

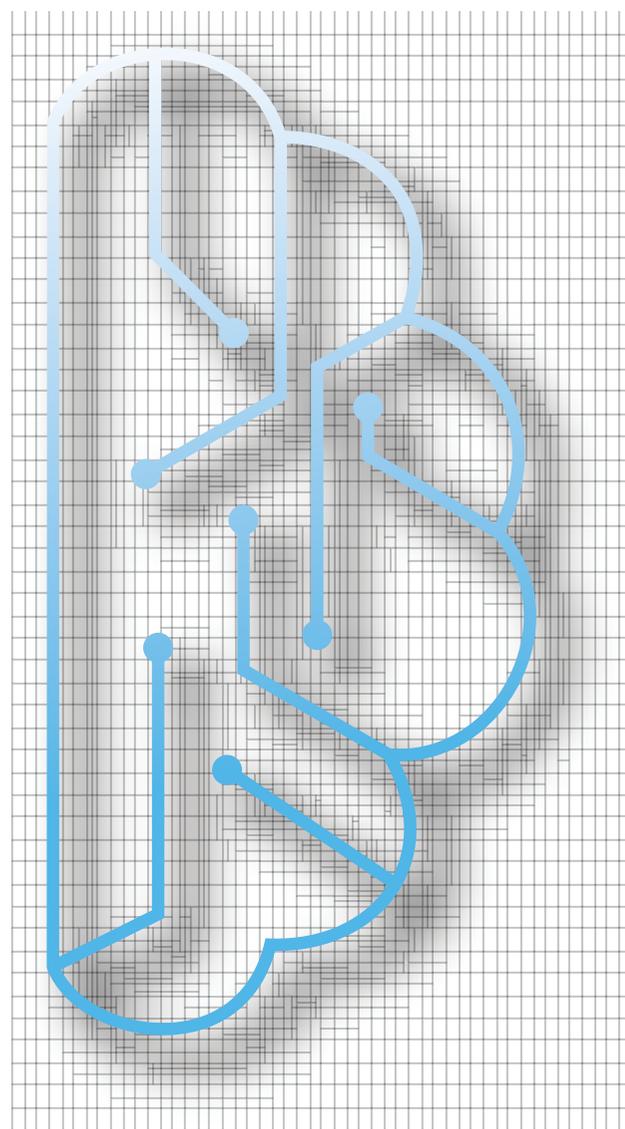
Cinco pasos

que la **IA** debería dar en 2021 para volverse «buena»

En los últimos años, la inteligencia artificial ha demostrado algunos avances para hacerse más justa, ética y capaz de resolver problemas reales, pero todavía tiene mucho camino por delante. Este año, se debería aumentar la inversión para enfoques minoritarios, la regulación y la inclusión de usuarios y colectivos afectados.

**MIT
Technology
Review**

Publicado por Opinno



KAREN HAO

TRADUCIDO POR ANA MILUTINOVIC
18 ENERO, 2021

Hace un año, antes de saber lo que traería 2020, reflexioné sobre el momento crucial en el que se encontraba la comunidad de la inteligencia artificial (IA). El año anterior, 2018, se habían detectado una serie de errores de alto nivel en la tecnología, como choques de coches sin conductor y herramientas de contratación discriminatorias. En 2019, el campo respondió con más charlas que nunca sobre la ética de la IA. Pero hablar, dije, no era suficiente. Necesitábamos tomar acciones reales. Dos meses después, el coronavirus paralizó el mundo.

En nuestra nueva realidad del distanciamiento social y vida en remoto, las conversaciones sobre los daños algorítmicos llegaron a un punto crítico. Los sistemas que habían estado al margen del debate, como los algoritmos de escaneo facial de HireVue y las herramientas de vigilancia en el trabajo, se volvieron dominantes. Otros, como las herramientas para monitorizar y evaluar a los estudiantes, se desarrollaban en tiempo real.

