How synthetic data changes the AI game

Synthetic data reduces dependency on real-world data, lowering costs, improving the quality of machine learning models and accelerating time to market

Miguel Ferreira
Head of Data Science, CVEDIA

DECIDING HOW TO LEVERAGE government data to increase the safety and security of our country is a central question in computer vision. Most solutions rely on machine learning models that have been trained with real-world data. Unfortunately, 80% of the work required for an artificial intelligence project is collecting and preparing data. As a result, capturing and labeling the right data becomes a heavy resource burden.

That's why synthetic data is a game-changer in AI. It reduces the time and costs involved in training the models because it removes the need for manual collection and labeling. When they start on their machine learning journeys, most agencies have limited data. For example, a state government might collect data from one or two cities and then use it to train algorithms that are expected to work across the entire state. These algorithms result in heavily biased, unreliable systems because of a lack of diverse data. By contrast, synthetic data is designed to reduce bias by introducing myriad diverse scenarios and conditions, so that the algorithm can operate anywhere. It also becomes relatively easy to add extra features to systems that have already been deployed.

Filling the gaps in available information

Synthetic data aids in situations where agencies don't have enough information. For example, what happens to computer vision models when a defense adversary introduces a new armored vehicle or warfighters encounter a new type of operating environment? If the algorithms rely exclusively on real data, they have a difficult time adapting to changes or responding to brand-new scenarios. In situations like these, CVEDIA can create a 3D model of the vehicle and use it to generate vast amounts of information, thereby increasing the AI's ability to detect and respond to those new features.

Reducing cognitive overload

In addition, synthetic data can help in situations where there's too much information. Given all the satellites, cameras and sensors that are continually producing visual data, it's easy to become overwhelmed. The use of synthetic data in computer vision solutions can reduce cognitive overload by extrapolating on important information, so agencies can understand what the data is telling them and where they need to look.

Most government leaders believe technology advances like these are only available to cutting-edge companies with high budgets for R&D. However, the government has the same access to neural networks and high-performing hardware that the private sector does — and CVEDIA's synthetic data can reduce AI project overhead by up to 90%. With the help of synthetic data, agencies can make the most of computer vision and machine learning algorithms today. ◼

Miguel Ferreira is head of data science at CVEDIA.