***Modelos de lenguaje basados en Transformers***

***Nombres:*** Juan Camilo Gutierrez Hoyos, Juan Camilo Espinosa Dominguez, Daniel Fernando Correa Pabon, Jose David Santacruz Guerrero,Carlos Enrique Peñaranda

* ***Breve explicación de la tendencia***

Los transformers se describieron por primera vez en un documento de 2017. Estos transformers son una de las clases más nuevas y potentes de modelos inventados hasta la fecha.

Un modelo transformer es una red neuronal que entiende los parámetros de entrada y genera un contexto con respecto a estos, dando una variedad de posibilidades de conjunción de las entradas, generando una salida lógica acorde a las entradas dadas, la estructura de este tipo de red neuronal se basa en una de sus predecesoras las redes recurrentes y modelos de atención usados en las redes de procesamiento de lenguaje, algo que resaltar de estas redes llamadas transformers es la forma en la que mantienen la posición en la que ingresan los datos, debido a que en esta red se ingresan todos los datos sin restricción, pero se usa un modelo en el que la aparición de la entrada está dada por la función seno y coseno las cuales dependiendo de la posición cambia la frecuencia de la función, con esto la red se valida no haber olvidado parámetros y le da un orden a la entrada de los mismos, generando así una salida mucho más completa y lógica.

Los transformers están siendo mayormente implementados en traducción de texto y charlas casi en tiempo real, además están ayudando a los investigadores a comprender las cadenas de genes en el ADN y los aminoácidos en proteínas de formas que pueden acelerar el diseño de fármacos, también pueden detectar tendencias y anomalías para prevenir fraudes, optimizar la manufactura, hacer recomendaciones en línea o mejorar el área de la salud.

* ***Impactos positivos:***

*Reducción de tiempos de procesamiento gracias a que la totalidad de entradas es procesada en paralelo por la red, permitiendo un entrenamiento más rápido de grandes modelos.*

*Gracias a su arquitectura basada en auto-atención, contribuyen a la solución del problema de memoria a largo plazo que sufren las principales redes neuronales recurrentes al tratar de acceder a información antigua.*

* ***Impactos negativos:***

Difíciles de implementar en hardware debido a la computación intensiva que requieren para su entrenamiento y funcionamiento. Así mismo esto tiene impacto a nivel ambiental por el consumo computacional que genera un mayor consumo energético.

Los sesgos es un problema que se viene dando en la inteligencia artificial y más con aplicaciones de procesamiento de lenguaje natural, algunas cosas comunes son sesgos ideológicos, raciales, clasistas, etc. Todo esto a raíz de las misma personas y que también depende en gran medida del dataset utilizado, con esto generando patrones tóxicos.

* ***Detallar al menos dos aplicaciones con sus posibles impactos***

**GPT-3**

Desarrollado por Open AI GPT-3 o Transformer Generativo Pre Entrenado 3 por sus siglas en inglés, es un modelo de procesamiento de lenguaje natural basado en la arquitectura de transformers entrenado con 175 billones de parámetros y que representa uno de los mayores avances en el campo del NLP hasta la fecha, las pruebas realizadas a GPT-3 demuestran que esta red tiene la capacidad de elaborar múltiples tareas a través de la interpretación del lenguaje natural, tareas como traducción, generación de textos, responder a preguntas u otras que implican conocimientos en ciertos dominios para llevarlas a cabo como la decodificación de palabras, mención a novelas, entre otros (Language Models are Few-Shot Learners, 2020).

Algunos impactos positivos de GPT-3 que fueron identificados son:

- La capacidad para automatizar procesos repetitivos que involucren la interpretación de grandes cantidades de texto como lo pueden ser la generación de reportes.

- La extracción de información a partir de grandes cantidades de texto para la toma de decisiones.

- Identificación de patrones en datos a partir de grandes cantidades de textos.

- Generación de nuevas soluciones que hagan uso de NLP para facilitar el desarrollo de tareas.

De la misma manera algunos impactos negativos o potenciales peligros de GPT-3 son:

- La posibilidad de generar y difundir textos que contengan prejuicios hacia diferentes grupos sociales.

- Su uso para cometer fraudes en línea y la difusión de noticias falsas.

- Su uso para dirigir publicidad a las personas en función de un perfil psicológico identificado a partir del análisis de texto.

- Su uso para influir en las emociones u opiniones de las personas.

**Megatron-Turing**

Resultado de la colaboración entre NVIDIA y Microsoft en la investigación para el entrenamiento de grandes modelos de IA, se presenta el modelo de procesamiento de lenguaje natural basado en transformers Megatron-Turing, descrito como el más grande y poderoso monolítico de lenguaje basado en Transformers entrenado hasta la fecha, con un total de 530 billones de parámetros (Alvi & Kharya, 2021), 3 veces más parámetros que el modelo de este tipo más grande previo a este.

Los impactos positivos y negativos identificados para el modelo de NLP Megatron-Turing son muy similares a los identificados en el modelo GPT-3 debido a la naturaleza de estos y la forma en la que han sido entrenados.

***Conclusiones***

Los modelos del lenguaje natural han venido evolucionando desde la propuesta de los transformers a un ritmo abismal, dejando atrás los modelos basados en redes neuronales recurrentes y abriendo el camino a modelos altamente paralelizables, pasando de modelos como ElMo con cientos de millones de parámetros entrenables a modelos como Megatron-Turing con 10 mil veces más parámetros entrenables en menos de una década. Esto por si solo nos da una idea del crecimiento y avance de esta tecnología, sin duda alguna son noticias positivas para el campo de la IA pero el uso de estas herramientas para fines malicioso no puede ser descartado, ya no son solo los sesgos producto de los datos de partida sino las posibles aplicaciones que pueden resultar en por ejemplo estafas masivas, IA entrenadas para engañar. Al fin y al cabo, estos modelos no poseen conciencia ni moral, por lo que esta responsabilidad ética cae sobre nosotros, quienes las diseñamos y usamos ¿Qué mecanismos se pueden diseñar para prevenir estos usos maliciosos? Si es técnicamente inviable evitarlo ¿Debemos los ingenieros y encargados de IA exigir mayor reglamentación? Son preguntas que debemos hacernos y encargarnos de solucionar.

***Referencias***

*- Language Models are Few-Shot Learners. (2020, 28 mayo). Arvix. Recuperado 9 de agosto de 2022, de* [*https://arxiv.org/pdf/2005.14165.pdf*](https://arxiv.org/pdf/2005.14165.pdf)

*- Alvi, A., & Kharya, P. (2021, 11 octubre). Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World’s Largest and Most Powerful Generative Language Model. Microsoft. Recuperado 9 de agosto de 2022, de* [*https://www.google.com/url?q=https://www.microsoft.com/en-us/research/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/&sa=D&source=docs&ust=1660167484009417&usg=AOvVaw275h-9NCxPPOnpC9lPgiGl*](https://www.google.com/url?q=https://www.microsoft.com/en-us/research/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/&sa=D&source=docs&ust=1660167484009417&usg=AOvVaw275h-9NCxPPOnpC9lPgiGl)

*- Attention is all you need. A. Vaswani, N. Shazeer, et all. (2017) 31st Conference on Neural Information Processing Systems.*