

There is a blind spot in AI research. Fears about the future impacts of artificial intelligence are distracting researchers from the real risks of deployed systems

Enviado por cristobalrn en Dom, 06/23/2019 - 15:27

Cita:

Crawford, Kate and Ryan Calo [2016], "There is a blind spot in AI research. Fears about the future impacts of artificial intelligence are distracting researchers from the real risks of deployed systems", *Nature*, (538):311-313, 20 de octubre, https://www.nature.com/news/polopoly_fs/1.20805!/menu/main/topColumns/to... [1]

Fuente:

Otra

Fecha de publicación:

Jueves, Octubre 20, 2016

Revista descriptores:

[Empresas transnacionales y gobernanza mundial](#) [2]

[Estudios de caso: actividades - empresas](#) [3]

[Fronteras del capital](#) [4]

[Relaciones entre empresas estados y sociedad](#) [5]

[Tecnologías militares - tecnologías de uso dual](#) [6]

Tema:

Enfoques para evaluar los impactos sociales de la inteligencia artificial

Idea principal:

Kate Crawford es investigadora principal en Microsoft Research, profesora visitante en el Massachusetts Institute of Technology e investigadora en la Universidad de Nueva York.

Ryan Calo es profesor asistente de leyes y de ciencias de la información, así como co-director del *Tech Policy Lab* en la Universidad de Washington en Seattle.

El 12 de octubre de 2016 la Casa Blanca publicó un informe sobre el futuro de la inteligencia artificial; el informe fue el resultado de cuatro talleres que se dieron entre mayo y julio de 2016. En estos eventos participaron expertos de diversas áreas para discutir cómo la inteligencia artificial cambiará la manera en que vivimos. Además de señalar las potencialidades del uso de las tecnologías con inteligencia artificial en la vida cotidiana, los talleres también destacaron un "punto ciego" muy importante en el pensamiento sobre la inteligencia artificial: los sistemas autónomos ya se utilizan en tareas clave en muchas de las instituciones más importantes de nuestra vida (desde los hospitales hasta las cortes judiciales), pero aún no hay métodos aceptados para evaluar los efectos de esas tecnologías sobre los individuos y la sociedad.

Aunque ha habido un enorme avance en el dominio técnico sobre la inteligencia artificial, es necesario que esas mejoras se complementen con lo que los autores llaman “análisis de sistemas sociales” para evaluar el impacto social, cultural y político de esas tecnologías. Un análisis de este tipo serviría para investigar, entre otras, cómo las tecnologías con inteligencia artificial modifican la relación médico-paciente, cuáles son los efectos sociales de la automatización de los préstamos bancarios o si el uso de datos históricos para predecir dónde ocurrirán los crímenes provoca que haya una vigilancia excesiva en los barrios marginales.

Los sistemas con inteligencia artificial ya son utilizados en numerosas áreas, desde los análisis químicos hasta la asignación de publicidad personalizada, pasando por las empresas legales y las instituciones financieras. Un problema es que el entusiasmo que provocan estas tecnologías ha llevado a que sean utilizadas sin que hayan sido suficientemente probadas y sin que sus efectos sociales hayan sido evaluados.

Los tecnólogos confían que la inteligencia artificial ayudará a reducir los errores, prejuicios y sesgos de los humanos. No obstante, estudios recientes muestran que los sistemas con inteligencia artificial tienden a reproducir esos sesgos y a afectar a grupos que ya son objeto de marginación o discriminación por motivos de raza, género, edad o condición socioeconómica.

¿Cómo evitar estos efectos perniciosos de las tecnologías? Hasta ahora ha habido tres maneras principales de responder a las preocupaciones sobre los impactos éticos y sociales de la inteligencia artificial: implementar y ajustar; introducir valores en el diseño; y los experimentos mentales [thought experiments]. Los autores consideran que aunque las tres aproximaciones son valiosas, son insuficientes.

Implementar y ajustar [deploy and comply]. Es una aproximación que consiste en poner a funcionar las tecnologías e introducir ajustes de manera reactiva cuando hay alguna dificultad. Un ejemplo: Google ajustó su algoritmo de reconocimiento de imágenes después de que el sistema confundió a una pareja de afroamericanos con gorilas. El problema es que aunque las corporaciones puedan introducir ajustes a sus sistemas, sigue sin haber investigación sobre sus efectos en el mediano plazo.

Valores en el diseño . A partir de este enfoque, los sistemas y productos deben incorporar en su diseño los valores de los usuarios. Las empresas de esta industria han hecho esfuerzos por crear marcos y estándares compartidos que conduzcan a la “innovación responsable”. El problema con este enfoque es que parte del supuesto de que los sistemas serán diseñados y utilizados, de tal manera que no permite a los hacedores de política ni a la sociedad decidir si un sistema debería construirse o si los prototipos aún no son confiables para su implementación.

