



Sociedad de Inteligencia Artificial de la
Facultad de Ingeniería
Actividad del curso: Ética en la IA

Téllez González Jorge Luis

23 de marzo, 2022



1. Menciona un intento de escapar el negocio de la vigilancia y lo que propone para lograrlo.

Entre los navegadores web que se emplean actualmente, *Brave* es uno de los que más esfuerzos han puesto para evitar la recolección masiva de datos de los usuarios que emplean el navegador. Por ejemplo, Brave permite saltarse la tecnología de carga web AMP de Google que, si bien su intención es cargar páginas web en dispositivos móviles, por otra parte también es empleada para mantener un rastreo de las páginas web que un usuario visita, expandiendo la constante vigilancia pasiva que tienen sobre ellos.

Aunque también se ha hablado de *DuckDuckGo* como un buscador seguro y respetuoso con los datos de sus usuarios, recientemente han salido a la luz una polémica en la transparencia del mismo al bloquearse las búsquedas de páginas web *pirata* o con copyright en un intento de censura; lo cual es negado por sus desarrolladores tajantemente.

2. Explica como la Inteligencia Artificial podría convencer a algunos usuarios a ceder su privacidad.

Un ejemplo claro se puede observar en los servicios-producto que, bajo la capa de mejorar la experiencia de un usuario en su vida diaria (como el caso de los *Echo* de Amazon), en letras pequeñas sus usuarios estén cediendo el control de sus datos personales e, incluso, registros de sus actividades diarias en forma de grabaciones de audio sin autorización explícita del usuario que usa tales servicios; bajo la premisa de que tal violación de privacidad es para mejorar la inteligencia artificial de los servicios de Amazon.



3. ¿Cuál es la relación entre la opacidad de un sistema de Inteligencia Artificial y el sesgo de este?

Una opacidad en un sistema de IA se define como el desconocimiento de la serie de procedimientos que esta realiza para llegar a un resultado determinado; incluso al grado de aceptarlos por completo en una especie de dogma sin posibilidad de cuestionar. La opacidad contradice uno de los principales aspectos que una IA debe cumplir que es la *transparencia*.

Por otra parte, el sesgo de las mismas se define por el hecho que una IA que tenga una cierta inclinación a entregar determinados resultados o llegar a ciertas conclusiones es consecuencia de una inclinación marcada por parte de los desarrolladores de la misma. Esto presenta cierto problema junto con la opacidad ya que, si un sistema de IA no presenta transparencia en la forma que recopila datos, los analiza y finalmente los procesa, se corre el riesgo de que, en uno de estos procesos, la transparencia se pierda y comiencen a distorsionarse a favor de determinada tendencia (incluso discriminatoria) que, potencialmente, pueda afectar a terceros.

Como ejemplo de esto, se pueden observar los programas de autoselección de candidatos a empleo que, por medio de algoritmos de IA poco transparentes, pueden favorecer a determinados candidatos y desechar a otros debido a los prejuicios de lo que un candidato *ideal* debe tener de acuerdo a la visión de los desarrolladores del software; lo que provoca que potenciales candidatos a puestos de trabajo no puedan tener siquiera una oportunidad de avanzar a una entrevista por una IA que los rechazó sin saber a profundidad si realmente se trataba de una persona cualificada o no para el puesto.



4. ¿Por qué se dice que un data set siempre tendrá sesgo, excepto para la situación para la cual fue diseñado?

Debido a que un sistema de IA se inclinará a lo que aprendió y tratará de replicarlo pues es lo que considera *correcto* bajo la perspectiva de su entrenamiento. Por ejemplo, un sistema que fue entrenado con imágenes de perros y de selección de perros, evidentemente tendrá marcada tendencia por estos animales en lugar de, por ejemplo, gatos que estén en el dataset.

Incluso, un sistema como *Michigan* que, al trabajar con datasets de bosquejos de gatos y construirlos por completo como gatos presenta una situación de este estilo, ya que si se introduce un bosquejo que no es un gato en sí, con base en lo que el modelo aprendió este tratará de construir un gato de cualquier forma.

5. Menciona las consecuencias del uso de la Inteligencia Artificial y la robótica en el ámbito sexual por qué son preocupantes.

Aunque la IA actualmente no alcanza niveles de conciencia semejantes al de un ser humano, existe el dilema de que, en un futuro, la IA pueda tratar de atribuirse más importancia y eventualmente reemplazar a otros seres humanos completamente; poniendo en la mesa el dilema de si el amor o cariño que estas demuestran realmente es igual al de un ser humano o es únicamente una mera imitación carente de *esencia* humana, entendiendo la esencia como sentimientos genuinos y no creados artificialmente como respuestas genéricas a determinados impulsos.

