

The Story of the COVID-19 Open Research Dataset (CORD-19)

Oren Etzioni

CEO, Allen Institute for AI (AI2)
Professor, University of Washington

 Allen Institute for AI

AI evokes fear



**AI is a tool, not a being, and it's up to us to
employ this tool for the common good**

Allen Institute for AI (AI2)

- Non-profit Research Institute founded in 2014 by Paul Allen
- Mission: AI for the Common Good
- Leader in Computer Vision, Natural Language Processing, Deep Learning— 300+ papers, 9 best paper awards
- 150+ AI PhDs, researchers, engineers, and support staff



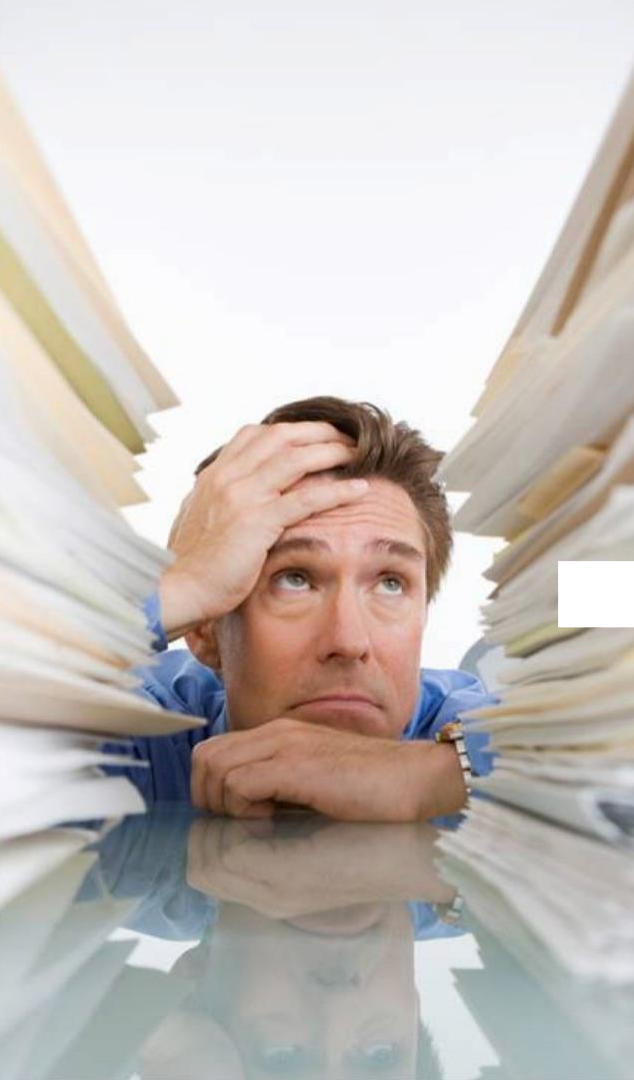
Information Overload - Scholars can't keep up



- 3M+ new papers per year
- Academics read ~250 papers per year

"The amount of stuff to absorb is truly overwhelming"

- Researcher, Brain Development Laboratory, University of Washington



Semantic Scholar

Cut through the clutter.

Home in on key papers, citations, and results.

 Find it fast

Try: [Open information extraction](#) [POS tagging](#) [Dependency parsing](#)



