Facebook's self-supervised AI offers potential benefits—if it can overcome bias concerns

Article

The news: Facebook revealed a self-supervised artificial intelligence model it claims can accurately learn to categorize Instagram images with less human assistance than before.

Here's how it works: Researchers at Facebook fed the AI, called SEER, over 1 billion unlabeled images extracted from public IG accounts. Using self-supervised learning—a method where a machine learns to train itself without human data labeling—SEER achieved a classification accuracy score of 84.2%. outperforming "the most advanced, state-of-the-art self-supervised systems," per Facebook.

What's next?: While SEER is still in its early stages, Facebook believes it can bring about real-world benefits. Here are some of SEER's possible use cases:

- **Reducing harmful content:** Facebook already relies on AI to detect child pornography, violence, and other harmful content on its platform. A more robust self-supervised AI could further help the company parse through troves of online detritus and reduce some of the burden placed on its beleaguered content moderators.

- **Scalability:** Facebook claims SEER and other self-supervised programs can sift through significantly larger and more complex datasets than AI trained using machine learning or deep learning approaches.

- **Speed:** Freed from the laborious task of creating labels for objects at the start of a project, Facebook's AI can potentially analyze data more quickly.

- **Auto-generating alt text:** A spokesperson told CNBC the new AI could analyze images and automatically generate text to assist visually impaired people and automatically create hashtags.

- **And of course, there's digital monetization:** The AI could help Facebook categorize items sold across in its Facebook Marketplace.

The Big Picture: Ever-increasing data sharing by users will likely lead to rapid AI advancement.

Self-supervised programs are data-intensive and reportedly require around 100x more images to achieve comparable levels of accuracy than AI's trained using human-generated labels. Luckily for Facebook and others interested in building AI, there's no shortage of data. A 2020 IDC report predicts the amount of data created globally in the next three years will exceed the amount created in the past three decades. At the same time, Cisco expects the total number of users on the internet to increase from 3.9 billion in 2018 to around 5.3 billion

in 2023. All this means future AI programs will benefit both from new users to harvest content from, and a willingness by those users to upload more photos, texts, videos, and voice recordings than ever before.
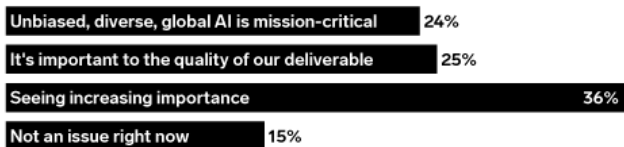
Why this could backfire: Facebook's technical advances in AI still leaves lingering foundational questions around algorithmic bias unaddressed.

Facebook's choice to train its algorithm on Instagram images could bias the AI toward younger demographics with greater access to social media and mobile apps, with no guarantee of accuracy for groups underrepresented in that data set, according to Nikita Aggarwal, a research associate at the Oxford Internet Institute. "There's a difference between developing AI systems that can identify correlations in data to classify images," Aggarwal told New Scientist, "and systems that can actually understand the meaning and context of what they're doing or indeed reason about it."

While Facebook claims self-supervised AI models removal of human-generated labels could "mitigate some of the biases" inherent in data curation, that wouldn't necessarily address biases baked into the society—and therefore the data—itself. Even if self-supervised AI programs overcome accuracy issues, they still risk replicating the same cultural and social biases endemic to the type of prejudiced data it's being trained on—garbage in, garbage out, resulting in AI reinforcing the same prejudices from the data it's fed.

How Do Companies Worldwide View Data Diversity, Bias Reduction and Global Scale for Their AI?
% of respondents, May 2020

Unbiased, diverse, global AI is mission-critical	24%
It's important to the quality of our deliverable	25%
Seeing increasing importance	36%
Not an issue right now	15%

Source: Appen, "The State of AI and Machine Learning," June 23, 2020

257111   eMarketer | InsiderIntelligence.com

INSIDER INTELLIGENCE | eMarketer

Page 3