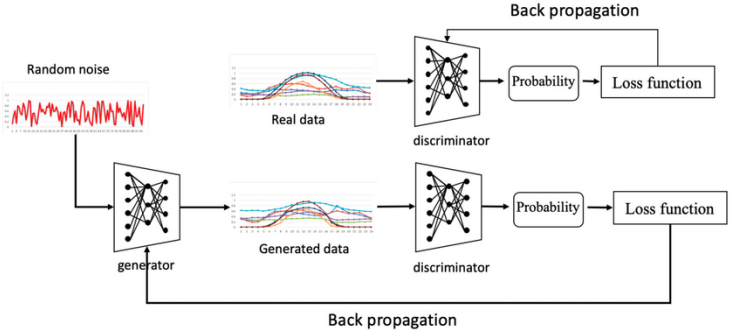
Sistemas de recomendación adaptativos:

- Ejemplo: En plataformas legales, sugiere jurisprudencia relevante de forma adaptativa durante una búsqueda.
- Por qué es útil: Mejora la experiencia del usuario, y puede guiar hacia la información más pertinente de manera eficiente.

REDES GENERATIVAS (GANS)

Autoencoders y GANs: Redes neuronales generativas que pueden crear nuevos datos que se asemejan a los datos de entrenamiento.

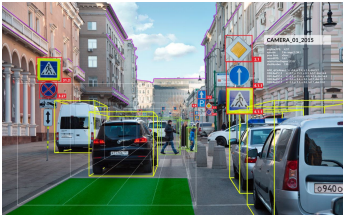
- Generación de nuevos ejemplos de cláusulas contractuales o escenarios de negociación con propiedades deseadas.
- Creación de datos sintéticos de transacciones o perfiles de clientes para aumentar los conjuntos de datos de entrenamiento sin comprometer la privacidad.

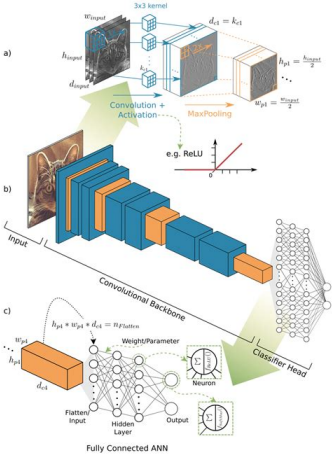


VISIÓN ARTIFICIAL

La Visión Artificial es una rama de la IA centrada en el desarrollo de algoritmos y técnicas que permiten a las máquinas comprender y analizar imágenes y videos.

- Permite a las máquinas percibir y comprender el mundo visual de forma similar a los humanos.
- Aplicaciones: Detección de objetos, seguimiento de movimiento, reconocimiento facial (ej. para control de acceso o identificación), vehículos autónomos, análisis de imágenes en logística, control de calidad en manufactura, entre otros.

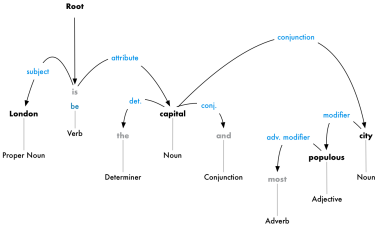




PROCESAMIENTO DEL LENGUAJE NATURAL (NLP)

Son técnicas de IA para comprender, interpretar y generar lenguaje humano. Esencial para:

- Extracción de información: Analizar vasta literatura jurídica (artículos, sentencias, contratos) o informes de negocio para descubrir relaciones ocultas, extraer entidades clave (nombres, fechas, importes) y hechos relevantes.
- Construcción de bases de datos: Automatizar la creación de bases de datos estructuradas a partir de texto no estructurado (ej. correos electrónicos, informes), acelerando la organización de la información para su posterior análisis.
- Generación de hipótesis y resúmenes: Identificar tendencias emergentes o posibles riesgos analizando grandes volúmenes de texto. Resumir automáticamente documentos largos.
- Clasificación y chatbots: Clasificar automáticamente documentos legales por área o tipo. Desarrollar asistentes virtuales para consultas legales básicas o atención al cliente.