En agosto, tras el espectacular fracaso del Gobierno de Reino Unido para sustituir los exámenes presenciales con un algoritmo para las admisiones universitarias, cientos de estudiantes se reunieron en la capital del país, Londres, para manifestarse bajo el lema: «A la mierda el algoritmo». Y cuando un manifestante de la Universidad de Stanford (EE. UU.) volvió a gritarlo como respuesta a una debacle diferente unos meses más tarde, la investigadora de las responsabilidades de IA Deborah Rajin tuiteó: «Esto se está convirtiendo en el grito de guerra de 2020».

Al mismo tiempo, efectivamente hubo más acción. Se produjo una gran victoria cuando Amazon, Microsoft e IBM prohibieron o suspendieron la venta de sus productos de reconocimiento

facial a las fuerzas del orden, después de que el asesinato de George Floyd provocara protestas mundiales contra la brutalidad policial.

Fue la culminación de dos años de lucha de investigadores y activistas de derechos civiles para demostrar los efectos ineficaces y discriminatorios de las tecnologías de las empresas. Y hubo otro cambio pequeño, pero notable: por primera vez, NeurIPS, una de las conferencias de investigación de IA más destacadas, requirió que los investigadores presentaran una declaración de ética con sus artículos.

Ahora estamos a principios de 2021, y la atención pública y regulatoria sobre la influencia de la IA es mayor que nunca. Mi deseo de año nuevo: aprovechémoslo. Presento las cinco esperanzas que tengo para la IA en el próximo año.

REDUCIR LA INFLUENCIA EMPRESARIAL EN LA INVESTIGACIÓN

Los gigantes tecnológicos tienen un control desproporcionado sobre el rumbo de la investigación de la IA. Esto ha cambiado la dirección del campo en su conjunto hacia enormes cantidades de datos y modelos cada vez más grandes, con varias consecuencias. Esta tendencia aumenta el impacto climático de los avances de la IA, impide que los laboratorios con recursos limitados participen en el campo y conduce a una investigación científica más perezosa al ignorar los demás enfoques posibles. Como mostró la destitución de Timnit Gebu por parte de Google, los gigantes tecnológicos también limitan fácilmente la capacidad de investigar otras consecuencias.

Pero gran parte de la influencia empresarial se reduce al dinero y a la falta de financiación alternativa. Como escribí el año pasado en mi reportaje sobre OpenAI, inicialmente el laboratorio quería depender solo de donaciones independientes y patronos. Esa apuesta resultó insostenible y, cuatro años después, OpenAI firmó un acuerdo de inversión con Microsoft. Espero que los gobiernos empiecen a cubrir este vacío para ofrecer a los investigadores opciones de financiación ajenas al mundo militar. No será una solución perfecta, pero sería el inicio. Los gobiernos están sujetos a la sociedad, no al resultado final.

VOLVER A CENTRARSE EN LA COMPRENSIÓN DEL SENTIDO COMÚN

La inmensa atención puesta en los modelos más grandes y mejores ha eclipsado uno de los objetivos centrales de la investigación de la IA: crear máquinas inteligentes que no solo busquen los patronos, sino que realmente entiendan el mundo. Si bien la influencia corporativa es el principal responsable de esta tendencia, también hay otros culpables. Las conferencias de investigación y las publicaciones revisadas por pares ponen un gran énfasis en conseguir resultados de «la última tecnología», pero eso a menudo se mide de manera deficiente mediante pruebas que pueden superarse con más datos y con modelos más grandes.

la atención puesta en modelos más grandes ha eclipsado uno de los objetivos centrales de la IA: crear máquinas inteligentes que no solo busquen los patronos, sino que entiendan el mundo.

No es que los modelos a gran escala puedan adquirir sentido común. Esa sigue siendo una pregunta abierta. Pero hay otras vías de investigación que merecen una mayor inversión. Algunos expertos han apostado por la IA neurosimbólica, que combina aprendizaje profundo con sistemas de conocimiento simbólico. Otros experimentan con técnicas probabilísticas, que utilizan muchos menos datos y se inspiran en la capacidad de un niño de aprender con muy pocos ejemplos.

