

GLOSARIO ALFABÉTICO DE INTELIGENCIA ARTIFICIAL

AGI (INTELIGENCIA ARTIFICIAL GENERAL): Es un concepto teórico que se refiere a una IA capaz de igualar o superar el rendimiento cognitivo humano en prácticamente cualquier tarea, poseyendo la capacidad de razonar y aprender a través de múltiples dominios de forma autónoma.

AJUSTE FINO (FINE-TUNING): Es el proceso de tomar un modelo de IA ya entrenado en un conjunto de datos general y realizar un entrenamiento adicional con datos específicos para optimizar su rendimiento en una tarea o dominio particular, como la medicina, el derecho, atención al cliente, educación. Suele ser más eficiente que crear un modelo nuevo

ALGORITMO (ALGORITHM): Un conjunto de reglas, fórmulas o instrucciones empleadas para procesar información y obtener un resultado. En el ámbito de la IA, el algoritmo es una indicación para el sistema que permite aprender de ejemplos, detectar patrones o clasificar información. Es un procedimiento ordenado para resolver un problema.

ALUCINACIÓN (HALLUCINATION): Ocurre cuando un modelo generativo produce información que suena muy convincente, pero es fácticamente falsa, inventada o confusa, debido a que el modelo predice la palabra más probable sin consultar una base de datos fidedigna. Las alucinaciones también pueden deberse a sesgos en los conjuntos de datos de entrenamiento o a la falta de acceso del modelo a información actualizada.

ANI (IA LIMITADA O DÉBIL): Se refiere a modelos de IA diseñados y entrenados para realizar una tarea específica y muy concreta, que no pueden adaptarse a otros dominios fuera de su programación. Son asistentes virtuales, traductores automáticos, reconocimiento de imágenes etc.

ANOTACIÓN DE DATOS: Proceso de añadir etiquetas, marcas o descripciones específicas a los datos sin procesar, como imágenes o textos, para que los algoritmos de IA los entiendan y puedan reconocer objetos, acciones o conceptos durante su entrenamiento.

APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING - ML): Es una rama de la IA que permite a los sistemas aprender y adaptarse sin necesidad de ser programados con instrucciones explícitas para cada tarea. Utiliza algoritmos para aprender de los datos, mejorando su precisión de forma gradual.

APRENDIZAJE NO SUPERVISADO: Es un enfoque de entrenamiento donde la máquina busca patrones, estructuras o grupos ocultos por sí sola dentro de conjuntos de datos que no tienen etiquetas ni instrucciones previas sobre lo que debe encontrar. Es útil cuando hay muchos datos disponibles, pero no se dispone de etiquetas ya preparadas.

APRENDIZAJE POR REFUERZO (REINFORCEMENT LEARNING): Es una tipología específica de aprendizaje automático en la que la máquina aprende a partir de los errores y aciertos cometidos en una tarea concreta mediante un sistema de recompensas y castigos que guían su comportamiento. La idea es parecida a aprender por experiencia: repetir lo que funciona y evitar lo que da mal resultado

APRENDIZAJE PROFUNDO (DEEP LEARNING): Es un subcampo del aprendizaje automático que utiliza redes neuronales artificiales con muchas capas para modelar y entender patrones complejos en grandes conjuntos de datos, imitando la estructura del cerebro humano. Es muy útil cuando la información es difícil de describir.

APRENDIZAJE SEMISUPERVISADO: Es una rama del machine learning que combina el aprendizaje supervisado y no supervisado mediante el uso de datos etiquetados y no etiquetados para entrenar modelos, siendo útil cuando hay grandes volúmenes de datos sin etiquetar

APRENDIZAJE SUPERVISADO: Es un modelo de entrenamiento donde la máquina aprende utilizando datos que ya han sido etiquetados con la respuesta correcta, permitiendo al sistema comparar su resultado con la solución real para aprender de forma guiada. Durante el entrenamiento, compara su predicción con la respuesta y va corrigiéndose de forma progresiva

CAJA NEGRA (BLACK BOX): Sistemas, incluidos varios algoritmos de aprendizaje automático, cuyo funcionamiento es inaccesible o demasiado complejo para que el usuario lo comprenda fácilmente, lo que implica falta de transparencia en sistemas de IA complejos donde las operaciones internas son tan intrincadas que resulta difícil para los humanos, incluidos sus creadores, explicar exactamente por qué el algoritmo ha tomado una decisión específica.

CÓDIGO ABIERTO: Modelo de desarrollo en el que el código de un sistema está disponible para ser revisado, reutilizado o modificado por otras personas. Favorece el escrutinio, la colaboración y en muchos casos, una mejora rápida de éste. En IA hablar de código abierto no siempre significa que todo este abierto. El código puede ser público mientras que los datos, los pesos del modelo o ciertas condiciones de uso sean restringidas

DATOS DE ENTRENAMIENTO (TRAINING DATASETS): Conjunto de datos que se utiliza para enseñar a los modelos de aprendizaje automático a identificar patrones y generar resultados. La calidad, diversidad, actualidad y equilibrio de estos datos influye directamente en el comportamiento del sistema. Si los datos son pobres o sesgados, el modelo tenderá a reflejar estas limitaciones.