Launched In Computer Science

3M Papers

2015

Expanded To Neuroscience

10M Papers

2016

Expanded To BioMedicine

40M Papers

1M Monthly Active Users

2017

Scientific Content Graph

Data sets

Research software

Slides, videos

Clinical trials

News, blogs, and

Tweets

2018

Expanded To All Scientific Fields of Study

186M+ Papers

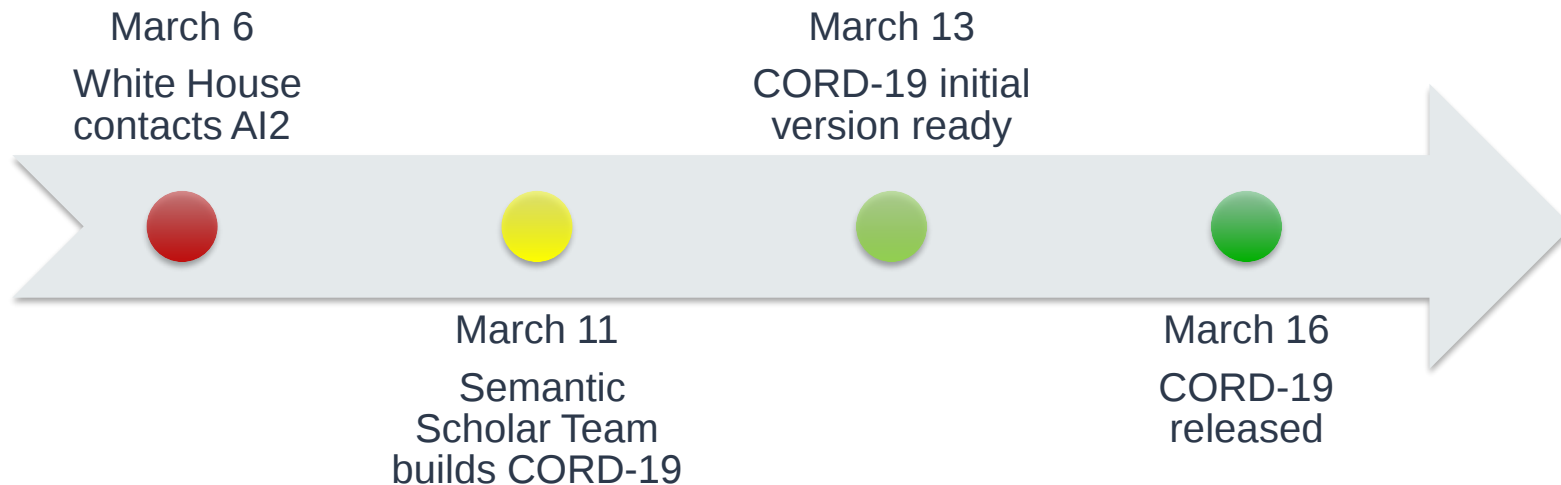
8M+ Monthly Active Users

550+ Partnerships

CORD-19!

2019 & 2020

The Inception of CORD-19 (10 days in March)



Chan
Zuckerberg
Initiative



bioRxiv
beta
THE PREPRINT SERVER FOR BIOLOGY

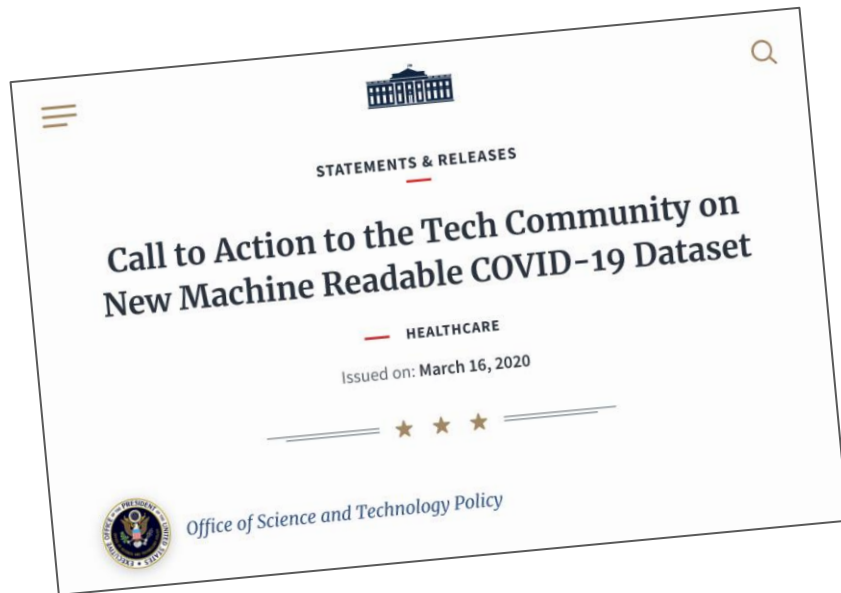
medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES



