Artık Yapay Zekâyı Şaşırtmanın Yolunu Arıyoruz

Yapay zekâyı kullanarak sahte fotoğraflar üretmek bazen eğlenceli olsa da bir sanatçının veya fotoğrafçının eserlerinin yapay zekâyı eğitmek için izinsiz kullanılması hiç hoş değil. Benzer şekilde, böyle bir fotoğrafta yer alan bir kişi olarak, yapay zekânın sizin benzerinizi tuhaf bir şekilde tasvir etmesini istemezsiniz.

İşte bu noktada, MIT Bilgisayar Bilimi ve Yapay Zekâ Laboratuvarı (CSAIL) tarafından geliştirilen "PhotoGuard" adlı yeni bir araç devreye giriyor. Bu araç, fotoğrafları izinsiz yapay zekâ kullanımına karşı korumayı hedefliyor.

PhotoGuard, bir resimdeki en küçük bilgi birimi olan pikselleri kullanarak çalışıyor. İnsan gözü tarafından algılanamayacak şekilde belirli pikselleri değiştirerek yapay zekâyı şaşırtıyor. Bu işlemi gerçekleştirmek için iki temel yöntem kullanıyor:

Encoder saldırısı yapay zekânın dijital dosyayı okumasını zorlaştırıyor.

Diffusion saldırısı ise yapay zekânın resmi farklı görmesine neden oluyor.

Görseldeki örnekte, normal fotoğraf takım elbiseli düğün fotoğrafına dönüştürülebilirken korumalı fotoğrafta yapay zekâ başarısız oluyor

PhotoGuard kesin bir çözüm sunmasa da bu alanda ortaya çıkacak diğer projelerin öncüsü olabilir. PhotoGuard benzeri araçların geliştirilmesi, yapay zekâ tarafından izinsiz kullanılan fotoğraflarla ilgili yaşanan birçok dava düşünüldüğünde oldukça makul görünüyor. Örneğin, Getty Images, Stability AI'a 12 milyondan fazla görselini izinsiz veya ücretsiz kopyaladığı için dava açtı.

Eğer PhotoGuard'u kendi başınıza denemek istiyorsanız, ilgili kodu GitHub'da bulabilirsiniz. Ancak bu teknolojinin henüz olgunlaşmadığını bilip ona göre hareket edin!

https://github.com/MadryLab/photoguard