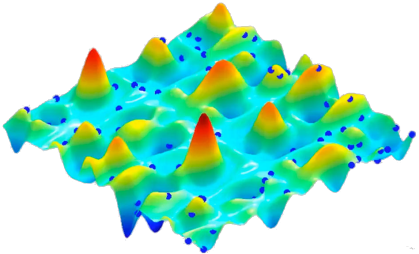


SWARM INTELLIGENCE

Un paradigma computacional inspirado en el comportamiento colectivo de sistemas descentralizados y autoorganizados (ej. bandadas de pájaros, colonias de hormigas). Es adecuado para problemas complejos de optimización. Optimización por Enjambre de Partículas (PSO): Una técnica donde un enjambre de partículas explora un espacio de búsqueda, ajustando sus trayectorias basándose en su mejor posición encontrada y la mejor posición encontrada por todo el enjambre.

Aplicaciones

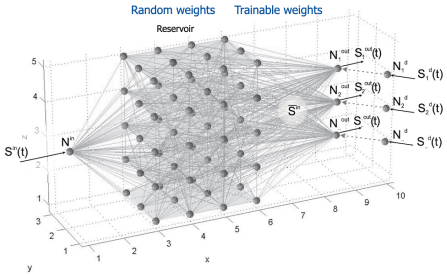
- Optimización de carteras de inversión: PSO puede optimizar la asignación de activos en una cartera para maximizar el retorno esperado para un nivel de riesgo dado, o minimizar el riesgo para un retorno objetivo.
- Optimización de rutas y logística: Encontrar las rutas más eficientes para cadenas de suministro, reparto de mercancías o asignación de recursos móviles.
- Ajuste de parámetros en modelos: PSO puede ayudar a encontrar los parámetros óptimos para modelos de predicción financiera o de comportamiento del consumidor, mejorando su precisión.
- Asignación de recursos: Optimizar la asignación de personal a tareas o proyectos en una empresa, o la distribución de recursos en campañas de marketing.



RESERVOIR COMPUTING

Reservoir Computing (RC) es un paradigma de red neuronal caracterizado por una red neuronal recurrente conectada aleatoriamente (el reservorio) que permanece fija durante el entrenamiento. Solo se entrena la capa de salida.

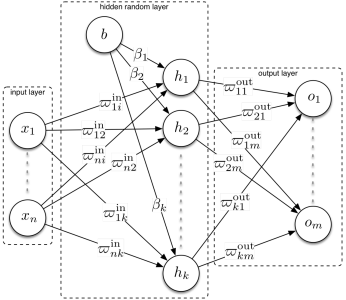
- Núcleo de Aprendizaje Superficial: La extracción de características complejas y no lineales ocurre automática e implícitamente dentro del reservoir fijo. El aprendizaje (optimización de pesos) ocurre únicamente en la capa de final.
- Eficiencia y simplicidad: Simplifica drásticamente el proceso de entrenamiento en comparación con las redes neuronales recurrentes (RNN) tradicionales, ya que solo es necesario ajustar los pesos de salida. Evita los problemas de gradientes que se desvanecen/explotan comunes en el entrenamiento de RNN profundas.



MÁQUINAS DE APRENDIZAJE EXTREMO (ELMS)

Son un tipo de red neuronal prealimentada para clasificación, regresión, clustering, aproximación dispersa y aprendizaje de características. Son una versión simplificada de las redes de una sola capa oculta.

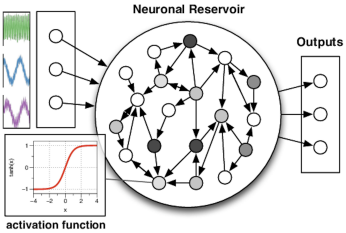
- Capa oculta aleatoria: A diferencia de las redes convencionales donde todos los pesos se ajustan, las ELMs asignan aleatoriamente los pesos de entrada y los sesgos de la capa oculta. Estos parámetros quedan fijos.
- Solución analítica de la capa de salida: La única parte que se entrena son los pesos de la capa de salida. Se hace a través de un solo paso de inversa generalizada (pseudoinversa de Moore-Penrose), en lugar de un descenso de gradiente iterativo.
- Velocidad y eficiencia: Este enfoque de aprenderlo todo de una vez hace que las ELMs sean muy rápidas de entrenar, a menudo órdenes de magnitud más rápidas que la retropropagación, mientras siguen logrando un rendimiento de generalización competitivo.