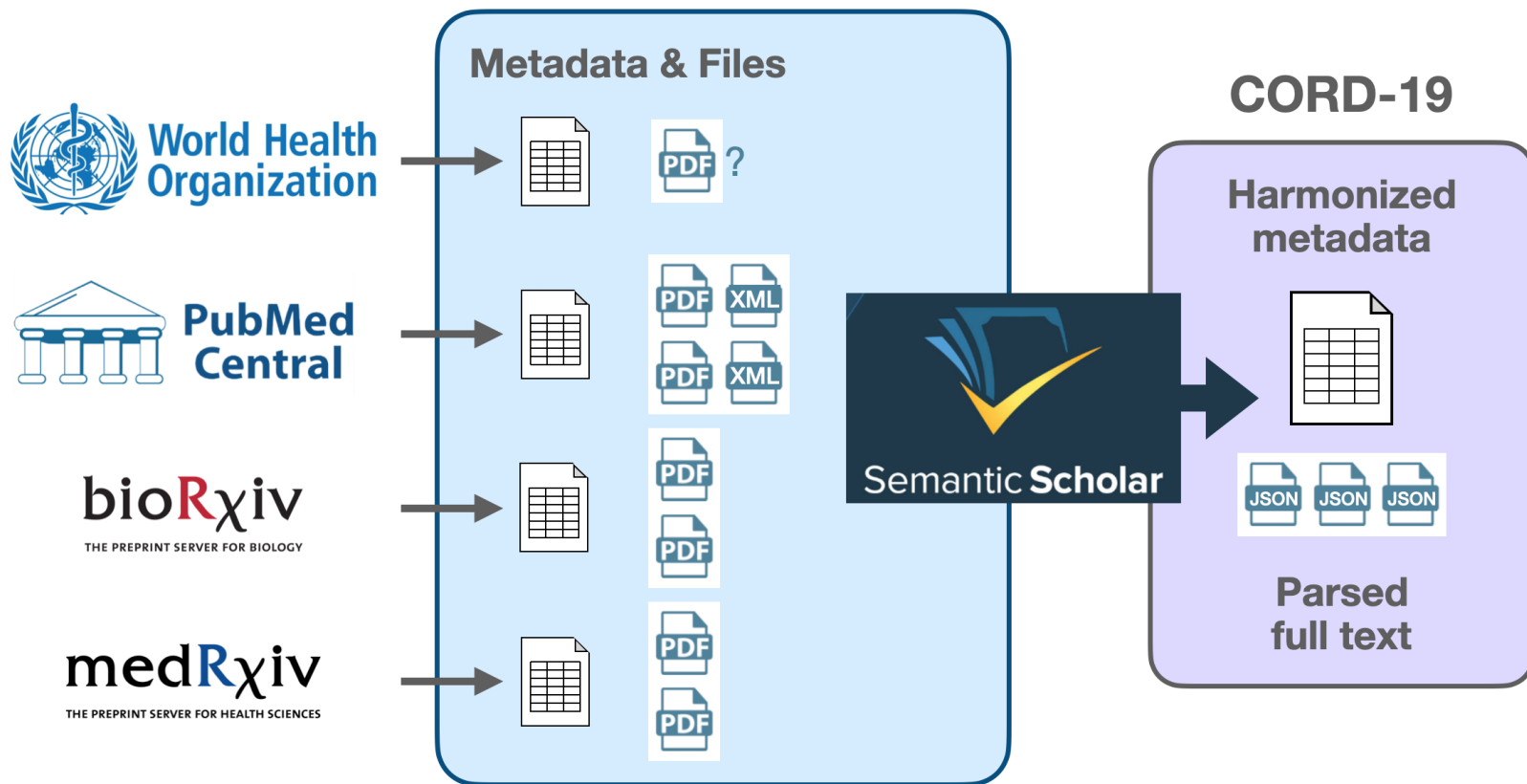
kaggle



COVID-19 Open Research Dataset (CORD-19)



From Human-Readable PDF Files to Machine-Readable Data



CORD-19: Current Status

- 128,000 papers citations
- Text and tables are now machine readable
- Meta data includes citations links
- Updated daily

Snapshots of a few AI tools using CORD-19

CovidAsk Question-Answering System from Korea University

Search

Latency: 1019ms

Answers | Related Entities

... It is stable in faeces at room temperature for at least 1-2 days and can be stable in infected patients for up to 4 days. **Heat at 56 degree Celsius kills the SARS coronavirus at around 10000 units per 15 minutes.** Thus, temperature is an important factor in survival of COVID19 virus and this article focuses on understanding the relationship between temperature and COVID19 transmission from the data available between January-March 2020.

