

Artificial intelligence produces data synthetically to help treat diseases like COVID-19

25 June 2020



The possibility to produce synthetic data solves many problems and helps develop for example better treatment methods. Credit: Matti Ahlgren / Aalto University

Data driven technologies and "big data" are revolutionizing many industries. However, in many areas of research—including health and drug development—there is too little data available due to its sensitive nature and the strict protection of individuals. When data are scarce, the conclusions and predictions made by researchers remain uncertain, and the coronavirus outbreak is one of these situations.

"When a person gets sick, of course, they want to get the best possible care. Then it would be important to have the best possible methods of personalized healthcare available," says Samuel Kaski, Academy Professor and the Director of the Finnish Center for Artificial Intelligence FCAI.

However, developing such methods of personalized healthcare requires a lot of [data](#), which is difficult to obtain because of ethical and [privacy issues](#) surrounding the large-scale gathering of personal data. "For example, I myself

would not like to give insurance companies my own genomic information, unless I can decide very precisely what the [insurance company](#) will do with the information," says Professor Kaski.

To solve this issue, researchers at FCAI have developed a new machine learning-based method that can produce [research data](#) synthetically. The method can be useful in helping develop better treatments and to understand the COVID-19 disease, as well as in other applications. The researchers recently released an application based on the method that allows academics and companies to share data with each other without compromising the privacy of the individuals involved in the study.

Many industries want to protect their own data so that they do not reveal trade secrets and inventions to their competitors. This is especially true in [drug development](#), which requires lots of financial risk. If pharmaceutical companies could share their data with other companies and researchers without disclosing their own inventions, everyone would benefit.

When researchers have synthetic data, they start understanding COVID-19 better

The ability to produce data synthetically solves these problems. In their previous study, which is currently being peer reviewed, FCAI researchers found that synthetic data can be used to draw as reliable statistical conclusions as the original data. It allows researchers to conduct an indefinite number of analyses while keeping the privacy of the individuals involved in the original experiment secure.

The application that was published at the end of June works like this: The researcher enters the

original data set into the application, from which the application builds the synthetic dataset. They can then share their data to other researchers and companies in a secure way.

The application was released on the fastest possible schedule so that researchers investigating the coronavirus pandemic would have access to it as early as possible. Researchers are further improving the application, to make it easier to use and add other functionality. "There are still many things we don't know about the new coronavirus: for example, we do not know well enough what the virus causes in the body and what the actual risk factors are. When researchers have synthetic data, we start understanding these things better," says Kaski.

FCAI researchers are now working on a project in which they use synthetic data to construct a model that, based on certain biomarkers, predicts whether a test subject's coronavirus test is positive or negative. Biomarkers can be for example certain types of molecules, cells, or hormones that indicate a disease.

"The original data set with which we do this has been publicly available. Now we are trying to reproduce the results of the original research with the help of synthetic data and build a predictive model from the synthetic data that was achieved in the original research," explains Joonas Jälkö, doctoral researcher at Aalto University.

More information: Jälkö et al., Privacy-preserving data sharing via probabilistic modeling, (2020). arXiv:1912.04439 [stat.ML].
arxiv.org/abs/1912.04439

Provided by Aalto University

APA citation: Artificial intelligence produces data synthetically to help treat diseases like COVID-19 (2020, June 25) retrieved 28 May 2022 from <https://techxplore.com/news/2020-06-artificial-intelligence-synthetically-diseases-covid-.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.