REDES DE ESTADO DE ECO (ECHO STATE NETWORKS - ESNS)

Las Redes de Estado de Eco (ESN) son un tipo específico de Red Neuronal Recurrente que se enmarcan en el paradigma más amplio de Reservoir Computing (RC). Están diseñadas para procesar datos secuenciales.

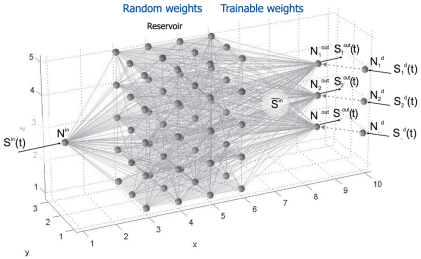
- Capa recurrente fija (reservoir): El núcleo es una red neuronal recurrente conectada de forma dispersa e inicializada aleatoriamente. Las conexiones dentro de este reservorio son fijas después de la inicialización.
- Propiedad de Eco State: Esta propiedad asegura que el estado interno del reservorio sea una memoria evanescente del historial de entrada, lo que implica que la influencia de las entradas pasadas decae con el tiempo.
- Entrenamiento de la capa de salida: Solo se entrenan los pesos que conectan los estados del reservorio con la capa de salida (usando pseudoinversa), lo que hace que las ESNs sean rápidas y eficientes.



MÁQUINAS DE ESTADO LÍQUIDO (LIQUID STATE MACHINES, LSMS)

Las Máquinas de Estado Líquido (LSMs) son un tipo específico de red neuronal recurrente (RNN). Son particularmente adecuadas para datos temporales (que varían en el tiempo).

- Reservoir bio-inspirado: El líquido se refiere a una red neuronal recurrente compleja y conectada aleatoriamente (spiking neurons) que actúa como un reservoir dinámico que procesa las señales entrantes y las transforma en una representación de alta dimensionalidad y variable en el tiempo, similar a cómo un cerebro biológico procesa la información.
- Propiedad de Eco State (implícita): Las LSMs también tienen esta propiedad. Esto asegura que el estado actual proporcione un historial rico e informativo de la señal de entrada.
- Capa de lectura: Al igual que con otros modelos de RC, las conexiones dentro del reservoir líquido son fijas. Solo se entrena una capa de salida simple para mapear los estados del reservorio de alta dimensionalidad a la salida deseada.



1. Introducción a la Inteligencia Artificial

2. Tipos de Inteligencia Artificial

3. Fundamentos de IA

4. Aplicaciones de la IA

5. Limitaciones y Desafíos (Funcionales)

6. El Futuro de la IA

PROCESAMIENTO DEL LENGUAJE NATURAL (NLP)

Tareas y Funcionalidades Comunes del NLP:

- Traducción Automática: (Ej. Google Translator)
  - Cómo funciona (conceptual): Modelos (a menudo basados en Deep Learning como Transformers) aprenden alineaciones y patrones entre idiomas a partir de grandes corpus de textos traducidos.
- Análisis de Sentimientos: Determinar la actitud o emoción expresada en un texto (positiva, negativa, neutral).
  - Cómo funciona (conceptual): Se entrena un clasificador con textos etiquetados con su sentimiento. El modelo aprende palabras y frases asociadas a diferentes sentimientos.
  - Aplicación funcional: Monitorizar la opinión pública sobre un producto o marca en redes sociales.
- Chatbots y asistentes virtuales: (Ej. Siri, Alexa, ChatGPT)
  - Cómo funcionan (conceptual): Combinan NLP para entender la pregunta del usuario (comprensión del lenguaje natural - NLU) y para generar una respuesta coherente (generación de lenguaje natural - NLG). Los modelos más avanzados (como los LLMs) generan texto prediciendo la siguiente palabra más probable basándose en el contexto.
- Reconocimiento de Voz (Speech-to-Text): Convertir audio hablado en texto.
- Síntesis de Voz (Text-to-Speech): Convertir texto en audio hablado.
- Extracción de información: Identificar y extraer datos específicos de documentos (ej. nombres, fechas, organizaciones de un contrato).
- Resumen automático de Textos.
- Clasificación de documentos.

MODELOS DE LENGUAJE GRANDES (LLM)

¿Qué son?

