OPINION

THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE MESSENGER

AI Data Grab: How Private Is Your Personal Information?

Published 09/26/23 07:00 AM ET

Sheldon H. Jacobson, Ph.D.

Morsa Images/ Getty Images

Anytime you visit a website and sign up for access privileges, you are asked to agree to certain terms and conditions of service. Most of us spend little time reading the fine print and perfunctorily click the requisite button so we can get on with what we need at the site. We effectively do so at our own risk.

Artificial intelligence (AI) is changing the privacy landscape, as some websites would like your permission to use your browsing and personal information data to train their AI engines.

This occurred recently with Zoom, which changed its terms of service to give the company such access and the ability to use user data to train its AI models. Once this became more widely known, some users called foul, with many searching for alternative platforms to virtual video meetings and communications. In response, Zoom retracted this clause in its service agreement. Time will determine if the situation was adequately repaired and how much the user backlash and privacy concerns potentially damaged its reputation.

If AI models are engines, then data is the fuel that drives them. Names like machine learning, deep learning, and reinforcement learning label such models for learning. Yet, where does the learning come from? It comes from data.

Data have become a precious commodity on the AI landscape. Without a rich and reliable source of data, AI models become at best ineffective, or at worst, useless. The value of AI is dependent on data.

When companies and their websites constantly have access to terabytes of data, they have hit a "data gusher." That is why organizations like Google and those who partner with them are ideally positioned to make the most significant AI advances. They have enormous streams of data available that allow AI models to learn, which in turn, makes these models more useful. Given that Google has over 92% of the search engine market, this creates an enormous platform for accessing data. Such a monopoly is one reason why the Department of Justice filed an antitrust lawsuit against the search engine giant.

Zoom was hoping to tap into the large amount of data that they have access to from its users. Unfortunately for the company, its hope was not well received by users in response to the perceived data-grab threat.

AI motivates organizations to monetize all information of users to gain competitive advantages in their market. The need for data will only grow with time, since without such data, AI models are ineffective. Relying on access to data that is only voluntarily provided may be insufficient in the AI arms race.

This begs the question: Should consumers expect compensation for the data that they generate for companies?

Much like how the gig economy has allowed people to monetize their belongings, like their cars (through ride services like Lyft and Uber) and homes (through Airbnb), should consumers be entitled to monetize their web activities and personal preferences that they share when online?

Polling companies routinely rely on people to voluntarily provide their opinions to inform organizations on purchasing trends, political preference, and attitudes. If people wish to monetize their views, compensation will be needed. The leap forward to include their personal information is but a small step within the monetization paradigm. The question is in what form does such compensation get transferred? Will access to information alone be sufficient?

With AI models in need of data, the race to secure any such data has already begun. For example, the Worldcoin project is paying people help develop digital ID (in the form of a promise of cryptocurrency) for their iris scan. Tangentially, Amazon has initiated biometric-based forms of payment using palm scans as a convenience for members.

The fact is that most people's personal information is widely available and already being shared, often without their knowledge. The concept of online privacy exists, but protecting all information is near impossible. Biometrics, the ultimate personal information, is clearly on the data-grab horizon.

The good news is that most of the information that we share and others access is not being used with nefarious intent. Using it to train an AI model will not inflict any personal harm and may even have some incremental benefits when the AI models inform people's decisions and choices.

The bigger concern is when cyber criminals steal personal information like passwords and financial information that exposes a person to risks that could lead to losses and damage.

Although Zoom managed its AI data collection without finesse, the potential harm of such actions is likely to be minimal, if any. For many users, that does not exonerate them from the way it was handled and poorly communicated. It does, however, remind everyone that the internet has been and will continue to be the "Wild West" for personal information, and escaping the consequences is exceedingly difficult, if at all possible.

Sheldon H. Jacobson, Ph.D., is a founder professor in computer science at the University of Illinois Urbana-Champaign. A data scientist, he applies his expertise in data-driven risk-based decision-making to evaluate and inform public policy and public health.