DEEPPAKES: Imágenes y vídeos alterados deliberadamente para que parezcan realistas. Puede imitar la voz, apariencia o los gestos de una persona. Plantea riesgos importantes de fraude, desinformación y suplantación.

DESINFORMACIÓN: Creación y difusión deliberada de información falsa y/o manipulada que tiene como objetivo influir, confundir o engañar a las personas, ya sea con el fin de causar daño o con fines políticos, personales o económicos. No es un error sino una acción orientada a producir una percepción de la realidad equivocada

IA EXPLICABLE (XAI): Describe la capacidad de presentar o explicar el proceso de toma de decisiones de un sistema de aprendizaje automático en términos que puedan ser entendidos por humanos para garantizar la transparencia y la confianza.

IA DE FRONTERA (FRONTIER IA): Según la Junta Asesora Científica de la ONU se refiere a modelos de IA de propósito general altamente capaces que pueden realizar muchas tareas igualando o superando las capacidades de los modelos más avanzados de hoy en día. Son los modelos más avanzados

IA GENERATIVA (GENERATIVE AI): Es un tipo de IA capaz de crear contenido original y nuevo, incluyendo texto, imágenes, audio, vídeo u otros medios, basándose en patrones aprendidos a partir de grandes volúmenes de datos existentes en respuesta a las instrucciones del usuario.

IA MULTIMODAL: La IA multimodal combina y analiza diferentes entradas de información a la vez como texto, imagen, audio o video, combina varias fuentes para producir la respuesta más completa y generar resultados más sólidos. Permite a la IA, por ejemplo, analizar una fotografía, leer una pregunta sobre esta y responder con texto.

IA DE PROPÓSITO GENERAL: Se refiere a un sistema avanzado de IA capaz de realizar eficazmente una serie de tareas distintas y adaptarse a una amplia gama de aplicaciones y no una sola aplicación. Puede resumir, redactar, traducir, clasificar extraer información. No hay que confundirla con la I.A. General ya que la IA de propósito general es más limitada

IA RESPONSABLE: es un conjunto de principios que ayudan a guiar el diseño, desarrollo, despliegue y uso de la IA, generando confianza en las soluciones de IA que tienen el potencial de empoderar a las organizaciones y a sus partes interesadas. No se centra sólo en que la tecnología funcione sino en sus efectos en la sociedad

INGENIERÍA DE PROMPTS (PROMPT ENGINEERING): Es la práctica de diseñar, refinar y optimizar las instrucciones de entrada para lograr que el modelo de IA devuelva el resultado más preciso, útil y alineado con los objetivos del usuario.

INTERPRETABILIDAD: Se suele describir como la capacidad de presentar o explicar el proceso de toma de decisiones de un sistema de aprendizaje automático en

términos que puedan ser entendidos por humanos. También se denomina transparencia o explicabilidad.

MECANISMO DE ATENCIÓN (ATTENTION MECHANISM): Una técnica que permite al modelo enfocarse en las partes más relevantes de un texto o dato para entender su significado profundo, asignando importancia a las palabras clave según el contexto de la frase.

MODELOS DE LENGUAJE GRANDE (LLM): Son un tipo de modelo fundacional entrenado con cantidades masivas de texto para comprender y generar lenguaje humano natural, siendo capaces de realizar tareas complejas como traducción, resumen y redacción creativa. Durante las fases de entrenamiento, los grandes modelos de lenguaje aprenden parámetros a partir de factores como el tamaño del modelo y los conjuntos de datos de entrenamiento. Los grandes modelos de lenguaje utilizan entonces los parámetros para inferir nuevo contenido.

MODELOS FUNDACIONALES (FOUNDATION MODELS): Modelos de aprendizaje automático masivos entrenados con conjuntos de datos gigantescos para adquirir conocimientos generales que luego pueden adaptarse a una gran variedad de aplicaciones y tareas específicas posteriores. Sirven como base o bloques de construcción para crear aplicaciones más especializadas.

NLP (PROCESAMIENTO DE LENGUAJE NATURAL): Es el campo de la IA que se ocupa de la interacción entre las computadoras y el lenguaje humano, permitiendo a las máquinas entender, interpretar y generar el lenguaje humano de forma natural y coherente.

PARÁMETROS: Son las conexiones internas, los ajustes numéricos, de un modelo que cambian durante el entrenamiento y determinan su capacidad para captar matices complejos y minimizar el error, su número suele ser un indicador de la potencia y sofisticación del sistema de IA.

PRIVACIDAD DIFERENCIAL: Es un conjunto de técnicas matemáticas utilizadas para entrenar modelos asegurando que no se puedan extraer ni identificar datos

personales específicos de los individuos incluidos en el conjunto de entrenamiento.