En 2021, espero que el campo reoriente sus incentivos para priorizar la comprensión sobre la predicción. Esto no solo podría conducir a sistemas técnicamente más sólidos, sino que las mejoras también tendrían importantes consecuencias sociales. La susceptibilidad de los actuales sistemas de aprendizaje profundo a ser engañados, por ejemplo, socava la seguridad de los vehículos autónomos y plantea peligrosas posibilidades para las armas autónomas. La incapacidad de los sistemas para distinguir entre la correlación y la causalidad también es la raíz de la discriminación algorítmica.

EMPODERAR A LOS INVESTIGADORES MARGINADOS

Si los algoritmos codifican los valores y los puntos de vista de sus diseñadores, su desarrollo debería involucrar a una muestra amplia y representativa de la humanidad. La mejor prueba de esto la vi en diciembre de 2019, cuando asistí a NeurIPS. Ese año, con un número récord de mujeres ponentes y representantes de minorías y asistentes, noté que el tono de los actos cambiaba de verdad. Hubo más debate que nunca sobre cómo lidiar con la influencia de la IA en la sociedad.

En ese momento alabé a la comunidad por su progreso. Pero el trato de Google a Gebru, una de las pocas mujeres negras destacadas en este sector, mostró lo mucho que queda todavía por hacer. La diversidad numérica no tiene sentido si esas personas no están empoderadas para llevar su experiencia vivida a su trabajo. Sin embargo, soy optimista de que el ambiente está cambiando. El foco marcado por el despido de Gebru se convirtió en un momento crítico de reflexión para la industria. Espero que este impulso lleve a un cambio sistémico duradero.

algunos expertos han apostado por la IA neurosimbólica, que combina aprendizaje profundo con sistemas de conocimiento simbólico.

CENTRAR LAS OPINIONES DE LAS COMUNIDADES AFECTADAS

También hay otro grupo del que hay que hablar. Una de las tendencias más emocionantes del año pasado fue la aparición del aprendizaje automático participativo. Es un intento de reinventar el proceso de desarrollo de la IA para incluir a aquellos que, en última instancia, acaban sujetos a los algoritmos.

En julio, el primer taller-conferencia dedicado a este enfoque recogió una amplia variedad de ideas sobre cómo podría llevarse a cabo. Las sugerencias incluyeron nuevos procedimientos directivos para solicitar comentarios de la comunidad; nuevos métodos de auditoría de modelos para informar e involucrar al público; y propuso rediseños de los sistemas de inteligencia artificial para ofrecer a los usuarios un mayor control de su configuración.

Mi esperanza para 2021 es ver más de estas ideas exploradas y adoptadas formalmente. Facebook ya ha empezado a hacerlo: si sigue adelante y permite que su junta de supervisión externa realice cambios vinculantes en las políticas de moderación de contenido de la plataforma, la estructura directiva podría convertirse en un mecanismo de retroalimentación digno de seguir como ejemplo.

CODIFICAR LAS BARRERAS DE PROTECCIÓN EN LA REGULACIÓN

Hasta ahora, los esfuerzos más básicos han logrado que los movimientos mitiguen algunos daños algorítmicos y responsabilizar a los gigantes tecnológicos. Pero establecer barreras de protección más permanentes dependerá de los reguladores nacionales e internacionales. La buena noticia es que los legisladores de todo el mundo han estado observando y están redactando la legislación. En EE. UU., los miembros del Congreso ya han presentado proyectos de ley para abordar el reconocimiento facial, el sesgo de la inteligencia artificial y los *deepfakes*. Varios de ellos también enviaron una carta a Google en diciembre expresando su intención de continuar con esta regulación.

Por eso, mi último deseo para 2021 es que veamos la aprobación de algunos de estos proyectos de ley. Es hora de que codifiquemos lo que hemos aprendido en los últimos años y nos alejemos de la ficción de la autorregulación. </>



La autora es reportera senior de IA en MIT Technology Review.