Can we stop AI outsmarting humanity?

Enviado por NardaCr en Lun, 04/08/2019 - 12:59
Cita:

Hvistendahl, Mara [2019], "Can we stop AI outsmarting humanity?", *The Guardian*, London, 28 de marzo, https://www.theguardian.com/technology/2019/mar/28/can-we-stop-robots-ou... [1]

Fuente:
Otra
Fecha de publicación:
Jueves, Marzo 28, 2019
Revista descriptores:
Estudios de caso: actividades - empresas [2]
Fronteras del capital [3]
Tecnologías militares - tecnologías de uso dual [4]
Tema:
Riesgos de la Superinteligencia en la Inteligencia Artificial.
Idea principal:

Jaan Tallin es un programador de informática nacido en Estonia y con una formación de físico, en 2007 encontró el texto de Eliezer Yudkowsky titulado “Staring into the Singularity” el cual habla sobre el avance tecnológico y como la civilización humana se vería rebasada en muchos aspectos hasta el grado de ser dominados por inteligencia artificial.

Tallin leyó más textos de Yudkowsky, de los cuales muchos habían sido dedicados a la inteligencia artificial. Al leer los artículos de Yudkowsky, Tallin se convenció de que la superinteligencia en la inteligencia artificial (IA) podría amenazar la existencia humana, ocupando nuestros espacios y dominándonos al grado de exterminarnos.

Por medio de un correo electrónico Tallin contactó a Yudkowsky, con el cual se reunió en California una semana después, en donde hablaron sobre diversos conceptos y los retos que la Inteligencia Artificial trae consigo. Tallin dio un cheque para el Singularity Institute for Artificial Intelligence, en donde Yudkowsky era investigador.

Después de encontrarse, Tallin dio charlas por el mundo sobre la amenaza planteada por la superinteligencia, a la vez que comenzó con el financiamiento de la investigación de métodos que la humanidad podría tener para evitar el colapso: la Inteligencia Artificial amigable[1].

Tallin da donaciones a 11 organizaciones en las que trabaja diferentes enfoques para la seguridad de IA. A su vez, otras personas de su círculo también han dado donaciones a instituciones de seguridad como Peter Thiel cofundador de PayPal al Machine Intelligence Research Institute y Elon Musk cofundador de Tesla al Future of Life Institute.

Dentro del Centre for the Study of Existential Risk (CSER) cofundado por Tallin, se estudian los riesgos existenciales, los cuales son amenazas para la supervivencia de la humanidad como accidentes, mal uso de la inteligencia artificial y carrera de armas. Otros temas que se abordan son el cambio climático, la guerra nuclear y las armas biológicas. Aunque estos temas no son el interés principal de Tallin, él espera que sean los que atraigan a los investigadores para después pasar al plano de la Inteligencia Artificial. Los investigadores de los fondos de Tallin están conscientes que la IA no tiene como fin el dominar el mundo, pero que como en todo, pueden existir errores que puedan llegar a eso.

A su vez, también existen detractores de las ideas de Tallin, incluso de la misma comunidad preocupada por la seguridad de la IA, algunos diciendo que aún es muy pronto para preocuparse por la restricción de la inteligencia artificial superinteligente debido a que todavía no la entendemos y otros que están desviando la atención de los verdaderos problemas como que la mayoría de los algoritmos están diseñados por hombres blancos.

Stuart Armstrong es un investigador del Future of Humanity Institute, instituto que ha recibido donaciones de Tallin. Armstrong es de los pocos investigadores que se centra de tiempo completo en la seguridad de la IA. Este investigador ha propuesto límites y medidas para poder frenar en su momento a la IA, como un gran botón de apagado, aunque evidentemente no será sencillo de construir. Otro enfoque, el que más entusiasma a los investigadores, sugiere que es necesario hacer que la IA adquiera los valores humanos, programándolos y enseñando a la propia IA a aprenderlos.

Durante su estancia en Cambridge, Tallin se reunió con Seán Ó héigeartaigh, director ejecutivo del CSER y con Matthijs Maas, investigador de inteligencia artificial de la Universidad de Copenhague. En esta reunión surgieron diversos cuestionamientos, de los cuales los más relevantes fueron si se quiere dominar a la IA y si la Inteligencia Artificial tendría derechos. Para Tallin lo importante de la IA es lo que hace y darse cuenta de que sus funciones están en manos de los seres humanos y es importante que sea así.

---

[1] La inteligencia artificial amigable (también llamada IA amigable o FAI por sus siglas en inglés) es una IA fuerte (IAF) e hipotética, puede tener un efecto positivo más que uno negativo sobre la humanidad. El término fue acuñado por Eliezer Yudkowsky para discutir acerca de los agentes artificiales súper inteligentes que de manera confiable implementan los valores humanos (https://es.wikipedia.org/wiki/Inteligencia_artificial_amigable [5]).

Datos cruciales:

1. Jann Tallin cofundó Skype en 2003, en donde desarrolló el backend (parte del desarrollo web que se encarga de que toda la lógica de una página web funcione).

2. El término Inteligencia Artificial fue acuñado en 1956, una década después de la creación de las primeras computadoras digitales.

3. Tallin ha donado al instituto donde era investigador Yudkwosky más de 600,000 dólares. Este instituto cambio su nombre en 2013 por Machine Intelligence Research Institute (MIRI). Asimismo ha donado más de 310,000 dólares al Future of Humanity Institute (FHI) de la

Universidad de Oxford.

4. Peter Thiel donó 1.6 millones de dólares a MIRI mientras que Elon Musk donó 10 millones a Future of Life Institute.

5. Tallin cofundó el Cambridge Centre for the Study of Existential Risk (CSER) en 2012 con un capital inicial de 200,000 dólares.

6. En 2012 Nick Bostrom quien fue confudador del FHI, propuso en un documento aislar a la superinteligencia en un tanque de retención a la vez de restringirlo al responder preguntas.

Nexo con el tema que estudiamos:

El desarrollo de la inteligencia artificial ha traído consigo el cambio en la forma de producción y de consumo desde que se empezó a implementar. Es importante mencionar que el desarrollo de inteligencia artificial ha sido prioritario para países como China, Rusia y Estados Unidos, haciendo saber que están dispuestos a usarla con otros fines, como la invasión de la privacidad o utilizarla como un arma bélica. La IA es fundamental para ver el papel que jugaran cada una de las naciones.

Pero, como se vio en el texto, el desarrollo de esta tecnología trae nuevos retos, los cuales están sumamente a problemas éticos. Ya que se ha visto que la se ha seguido con el avance de la investigación en aspectos técnicos de la IA pero no en aspectos éticos.

Links
[1] https://www.theguardian.com/technology/2019/mar/28/can-we-stop-robots-outsmarting-humanity-artificial-intelligence-singularity
[2] http://let.iiec.unam.mx/taxonomy/term/16
[3] http://let.iiec.unam.mx/taxonomy/term/18
[4] http://let.iiec.unam.mx/descriptores-let/tecnolog%C3%ADas-militares-tecnolog%C3%ADas-de-uso-dual
[5] https://es.wikipedia.org/wiki/Inteligencia_artificial_amigable