Experimentos mentales. En los últimos años, el debate sobre las implicaciones éticas y sociales de la inteligencia artificial ha estado dominada por situaciones hipotéticas o lo que los autores llaman “experimentos mentales”. Un ejemplo es la extensión del “dilema del tranvía” [1] hacia los vehículos autónomos. La formulación de esta situación hipotética busca poner a debate las decisiones éticas, así como la responsabilidad y la culpabilidad: si un vehículo autónomo atropella a una persona, ¿quién es el responsable? ¿La empresa que lo fabricó, el dueño de ese vehículo, la persona que va a bordo de él? El problema que los autores ven en este enfoque es que distrae la atención de otros problemas sociales más amplios: ¿debe priorizarse la inversión en

vehículos autónomos o en el transporte público? ¿Cuál será el impacto de los vehículos autónomos sobre el ambiente y el empleo?

Los autores consideran que es necesaria una cuarta aproximación: el análisis de sistemas sociales. Este análisis se cuestiona por “todos los posibles efectos de los sistemas con inteligencia artificial sobre todas las partes” involucradas. Asimismo, indaga el impacto social de todas las etapas de la inteligencia artificial: concepción, diseño, implementación y regulación.

Para los autores, lo primero que se tendría que hacer sería investigar cómo las diferencias de ingreso, información y servicios básicos al interior de una comunidad y entre distintas comunidades inciden y dan forma a los datos con los que se entrenan los sistemas con inteligencia artificial. Esto implica considerar la historia social y política de los datos que sirven de alimento para los algoritmos. Asimismo, el análisis de sistemas sociales exige considerar las características de los grupos donde se están implementando esas tecnologías para que funcionen adecuadamente.

El análisis de sistemas sociales debe ser interdisciplinario y recurrir a la filosofía, el derecho, la sociología, la antropología y los estudios sobre ciencia y tecnología, entre otras disciplinas. Debe también preguntarse cómo el cambio tecnológico afecta los valores sociales, políticos y culturales de las personas. Eso permitiría tener una comprensión holística e integrada de la inteligencia artificial.

“La inteligencia artificial representa un cambio cultural tanto como un cambio tecnológico”. Los sistemas con esta tecnología están cambiando nuestra vida cotidiana, del trabajo al entretenimiento. “Debemos asegurarnos que esos cambios sean benéficos”.

[1] El dilema del tranvía es un experimento mental en ética que afirma lo siguiente: “Un tranvía corre fuera de control por una vía. En su camino se hallan cinco personas atadas a la vía por un filósofo malvado. Afortunadamente, es posible accionar un botón que encaminará al tranvía por una vía diferente, por desgracia, hay otra persona atada a ésta. ¿Debería pulsarse el botón?”. Al respecto véase: https://es.wikipedia.org/wiki/Dilema_del_tranv%C3%ADa [7]

Datos cruciales:

En su libro *The Master Algorithm* el científico de la computación Pedro Domingos afirma que “a la gente le preocupa que las computadoras se puedan volver demasiado inteligentes y que tomen el control del mundo; el problema real es que son demasiado estúpidas y ya tomaron el control del mundo”.

Los análisis de mercado asignan a las aplicaciones de inteligencia artificial un valor de aproximadamente mil millones de dólares. La directora ejecutiva de IBM, Ginni Rometty, considera que en la próxima década los sistemas con inteligencia artificial tendrán un valor de 2 billones [trillion] de dólares.

Nexo con el tema que estudiamos:

Aunque la investigación técnica sobre inteligencia artificial ha avanzado a pasos agigantados, los

efectos de su aplicación sobre la sociedad, la economía y la cultura no han sido suficientemente explorados. En la mayoría de los estudios que se proponen investigar las implicaciones sociales de estas tecnologías, el abordaje suele centrarse en los aspectos éticos de su implementación (los sesgos y prejuicios que estas tecnologías reproducen, entre otras) y en el impacto de la automatización sobre el empleo. Desde nuestra perspectiva, es necesario profundizar en la comprensión sobre cómo esta tecnología que forma parte del núcleo tecnológico de vanguardia del capitalismo contemporáneo contribuye a reproducir el poder y las ganancias de las grandes corporaciones que las utilizan.

Source URL (modified on 24 Junio 2019 - 6:41 pm): <http://let.iiec.unam.mx/node/2282>

Links

- [1] https://www.nature.com/news/polopoly_fs/1.20805!/menu/main/topColumns/topLeftColumn/pdf/538311a.pdf
- [2] <http://let.iiec.unam.mx/taxonomy/term/14>
- [3] <http://let.iiec.unam.mx/taxonomy/term/16>
- [4] <http://let.iiec.unam.mx/taxonomy/term/18>
- [5] <http://let.iiec.unam.mx/taxonomy/term/20>
- [6] <http://let.iiec.unam.mx/descriptores-let/tecnolog%C3%ADas-militares-tecnolog%C3%ADas-de-uso-dual>
- [7] https://es.wikipedia.org/wiki/Dilema_del_tranv%C3%ADa