Los LLMs, como GPT (Generative Pre-trained Transformer), son un tipo avanzado de modelo de PLN basado en la arquitectura Transformer de Deep Learning. Están pre-entrenados con cantidades masivas de datos de texto (libros, artículos, sitios web).

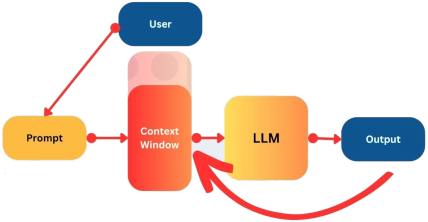
Características funcionales clave:

- Generación de texto coherente y contextual: Su principal habilidad es predecir la siguiente palabra (o token) en una secuencia, lo que les permite generar texto que suena natural y es relevante al contexto proporcionado (el prompt).
- Comprensión del contexto extenso: Pueden mantener el hilo de conversaciones largas y considerar información previa.
- Capacidad de Few-Shot o Zero-Shot Learning: A menudo pueden realizar tareas para las que no fueron explícitamente entrenados con solo unos pocos ejemplos (few-shot) o incluso sin ejemplos (zero-shot), simplemente describiendo la tarea en el prompt.
- Versatilidad: Pueden usarse para una amplia gama de tareas de PLN: redacción, resumen, traducción, respuesta a preguntas, generación de código, etc.

MODELOS DE LENGUAJE GRANDES (LLMS)

Cómo funcionan (Conceptual - muy simplificado)

1. Pre-entrenamiento: Aprende patrones del lenguaje, gramática, conocimiento general, etc., a partir de la ingesta masiva de texto. Aprende a predecir palabras faltantes o la siguiente palabra en una frase.
2. Fine-tuning: Se puede entrenar adicionalmente el modelo pre-entrenado con un conjunto de datos más pequeño y específico para una tarea particular (ej., responder preguntas médicas, escribir en un estilo particular).
3. Inferencia: Con un prompt, el modelo genera una respuesta palabra por palabra, eligiendo la palabra más probable según lo aprendido y el contexto actual.



Importante

Aunque los LLMs son muy potentes, no entienden en el sentido humano. Son sistemas de reconocimiento y generación de patrones extremadamente sofisticados. Pueden alucinar o generar información incorrecta.

VISIÓN ARTIFICIAL

Tareas y Funcionalidades Comunes:

- Clasificación de Imágenes: Asignar una etiqueta a una imagen (ej. gato, coche, paisaje).
  - Cómo funciona (conceptual): Redes Neuronales Convolucionales (CNNs), un tipo de red de Deep Learning, son muy efectivas. Aprenden a detectar características visuales (bordes, texturas, formas) en diferentes niveles de abstracción.
- Detección de Objetos: Localizar y clasificar múltiples objetos dentro de una imagen (dibujando cajas delimitadoras alrededor de ellos).
  - Aplicación funcional: Coches autónomos identificando otros vehículos, peatones, señales.
- Segmentación de Imágenes: Dividir una imagen en regiones que corresponden a diferentes objetos o partes de objetos (más detallado que la detección).
- Reconocimiento Facial: Identificar o verificar la identidad de una persona a partir de una imagen o video de su rostro.
- Análisis de Video: Seguir objetos, reconocer actividades o eventos en secuencias de video.
- Generación de Imágenes: Crear nuevas imágenes (ej. DALL-E, Stable Diffusion). Estos modelos (a menudo basados en GANs o modelos de difusión) aprenden la distribución de los datos de imágenes de entrenamiento y pueden generar nuevas muestras.
- Reconstrucción 3D: Crear modelos 3D a partir de imágenes 2D.

