

March 16, 2020

Contact: Lynne Weil, 202-440-2020

New National Online Coronavirus Resource: CSET Experts Lead in Creating National Database Assisted by Machine Learning

Washington, DC -- The White House is tapping the expertise of researchers from Georgetown's Center for Security and Emerging Technology (CSET) to determine how data and open research can be used to address the COVID-19 pandemic.

Today the Allen Institute for AI, working with the administration's Office of Science and Technology Policy (OSTP), CSET and other partners, released [CORD-19](#), a publicly available repository of information to assist in tracking the virus. CORD-19 stands for COVID-19 Open Research Dataset.

The repository is linked to the World Health Organization (WHO) database of publications on coronavirus disease and other resources.

Speeding Up Response

Along with the Allen Institute, Chan Zuckerberg Initiative, Microsoft Research, and the National Library of Medicine of the National Institutes of Health, CSET is working to prepare and distribute a set of full-text scholarly literature about the coronavirus family of viruses.

"With this step, we've made available full-text, machine-readable resources to help speed response to this global crisis," said Dewey Murdick, CSET's director of data science, who has led the team at the request of OSTP. "The worldwide machine learning community now has the opportunity to apply recent advances in natural language processing to find answers to important questions about this infectious disease."

"Once the crisis has passed," he added, "we hope this project will offer proof of concept for ways to use machine learning to advance scientific research."

All Hands on Deck

WHO, which characterized coronavirus as a pandemic on March 11, indicates that as of today there are nearly 165,000 cases in 146 countries and 6,470 deaths.

The pandemic is drastically changing how people go about their daily lives, with increasing calls for social distancing. Georgetown [moved all its classroom instruction to virtual learning environments](#) today, as have many universities across the country.

U.S. Chief Technology Officer Michael Kratsios previewed the new database at a Wednesday meeting.

“The White House’s top priority is ensuring the safety and health of the American people amid the COVID-19 outbreak,” Kratsios said in a statement. “Cutting-edge technology companies and major online platforms will play a critical role in this all-hands-on-deck effort. Today’s meeting outlined an initial path forward and we intend to continue this important conversation.”

CORD-19’s more than 29,000 articles, 13,000 of which have full text, contain a wealth of information about the novel coronavirus (also known as COVID-19 and SARS-CoV-2) and related viruses, and will continue to be updated as new insights are published in archival services and peer-reviewed publications.

The other resources the repository will be linked to include [Microsoft Academic Graph \(COVID-19 resource page\)](#), [Dimensions \(COVID-19 resource page\)](#), [PubMed](#) and [Semantic Scholar](#) services.