شرکت IBM با واترمارک از مالکیت معنوی مدل هوش مصنوعی خود محافظت می کند - دیجیاتو

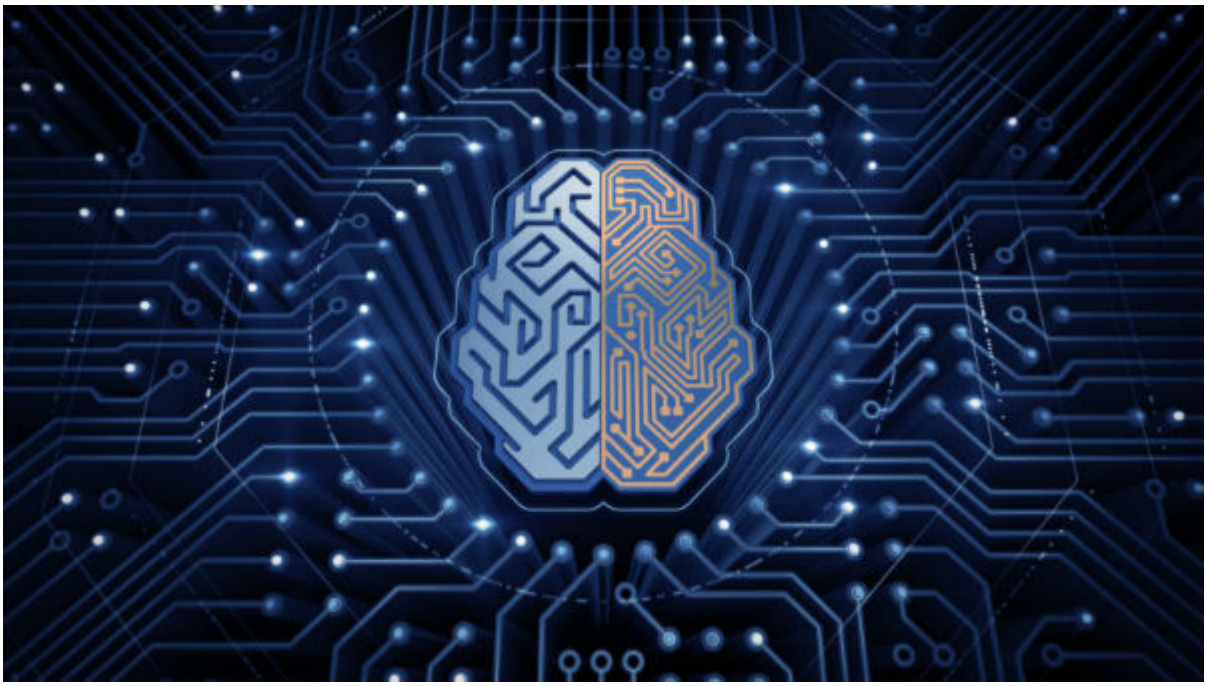
علی باقرزاده | یکشنبه، ۳۱ تیر ۱۳۹۷

محتوای دیجیتال از جمله تصاویر و ویدیوها به راحتی سرقت می شوند و به همین دلیل بسیاری از تولید کنندگان محتوا از روش هایی از جمله واترمارک استفاده می کنند تا در صورت سرقت محتوا، ساده تر مالکیت خود را ثابت کنند. اکنون IBM نیز قصد دارد از روشی مشابه برای جلوگیری از به سرقت رفتن مدل های هوش مصنوعی خود بهره ببرد.

به عقیده IBM مدل های یادگیری عمیق با پیاده سازی بیشتر، با تهدیدهای بیشتری هم مواجه می شوند. با ایده این شرکت که اکنون در انتظار ثبت اختراع است و با الهام از روش های واترمارک ویدیوها و تصاویر ایجاد شده اند، امکان محافظت از مدل ها فراهم می شود.

استفاده از یک واترمارک در عکس شامل دو مرحله است: «تعبیه کردن آن در عکس» و «کشف» آن. در مرحله اول، مالک تصویر عبارت «کپی رایت» یا موردی که به چشم انسان ها غیر قابل تشخیص باشد، را در تصویر قرار می دهد.

در صورتی که این تصویر سرقت شود و توسط دیگران مورد استفاده قرار بگیرد، مالک می تواند در مرحله تشخیص، مالکیت خود را ثابت کند. همین ایده می تواند بر شبکه عصبی یادگیری عمیق نیز اعمال شود. با این حال روش مورد استفاده توسط IBM تا حدودی متفاوت است.



بر اساس روشی که IBM توسعه داده، امکان تأیید سرویس های مبتنی بر مالکیت شبکه عصبی یادگیری عمیق (DNN) به واسطه پرسش های ساده رابط برنامه نویسی کاربردی (API) فراهم می شود. به این منظور سه الگوریتم تولید واترمارک توسعه داده شده اند:

- تعبیه کردن داده های معنا دار در کنار داده های یادگیری به عنوان واترمارک
- تعبیه کردن نمونه داده های غیر ضروری به عنوان واترمارک
- تعبیه کردن نویز به عنوان واترمارک

این شرکت مدعی است آزمایش هایی با واکنش هایی غیر قابل انتظار ولی کنترل شده انجام داده و از تعدادی مجموعه داده از جمله MNIST، یک مدل DNN واترمارک بهره برده است.

IBM می گوید این اولین باری نیست که چنین ایده ای مطرح شده، اما روش های قبلی به علت نیاز به دسترسی به مؤلفه هایی از مدل برای تعیین مالکیت، با محدودیت هایی همراه بودند. از سوی دیگر روش IBM در برابر حذف واترمارک ایمن است ولی در صورت پیاده سازی مدل به صورت بدون دسترسی آنلاین امکان تشخیص مالکیت مدل وجود ندارد.

این شرکت در نهایت اعلام کرده که در حال حاضر تصمیم گرفته از فریم ورک واترمارک خود به صورت داخلی بهره ببرد و در جستجوی این است که چگونه آن را به مشتریانش ارائه دهد.

تماشا کنید: پلان؛ هوش مصنوعی چگونه ما را زیر نظر می گیرد؟

دیجیاتو