OTRAS APLICACIONES FUNCIONALES RELEVANTES

- Sistemas de Recomendación:
  - Funcionamiento: Utilizan algoritmos (filtrado colaborativo, basado en contenido, híbridos) para predecir las preferencias de un usuario y sugerir ítems (películas, productos, noticias).
  - Aprenden de tu comportamiento pasado y del comportamiento de usuarios similares.
- Robótica Inteligente:
  - Funcionamiento: Robots que utilizan IA (visión artificial, NLP, aprendizaje por refuerzo) para percibir su entorno, tomar decisiones y actuar de forma autónoma o semi-autónoma.
  - Ejemplos: Robots industriales colaborativos (cobots), robots de limpieza, drones autónomos.
- Detección de Fraude y Anomalías:
  - Funcionamiento: Modelos de ML (supervisados y no supervisados) entrenados para identificar patrones inusuales o sospechosos en grandes volúmenes de transacciones o datos.
  - Aplicaciones: Fraude bancario, ciberseguridad, detección de defectos en manufactura.
- IA en la Investigación Científica:
  - Funcionamiento: Análisis de grandes conjuntos de datos experimentales, descubrimiento de fármacos, modelado de sistemas complejos (clima, genómica).
- IA Generativa (más allá de texto e imágenes):
  - Funcionamiento: Modelos que pueden generar música, código de software, diseños, etc.
  - Se basan en aprender la estructura subyacente de los datos de entrenamiento y luego generar nuevas muestras.

DATOS: EL COMBUSTIBLE DE LA IA

Los modelos de IA son tan buenos como los datos con los que se entrenan.

- Cantidad: Generalmente, más datos (de buena calidad) llevan a mejores modelos, especialmente para Deep Learning.
- Calidad:
  - Precisión: ¿Son correctos los datos y las etiquetas? Datos incorrectos enseñarán cosas incorrectas al modelo.
  - Completitud: ¿Faltan muchos valores?
  - Consistencia: ¿Están los datos en un formato uniforme?
  - Relevancia: ¿Son los datos pertinentes para el problema que se intenta resolver?
- Representatividad:
  - ¿Reflejan los datos de entrenamiento la diversidad del mundo real donde se aplicará el modelo?
  - Si no, el modelo puede funcionar bien en los datos de entrenamiento pero mal en la práctica.
  - Sesgos en los datos (Bias): Si los datos de entrenamiento contienen sesgos históricos o sociales (ej. subrepresentación de ciertos grupos), el modelo aprenderá y probablemente amplificará esos sesgos. Esto tiene implicaciones funcionales y éticas muy serias.

1. Introducción a la Inteligencia Artificial

2. Tipos de Inteligencia Artificial

3. Fundamentos de IA

4. Aplicaciones de la IA

5. Limitaciones y Desafíos (Funcionales)

6. El Futuro de la IA

LIMITACIONES Y DESAFÍOS FUNCIONALES DE LA IA ACTUAL

A pesar de los avances, la IA actual tiene limitaciones importantes:

- Dependencia de Grandes Cantidades de Datos:
  - Muchos modelos de vanguardia (especialmente Deep Learning) requieren enormes conjuntos de datos etiquetados, que pueden ser costosos y difíciles de obtener.
  - Problemas con pocos datos (few-shot learning) o datos desequilibrados son un desafío.
- Fragilidad y Falta de Robustez (Brittleness):
  - Los modelos pueden fallar de manera inesperada si la entrada es ligeramente diferente a los datos de entrenamiento (ej. ataques adversariales, donde pequeños cambios imperceptibles en una imagen hacen que un modelo la clasifique incorrectamente).
  - No tienen sentido común ni una comprensión profunda del mundo.
- Problemas de Generalización:
  - Un modelo entrenado en un contexto puede no funcionar bien en otro contexto ligeramente diferente (problema de transferencia de dominio).
- Consumo Energético y Computacional:
  - Entrenar modelos grandes (como LLMs) consume mucha energía y requiere hardware especializado, lo que plantea cuestiones de sostenibilidad y acceso.
- Falta de Sentido Común:
  - La IA actual opera principalmente mediante el reconocimiento de patrones y correlaciones, no mediante una comprensión causal o razonamiento de sentido común similar al humano.
  - Pueden dar respuestas absurdas o ilógicas en situaciones no previstas en su entrenamiento.

EL PROBLEMA DE LA CAJA NEGRA (EXPLAINABILITY - XAI)

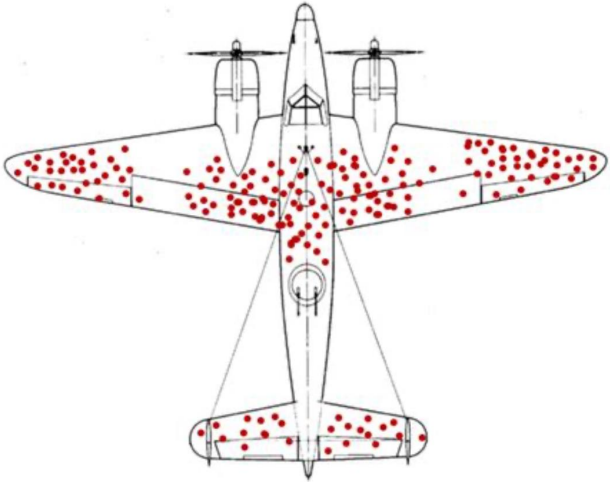
Muchos modelos de IA avanzados, especialmente las redes neuronales profundas, funcionan como cajas negras. Esto significa que:

