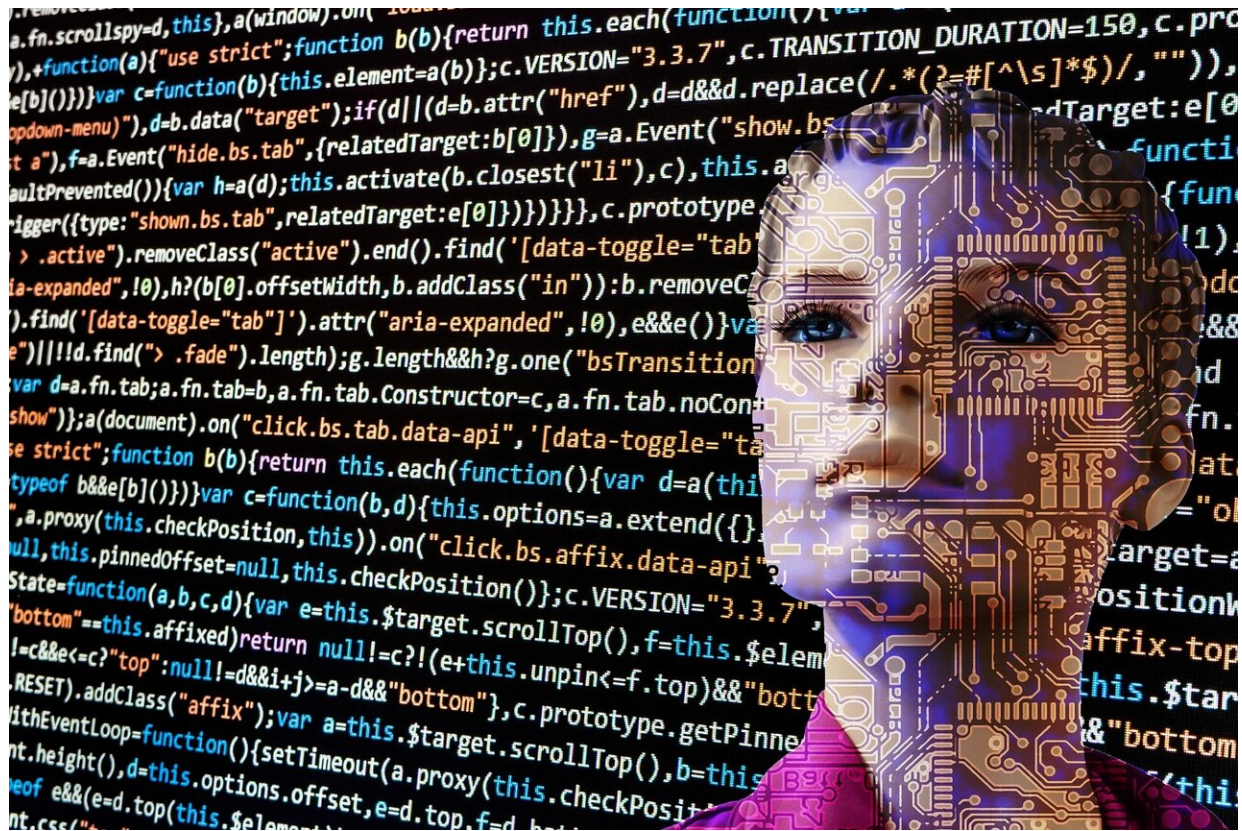# Researchers offer standards for studies using machine learning

August 30 2021

Credit: CC0 Public Domain

Researchers in the life sciences who use machine learning for their studies should adopt standards that allow other researchers to reproduce their results, according to a comment article published today in the

journal *Nature Methods.*

The authors explain that the standards are key to advancing scientific breakthroughs, making advances in knowledge, and ensuring research findings are reproducible from one group of scientists to the next. The standards would allow other groups of scientists to focus on the next breakthrough rather than spending time recreating the wheel built by the authors of the original study.

Casey S. Greene, Ph.D., director of the University of Colorado School of Medicine's Center for Health AI, is a corresponding author of the article, which he co-authored with first author Benjamin J. Heil, a member of Greene's research team, and researchers from the United States, Canada, and Europe.

"Ultimately all science requires trust—no scientist can reproduce the results from every paper they read," Greene and his co-authors write. "The question, then, is how to ensure that machine-learning analyses in the life sciences can be trusted."

Greene and his co-authors outline standards to qualify for one of three levels of accessibility: Bronze, silver, and gold. These standards each set minimum levels for sharing study materials so that other life science researchers can trust the work, and if warranted, validate the work and build on it.

To qualify for a bronze standard, life science researchers would need to make their data, code, and models publicly available. In machine learning, computers learn from training data and having access to that data enables scientists to look for problems that can confound the process. The code tells future researchers how the computer was told to carry out the steps of the work.

In machine learning, the resulting model is critically important. For future researchers, knowing the original research team's model is critical for understanding how it relates to the data it is supposed to analyze. Without access to the model, other researchers cannot determine biases that might influence the work. For example, it can be difficult to determine whether an algorithm favors one group of people over another.

"Being unable to examine a model also makes trusting it difficult," the authors write.

The silver standard calls for the data, models, and code provided at the bronze level, and adds more information about the system in which to run the code. For the next scientists, that information makes it theoretically possible that they could duplicate the training process.

To qualify for the gold standard, researchers must add an "easy button" to their work to make it possible for future researchers to reproduce the previous analysis with a single command. The original researchers must automate all steps of their analysis so that "the burden of reproducing their work is as small as possible." For the next scientists, this information makes it practically possible to duplicate the training process and either adapt or extend it.

Greene and his co-authors also offer recommendations for documenting the steps and sharing them.

The *Nature Methods* article is an important contribution to the continuing refinement of the use of machine learning and other data-analysis methods in health sciences and other fields where trust is particularly important. Greene is one of several leaders recently recruited by the CU School of Medicine to establish a program in developing and applying robust data science methodologies to advance biomedical research,

education, and clinical care.

More information: Benjamin J. Heil et al, Reproducibility standards for machine learning in the life sciences, *Nature Methods* (2021). DOI: 10.1038/s41592-021-01256-7