

Are You Still Using Real Data to Train Your AI? | IEEE Spectrum

By Eliza Strickland

April 26, 2022

It may be counterintuitive. But some argue that the key to training AI systems that must work in messy real-world environments, such as self-driving cars and warehouse robots, is not, in fact, real-world data. Instead, some say, synthetic data is what will unlock the true potential of AI. Synthetic data is generated instead of collected, and the consultancy Gartner has estimated that 60 percent of data used to train AI systems will be synthetic. But its use is controversial, as questions remain about whether synthetic data can accurately mirror real-world data and prepare AI systems for real-world situations.

Nvidia has embraced the synthetic data trend, and is striving to be a leader in the young industry. In November, Nvidia founder and CEO Jensen Huang announced the launch of the Omniverse Replicator, which Nvidia describes as “an engine for generating synthetic data with ground truth for training AI networks.” To find out what that means, *IEEE Spectrum* spoke with Rev Lebedev, vice president of simulation technology and Omniverse engineering at Nvidia.

Source: [Are You Still Using Real Data to Train Your AI? | IEEE Spectrum](#)