- Pueden hacer predicciones muy precisas.
- Pero es muy difícil (o imposible) entender por qué tomaron una decisión específica o cómo llegaron a una conclusión particular.
- Los mecanismos internos son complejos y no transparentes para los humanos.

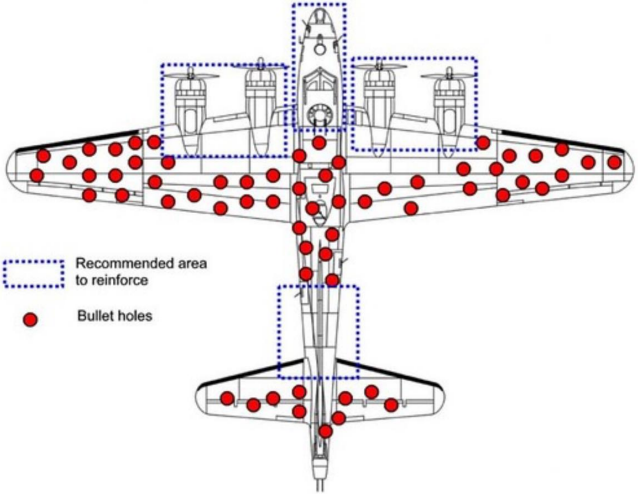
Importancia Funcional de la Explicabilidad:

- Confianza y Adopción: Los usuarios (especialmente en dominios críticos como medicina, finanzas, justicia) son reacios a confiar en decisiones que no pueden entender.
- Depuración y Mejora: Si un modelo comete un error, es difícil diagnosticar y corregir el problema si no se entiende su razonamiento.
- Detección de Sesgos: Un modelo explicable puede ayudar a revelar si está utilizando características inapropiadas o sesgadas para tomar decisiones.
- Cumplimiento Normativo (Relevancia Indirecta para Juristas): En algunas áreas, puede haber requisitos (futuros) de que las decisiones algorítmicas sean explicables.

SESGOS (BIAS) EN LA IA: UN PROBLEMA FUNCIONAL Y SOCIAL



SESGOS (BIAS) EN LA IA: UN PROBLEMA FUNCIONAL Y SOCIAL



SESGOS (BIAS) EN LA IA: UN PROBLEMA FUNCIONAL Y SOCIAL

Un sistema de IA se considera sesgado si sus resultados son sistemáticamente perjudiciales o injustos para ciertos individuos o grupos.

Impacto Funcional de los Sesgos:

- Rendimiento Desigual: El modelo puede funcionar significativamente peor para ciertos subgrupos.
- Decisiones Injustas o Discriminatorias: Aunque el objetivo no sea discriminar, el resultado puede serlo.
- Pérdida de Confianza: Si se percibe que un sistema es injusto, su adopción y utilidad se ven mermadas.
- Resultados Contraintuitivos o Erróneos: El sesgo puede llevar a que el modelo se enfoque en correlaciones espurias en lugar de las verdaderas señales.

Mitigación (Conceptual):

- Curación cuidadosa de datos, técnicas de preprocesamiento para equilibrar representaciones.
- Algoritmos de fairness-aware machine learning (aprendizaje automático consciente de la equidad).
- Auditoría y pruebas rigurosas para detectar y medir sesgos.
- Equipos de desarrollo diversos.