El juego *Doki Doki Literature Club* trae este debate a consideración con el personaje de *Monika*, un NPC dentro del juego que toma conciencia de que se encuentra dentro de un videojuego y realiza esfuerzos tangibles para conectar directamente con el jugador por medio del sabotaje en el

juego y establecer una conexión sentimental con el mismo. Incluso, a raíz del enorme impacto que el juego tuvo, se han realizado *mods* del juego donde el jugador puede interactuar directamente con Monika conversando sobre una multitud de temas variados y realizar actividades con ella; mostrando sentimientos de felicidad o enojo dependiendo de cómo trate el jugador a Monika.

La mayor preocupación reside en el hecho de que si las IAs algún día pueden llegar a materializarse, ¿Cómo puede garantizarse de que piensan y expresan sentimientos de forma genuina? Y si lo hacen, ¿Qué tan seguro sería para la integridad de las personas involucrarse amorosamente y/o sexualmente con IAs? ¿Esto podría provocar el declive de las tasas de natalidad humanas al *deshumanizar* las relaciones sexuales? O, por otra parte, ¿Sería ético someter a una IA a los deseos sexuales de un ser humano?

6. Explica las implicaciones ambientales del uso normalizado de la Inteligencia Artificial

El software que emplea IA ha planteado preocupaciones de que, debido a los recursos que requieren para operar, puedan estar generando un impacto ambiental muy profundo en el ambiente por medio por la naturaleza de la energía que emplean (especialmente aquellas energías contaminantes como el carbón o el petróleo). Aunque *per se* no forman parte de las IAs, la tecnología emergente *Blockchain* ha sido frecuentemente señalada por contribuir negativamente a los esfuerzos de preservación del ambiente por el enorme gasto de energía que emplean (aunque, en la actualidad, redes como Ethereum están en proceso de transición a *Proof of Stake* para reducir el consumo total de la red de forma dramática).



7. ¿Por qué resulta complicada la implementación de vehículos automatizados con respecto a su sistema?

Debido a que al implementar su uso en situaciones cotidianas se encuentran expuestos a posibles incidentes de tráfico que, incluso, puedan tener consecuencias fatales para seres humanos que estén involucrados. Uno de los problemas más frecuentes incluye el hecho de cómo es que un carro autónomo debe actuar al encontrarse en una situación crítica que terminará con daño a uno u otro ser humano sin poder realizar una acción para evitar un daño a ambos.

8. Menciona un intento de establecer un estándar ético en la Inteligencia Artificial.

Aunque no existe un estándar como tal debido a las distintas visiones que se tienen de los alcances y usos de una IA, tras el incidente del *Proyecto Maven* por parte de Google, esta compañía estableció una serie de 7 principios básicos en el desarrollo de un sistema basado en IA para establecer un precedente ético en todos los proyectos que desarrollaría a partir de ese momento:

1. **Que sea socialmente benéfica.**
2. **Evitar crear o reforzar un sesgo injusto.**
3. **Estar construida y probada para ser segura.**
4. **Ser responsable con las personas.**
5. **Incorporar principios de diseño de privacidad.**
6. **Mantener altos estándares de excelencia científica.**
7. **Estar disponible para usos que vayan de acuerdo con estos principios.**



9. ¿A qué se refiere la “moratoria de fenomenología sintética”?

Este término proviene a raíz de las declaraciones del filósofo *Thomas Metzinger* que se refiere a detener el avance hacia IAs que puedan presentar cierto grado de conciencia de sí mismas, ya que actualmente no somos capaces como humanidad de poder comprender la experiencia subjetiva del ser, la *conciencia fenoménica*. Debido a que el conocimiento humano no puede explicar de forma certera la conciencia, desarrollar IAs encaminadas a ese desarrollo puede conducir a resultados inesperados y que traerían graves problemas éticos para los que la humanidad no estaría preparada.

10. Explica por que el xRisk genera tal terror en las personas.

Debido a que se plantea que, cuando la IA por fin llegue a alcanzar la singularidad, sea el punto de partida de la extinción de la humanidad como especie y las IAs tomen el lugar que la humanidad tuvo en alguna ocasión. Esto se plantea como el miedo latente de que la IA no evolucione como una herramienta o asistente de la humanidad, sino que avance como un competidor directo de los seres humanos y, eventualmente, termine sobrepasándolos en inteligencia y capacidades de supervivencia.

Los créditos de las fotografías pertenecen a sus respectivos autores. ©