IBM to release world's largest facial analytics dataset

June 27 2018, by Aleksandra Mojsilovic, John R. Smith

Society is paying more attention than ever to the question of bias in artificial intelligence systems, and particularly those used to recognize and analyze images of faces. At IBM, we are taking the following actions to ensure facial recognition technology is built and trained responsibly:

(1) One of the biggest issues causing bias in the area of facial analysis is the lack of diverse data to train systems on. So, this fall, we will make publicly available the following as a tool for the technology industry and research community:

1. A facial attribute and identity training dataset of over 1 million images to improve facial analysis system training built by IBM Research scientists. It's annotated with attributes and identity, leveraging geo-tags from Flickr images to balance data from multiple countries and active learning tools to reduce sample selection bias. Currently, the largest facial attribute dataset available is 200,000 images so this new dataset with a million images will be a monumental improvement. Additionally, data sets available today only include attributes (hair color, facial hair, etc) or identity (identifying that 5 images are of the same person)—but not both. This new dataset changes that to make a single capability to match attributes to an individual.
2. A dataset which includes 36,000 [facial images](#) – equally distributed across all ethnicities, genders, and ages to provide a more diverse dataset for people to use in the evaluation of their technologies. This will specifically help algorithm designers to

identify and address bias in their facial analysis systems. The first step in addressing bias is to know there is a bias—and that is what this dataset will enable.

(2) Earlier this year, we substantially increased the accuracy of our Watson Visual Recognition service for facial analysis, which demonstrated a nearly ten-fold decrease in error-rate for facial analysis. And, we are continuing to drive continual improvements. A technical workshop is being held (by IBM Research in collaboration with University of Maryland) to identify and reduce bias in facial analysis on Sept 14, 2018 in conjunction with ECCV 2018. The results of the competition using the IBM facial image dataset will be announced at the workshop. Furthermore, our researchers continue to work with a broad range of stakeholders, users and experts to understand other biases and vulnerabilities that can affect AI decision-making, so that we can continue to make our systems better."

AI holds significant power to improve the way we live and work, but only if AI systems are developed and trained responsibly, and produce outcomes we trust. Making sure that the system is trained on balanced data, and rid of biases is critical to achieving such trust.

As the adoption of AI increases, the issue of preventing bias from entering into AI systems is rising to the forefront. We believe no technology—no matter how accurate—can or should replace human judgement, intuition and expertise. The power of advanced innovations, like AI, lies in their ability to augment, not replace, human decision-making. It is therefore critical that any organization using AI—including visual recognition or video analysis capabilities—train the teams working with it to understand bias, including implicit and unconscious bias, monitor for it, and know how to address it.

As a company that leads in driving diversity and inclusion in the

corporate world, discrimination of any kind is against IBM's values. We are deeply committed to ensuring AI technologies are developed without bias.

For more than a century, IBM has responsibly ushered revolutionary technologies into the world. We are dedicated to delivering AI services that are built responsibly, are unbiased and explainable. Our business has been guided by a set of Trust and Transparency Principles, which includes our firm belief that companies advancing AI have a responsibility to address the issue of bias head-on. And we are continually working to evaluate and update our services, advancing them in a way that is trustworthy and inclusive.

Provided by IBM