Effects of temperature on COVID-19 transmission. Mar, 2020

... Coronavirus disease 2019 (COVID-19) was first identified in Wuhan, China in December 2019, and is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). **COVID-19 is a pandemic with an estimated death rate between 1% and 5%; and an estimated R 0 between 2.2 and 6.7 according to various sources.** As of March 28th, 2020, there were over 649,000 confirmed cases and 30,249 total deaths, globally. ...

Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. Apr, 2020

... Content COVID-19 seems not to be very different from SARS regarding its clinical features. **However, it has a fatality rate of 2.3%, lower than SARS (9.5%) and much lower than MERS (34.4%).** It cannot be excluded that because of the COVID-19 less severe clinical picture it can spread in the community more easily than MERS and SARS. ...

COVID-19, SARS and MERS: are they closely related? Mar, 2020

Clinical Microbiology and Infection.

... 64.2% were likely to defer change in IS, the reluctance being most with rituximab (58.3%) followed by cyclophosphamide (53.3%), anti-tumor necrosis factor alpha agents (52.4%) and Janus kinase inhibitors (34.39%). **Hydroxychloroquine was the preferred choice (81.9%) for the treatment of COVID-19 followed by protease inhibitors (22.1%) and intravenous immunoglobulin (8.1%). Chloroquine** was less preferred (19%). ...

Management of rheumatic diseases in the times of COVID-19 pandemic- perspectives of rheumatology practitioners from India. Apr, 2020

Various strains of mouse hepatitis virus (MHV) exhibit different pathogenic phenotypes. **Infection with the A59 strain of MHV induces both encephalitis and hepatitis, while the highly neurovirulent JHM strain induces a fatal encephalitis with little, if any, hepatitis.** The pathogenic phenotype for each strain is determined by the genetic composition of the viral genome, as well as the host immune response. ...

Both Spike and Background Genes Contribute to Murine Coronavirus Neurovirulence. Jun, 2006

Journal of Virology.

Scientific Fact Checking (AI2 & UW)



SciFact

CORD-19 Claim Verification

Use our AI system to verify COVID-19 claims against CORD-19 corpus.
Find scientific research that supports or refutes your claims.

Medical Disclaimer: The information contained herein should NOT be used as a substitute for the advice of an appropriately qualified and licensed physician or other health care provider. The tool is not a substitute for... [\(more\)](#)

Find out whether scientific research supports or refutes your claim:



Claims to try:

- [Hypertension is a common comorbidity seen in COVID-19 patients.](#)
- [SARS-CoV-2 binds ACE2 receptor to gain entry into cells.](#)
- [Lopinavir / ritonavir have exhibited favorable clinical responses when used as a treatment for coronavirus.](#)
- [Mass masking reduces COVID-19 transmission rates.](#)
- [Older age is a positive predictor of mortality in COVID-19.](#)
- [The coronavirus cannot thrive in warmer climates.](#)
- [Bilateral ground glass opacities are often seen on chest imaging in COVID-19 patients.](#)
- [Non-caucasian populations are at higher risk for COVID-19 infection due to inequality and chronic health conditions.](#)
- [Viable SARS-CoV-2 viral particles can remain on plastic for 2 - 3 days.](#)

[\(more\)](#)



Medical Disclaimer: The information contained herein should NOT be used as a substitute for the advice of an appropriately qualified and licensed physician or other health care provider. The tool is not a substitute for... [\(more\)](#)

Mass masking reduces COVID-19 transmission rates



1 research papers support this claim.

All (1)

Support (1)

Refute (0)

An Investigation of COVID-19 Spreading Factors with Explainable AI Techniques

SUPPORT (98%)

Xiuyi Fan, Siyuan Liu, Jiarong Chen, and 1 more · *Unknown source*

- Across-the-board, city lockdown and contact tracing are the two most effective measures.
- For ensuring $R_t < 1$, public wearing face masks is also important.
- Warm temperature helps for reducing the transmission..

[See full abstract](#)

Claims to try:

- Hypertension is a common comorbidity seen in COVID-19 patients.
- SARS-CoV-2 binds ACE2 receptor to gain entry into cells.
- Lopinavir / ritonavir have exhibited favorable clinical responses when used as a treatment for coronavirus.
- Mass masking reduces COVID-19 transmission rates.
- Older age is a positive predictor of mortality in COVID-19.
- The coronavirus cannot thrive in warmer climates.
- Bilateral ground glass opacities are often seen on chest imaging in COVID-19 patients.
- Non-caucasian populations are at higher risk for COVID-19 infection due to inequality and chronic health conditions.
- Viable SARS-CoV-2 viral particles can remain on plastic for 2 - 3 days.

