Microsoft pulls open facial recognition dataset after Financial Times investigation

Earlier this week, Microsoft removed a database of more than 10 million faces, intended as a test and training dataset for facial recognition algorithms, according to a report by the *Financial Times*. Known as MS Celeb, the database contained more than 10 million images of roughly 100,000 people, largely scraped from publicly available online sources. While no individual photo in the dataset was difficult to find, the volume of images and the structured data accompanying them made the dataset extremely useful in training programs to recognize a person's face across different photos.

The takedown came after an earlier *Financial Times* investigation found that many of the people represented in the dataset were not aware of it and did not consent to having their pictures used. A number of experts speculated that the dataset might encounter legal issues under the General Data Protection Regulation, which imposes significant requirements for the storage and transfer of a subject's personal data.

Notably, Microsoft did not announce the removal of the dataset, and downplayed its significance in a comment to *FT*. "The site was intended for academic purposes," the spokesperson said. "It was run by an employee that is no longer with Microsoft and has since been removed."

Two similar datasets run by Duke University and Stanford were also taken down in the wake of the *FT*'s reporting.

https://www.theverge.com/2019/6/7/18656800/microsoft-facial-re

cognition-dataset-removed-privacy