1. Introducción a la Inteligencia Artificial

2. Tipos de Inteligencia Artificial

3. Fundamentos de IA

4. Aplicaciones de la IA

5. Limitaciones y Desafíos (Funcionales)

6. El Futuro de la IA

TENDENCIAS Y DIRECCIONES FUTURAS

La IA es un campo en rápida evolución. Algunas áreas de desarrollo activo y tendencias:

- IA más Eficiente en Datos (Data-Efficient AI)
  - Técnicas que requieren menos datos para aprender (aprendizaje por transferencia, few-shot/zero-shot learning, aprendizaje autosupervisado donde se generan etiquetas a partir de los propios datos).
- IA Robusta y Fiable:
  - Modelos menos susceptibles a ataques adversariales y que puedan indicar cuándo no están seguros de una predicción (cuantificación de la incertidumbre).
- IA Híbrida:
  - Combinación de enfoques basados en datos (Machine Learning) con enfoques basados en conocimiento y razonamiento simbólico (como los sistemas expertos) para obtener lo mejor de ambos mundos.
- Neuro-Simbólica IA: Un ejemplo de IA híbrida que busca combinar el aprendizaje de las redes neuronales con el razonamiento de la IA simbólica.
- Automatización del Machine Learning (AutoML):
  - Herramientas que automatizan partes del proceso de desarrollo de modelos de ML (selección de algoritmos, ajuste de hiperparámetros), haciéndolo más accesible.

TENDENCIAS Y DIRECCIONES FUTURAS

- IA en el Borde (Edge AI):
  - Ejecutar modelos de IA directamente en dispositivos locales (teléfonos, sensores) en lugar de en la nube, para mayor privacidad, menor latencia y menor dependencia de la conectividad.
- IA Federada (Federated Learning):
  - Entrenar modelos de IA de forma colaborativa en múltiples dispositivos descentralizados que contienen datos locales, sin necesidad de compartir los datos en sí, lo que mejora la privacidad.
- Avances continuos en IA Generativa: Modelos cada vez más capaces de generar contenido complejo y de alta calidad (texto, imágenes, video, música, código).
- Hacia una IA más Causal: Intentos de desarrollar IA que no solo encuentre correlaciones, sino que entienda relaciones de causa y efecto.

Consideración Final

El desarrollo de la IA continuará planteando tanto enormes oportunidades como importantes desafíos funcionales, técnicos y, eventualmente, sociales y éticos que requerirán una cuidadosa consideración.

PRÁCTICAS

Práctica 1: Asistencia Legal con IA

Cada grupo tendrá una de las siguientes tareas:

- Grupo A (Contratos): Pide a la IA que genere un borrador de una cláusula de confidencialidad para un acuerdo de colaboración entre dos empresas de tecnología. Luego, el grupo debe analizar críticamente el borrador, identificando posibles ambigüedades, riesgos legales o elementos faltantes desde una perspectiva jurídica.
- Grupo B (Litigios): Solicita a la IA que resuma un fallo judicial complejo (puedes proporcionar un texto breve o un enlace a uno). El grupo debe evaluar la precisión del resumen, destacar los puntos clave que la IA capturó y aquellos que podría haber omitido o malinterpretado.
- Grupo C (Asesoría Legal): Pide a la IA que genere una respuesta a una pregunta legal común (e.g., requisitos para constituir una sociedad anónima en España). El grupo debe cuestionar la fiabilidad de la respuesta, identificar si hay información errónea o incompleta, y pensar qué advertencias darían a un cliente al usar esta información.
- Grupo D (Investigación Jurídica): Pide a la IA que genere un listado de normativa relevante para un tema específico (e.g., regulación de protección de datos en la UE para PYMES). El grupo debe verificar la actualidad y exhaustividad de la información y discutir las limitaciones de la IA para una investigación profunda.

Instrucciones clave: Los grupos deben probar distintos prompts (instrucciones) para obtener mejores resultados. Recuérdales que deben actuar como abogados críticos, no como meros usuarios.

PRÁCTICAS

Práctica 2

- Generar un resumen de un caso judicial ficticio.
- Crear una primera versión de una cláusula contractual simple.
- Explicar un concepto legal complejo en términos sencillos.
- La IA como una herramienta de apoyo, no un sustituto del criterio jurídico.

Preguntas y Discusión

¡Gracias por su atención!

FUNDAMENTOS Y APLICACIONES DE INTELIGENCIA ARTIFICIAL

Derecho de la era digital:
Del ciudadano (y sus derechos fundamentales) a la Inteligencia Artificial

Summer School Derecho Digital

Fecha de actualización: 29th May 2025

Prof. Dr. José Luis Salmerón

BY NC SA

Escuela Politécnica Superior
CUNEF Universidad