PROMPT: Es la instrucción, pregunta o texto de entrada que el usuario proporciona a una IA generativa para iniciar una tarea u obtener una respuesta específica del modelo. El Prompt puede ser muy breve o muy detallada, cuanto mejor exprese el objetivo, el contexto y el formato esperado más probable es que la respuesta sea útil

RED GENERATIVA ADVERSARIAL (GAN): Estas redes son un desarrollo reciente en IA. Están formadas por dos subredes neuronales artificiales: una red generadora y una red discriminadora. La red generadora recibe datos de entrenamiento y produce datos artificiales basados en patrones. La red discriminadora compara los datos generados artificialmente con los datos de entrenamiento 'reales' y transmite a la red generadora donde ha detectado diferencias. El generador entonces altera sus parámetros. Con el tiempo, la red generadora aprende a dar datos más realistas, hasta que la red discriminadora no puede distinguir qué es artificial y qué es 'real'. El objetivo es mejorar y que el sistema llegue a producir resultados cada vez más realistas.

RED NEURONAL ARTIFICIAL (ARTIFICIAL NEURAL NETWORK): Es una arquitectura informática inspirada en el cerebro biológico compuesta por unidades computacionales interconectadas llamadas neuronas organizadas por capas que trabajan juntas para procesar información.

RED TEAM (EQUIPO ROJO): Es la práctica de realizar pruebas de seguridad donde expertos intentan encontrar de manera deliberada vulnerabilidades, sesgos o comportamientos peligrosos en un sistema de IA para corregirlos antes de su despliegue público. Consiste en intentar que el sistema falle para entender mejor sus riesgos.

RLHF (APRENDIZAJE POR REFUERZO A PARTIR DE RETROALIMENTACIÓN HUMANA): Es una técnica de machine learning en la que un "modelo de recompensa" se entrena con comentarios humanos directos y luego se utiliza para optimizar el rendimiento de un agente de inteligencia artificial mediante el aprendizaje por refuerzo.

SESGO ALGORÍTMICO (ALGORITHMIC BIAS): Una anomalía o prejuicio en los resultados de la IA producido porque los datos de entrenamiento contenían prejuicios humanos, históricos o por suposiciones erróneas durante el diseño del algoritmo.

SOBERANÍA TECNOLÓGICA: Capacidad de un país, institución o sector para disponer de la infraestructura, conocimiento y control sobre tecnologías clave para desarrollar IA. Supone no depender por completo de proveedores externos para datos, computación, modelos o servicios críticos

.

SUPERINTELIGENCIA: es un hipotético sistema de inteligencia artificial (IA) basado en software con un alcance intelectual más allá de la inteligencia humana. En el nivel más fundamental, esta IA superinteligente dispone de funciones cognitivas de vanguardia y capacidades de pensamiento muy desarrolladas, más avanzadas que las de cualquier ser humano.

TECNOLOGÍA EDUCATIVA: Conjunto de herramientas digitales destinadas a apoyar la enseñanza y el aprendizaje. No toda tecnología educativa utiliza IA, pero muchas herramientas actuales la incorporan en las plataformas de formación, sistemas de tutoría, contenidos interactivos...

TEMPERATURA: Es un parámetro de configuración que controla la aleatoriedad y creatividad del modelo, donde una temperatura baja da respuestas más predecibles y un alta genera resultados más variados y originales.

TOKENIZACIÓN: Es el proceso de dividir el texto en unidades más pequeñas llamadas tokens, que pueden ser palabras, sílabas o caracteres, para que los modelos de IA puedan procesar el lenguaje de forma numérica y matemática. 1000 tokens equivalen a unas 750 palabras

TOMA DE DECISIONES AUTOMATIZADA: Se refiere al uso de la tecnología, especialmente algoritmos de inteligencia artificial y aprendizaje automático, para

tomar decisiones de manera autónoma. Implica procesar datos, analizarlos y evaluar opciones sin intervención humana constante.

TRANSFORMADORES (TRANSFORMERS): Es la arquitectura de red neuronal que permite procesar secuencias de datos de forma paralela y analizar frases completas a la vez, facilitando una comprensión muy superior del contexto en comparación con modelos anteriores.

VENTANA DE CONTEXTO (CONTEXT WINDOW): Es la cantidad máxima de información o tokens que el modelo puede tener en cuenta en una interacción determinada. Incluye lo que la persona escribe y, según el sistema, también parte de lo dicho anteriormente. Cuánto mayor es la ventana, más texto o datos, puede considerar el sistema antes de perder detalles relevantes

VISIÓN ARTIFICIAL (COMPUTER VISION): Subcampo de la inteligencia artificial que dota a las máquinas de la capacidad de procesar, analizar e interpretar entradas visuales como imágenes y vídeos. Utiliza machine learning para ayudar a los ordenadores y otros sistemas a obtener información significativa a partir de datos visuales.