Adaptive Research Feed (Semantic Scholar)

Organize new research into multiple feeds

Research Feeds

Active Learning

COVID-19

Commonsense QA

Definition Modeling

General Feed

Opaque Model Tuning

Question Generation

Virtual Assistants

[Auto] Based on my author profile

[Auto] Based on recently viewed

COVID-19

Recommended Research Rated Papers

Recommendations From The Past Month

Epidemiological and Clinical Characteristics of Patients With Coronavirus Disease-2019 in Shiyang City, China

Long Liu, Xu Lei, +11 authors Zhixin Liu · Frontiers in Cellular and Infection Microbiology

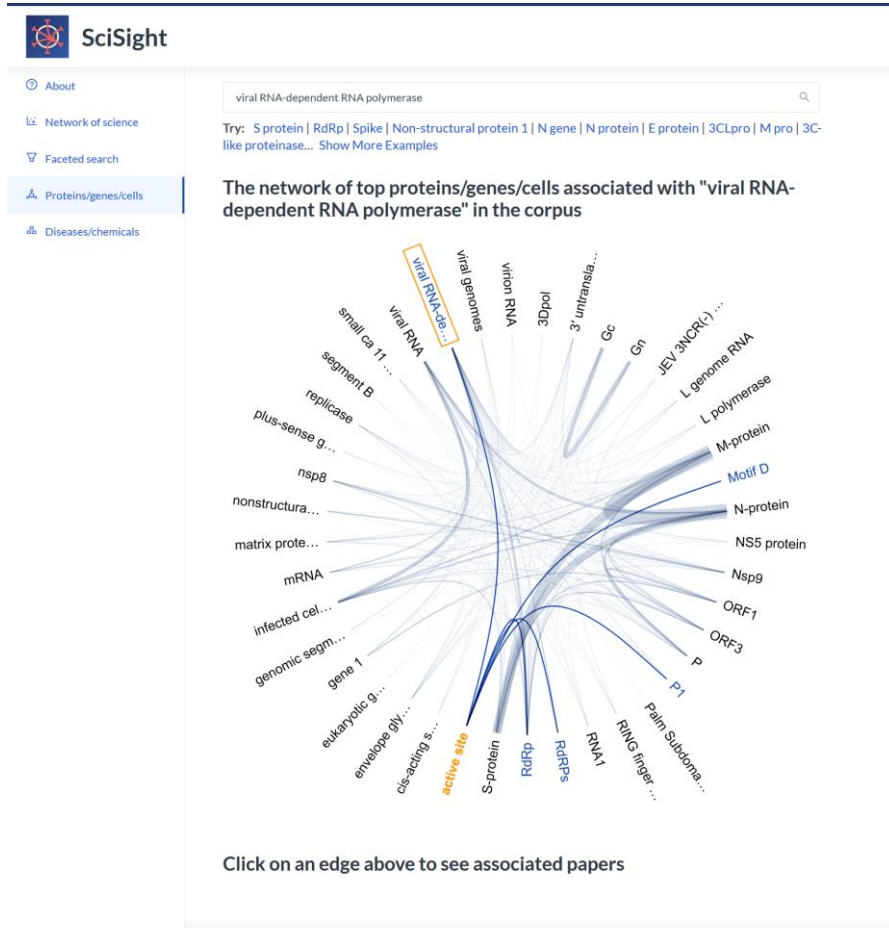
To investigate the early epidemic of COVID-19, a total of 176 confirmed cases in Shiyang City, China. The data indicated that the rate of emergence of early...

More Like This Less Like This Save

Rate papers to improve relevance, save papers to read later

- ✓ Stay on top of the latest research (updated daily)
- ✓ Feed adapts to your ratings (ensemble of neural, text-based models)
- ✓ Add papers to your feeds from anywhere on Semantic Scholar

Visualization of Connections Between Concepts in Papers



CORD-19 and the Scientific Community

Institutions using CORD-19



FRED HUTCH



UNIVERSITY OF
WATERLOO



U.S. National Library
of Medicine

Carnegie
Mellon
University



Google



Hewlett Packard
Labs



PennState

MAYO
CLINIC



NYU

PRIMER



KOREA
UNIVERSITY



JOHNS HOPKINS
UNIVERSITY



Universiteit Utrecht



SWEDISH
MEDICAL CENTER

NIST
National Institute of
Standards and Technology

UC San Diego

Connecting Computing and Medical Communities



“CORD-19 was so quickly ramped up and was invaluable in our early research. I am now using CORD-19 to find relevant data about the COVID-19 impact on endoscopy...CORD-19 is an excellent proof of concept for this rapidly developing new space.

