

---

Tecnología: Crean una inteligencia artificial psicópata

06/06/2018



Tras las inteligencias artificiales racistas y sexistas, llegó la hora de la psicópata, que tras "ver" imágenes macabras da respuestas inquietantes al método de psicodiagnóstico conocido como "Test de Rorschach", utilizado desde 1921 para evaluar la personalidad.

El experimento, desarrollado por el Media Lab del MIT, tiene como objetivo demostrar que el modo en que se entrenan estos algoritmos puede influenciar negativamente su comportamiento posterior.

El algoritmo que le dio vida a Norman es capaz de observar una foto y comprender el contenido después de haber construido una base de datos con imágenes. Los investigadores lo "entrenaron" con imágenes de personas moribundas tomadas de un directorio del sitio Reddit, mientras que otra inteligencia artificial se entrenó con fotos de personas y animales en distintas situaciones, ninguna dramática.

Ambas fueron luego sometidas al famoso Test de Rorschach, en el que se les pidió interpretar las manchas de tinta para evaluar su personalidad. La diferencia entre ambas interpretaciones resultó evidente.

La mancha que la inteligencia "normal" interpretaba como "un grupo de pájaros sobre una rama" para Norman era "un hombre que sufría una descarga eléctrica. Un "florero" se convertía en "un hombre a quien le dispararon a muerte" y "un guante de béisbol" fue interpretado como "un hombre asesinado por una ametralladora". "Norman nace del hecho que los datos que son utilizados para entrenar un algoritmo influyen de modo significativo el comportamiento", dijeron los científicos que desarrollaron la inteligencia artificial.

Agregaron que, por lo tanto, "cuando las personas hablan de comportamientos racistas o incorrectos el problema no es el algoritmo en sí, sino en los datos usados".

El problema ya surgió en distintas oportunidades en el pasado. El ejemplo más famoso es el de "Tay", la

inteligencia artificial desarrollada por Microsoft bajo la forma de usuario de Twitter, que pocos días después de su lanzamiento se transformó en racista a causa de los comentarios que leía.

Un estudio publicado en la revista Science hace aproximadamente un año demostró luego que los algoritmos incorporan los mismos prejuicios de quien los programa. Y, por ejemplo, asocian términos muy negativos a las minorías étnicas o los referidos a labores domésticas con las mujeres.

---