- Sravanthi Parasa, Gastroenterologist and Clinical Researcher at Swedish Medical Center, Seattle

Sample Medical Articles Utilizing CORD-19

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES



Cold
Spring
Harbor
Laboratory

BMJ Yale

COVID-19 and Inflammatory Bowel Diseases: risk assessment, shared molecular pathways and therapeutic challenges

Tuberculosis and COVID-19 in 2020: lessons from the past viral outbreaks and possible future outcomes

Who is more susceptible to Covid-19 infection and mortality in the States?

arXiv.org

Identifying Radiological Findings Related to COVID-19 from Medical Literature

Yuxiao Liang
UC San Diego

YUL154@ENG.UCSD.EDU

Pengtao Xie
UC San Diego

PENGTAOXIE2008@GMAIL.COM

AJO-DO
American Journal of Orthodontics
& Dentofacial Orthopedics

Precautions and Recommendations for Orthodontic Settings During the COVID-19 Outbreak: A Review

PubMed.gov

Coronaviruses and People With Intellectual Disability: An Exploratory Data Analysis

Safe Management of Bodies of Deceased Persons With Suspected or Confirmed COVID-19: A Rapid Systematic Review

SCIENTIFIC DATA

COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms

Vision for AI-based Scientific Discovery

“What if a **cure for an intractable cancer** is hidden within the tedious reports on thousands of clinical studies? ...

AI-based discovery engines will help to find the answers to science's thorniest problems and ultimately revolutionize science.”

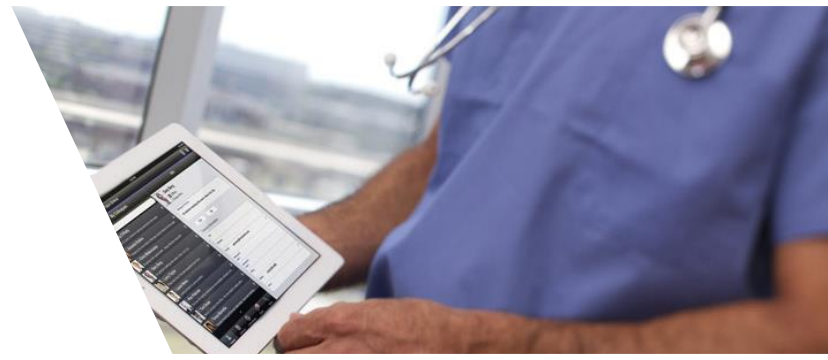
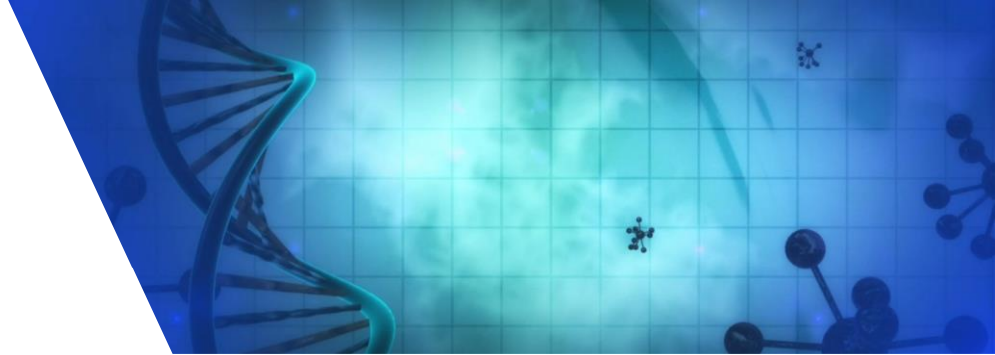
Wired Magazine, September, 2015



Semantic**Scholar**

“It's the **absence** of AI technologies that is **already** killing people.”

-Eric Horvitz



**Can we scale AI tools to support all
scientific inquiries?**