



AI for Health: Covid19 Learnings

Geralyn Miller
Senior Director
AI for Good Research Lab
Microsoft

Please Do Not Reproduce Without Permission





AI for Good

Empowering the people working to solve humanity's greatest issues

AI for
Accessibility

AI for
Earth

AI for
Humanitarian
Action

AI for
Cultural Heritage

AI for
Health

Microsoft AI for Health

A \$60 million, five-year philanthropic program created to empower those working to tackle some of the toughest challenges in health.



AI for Health

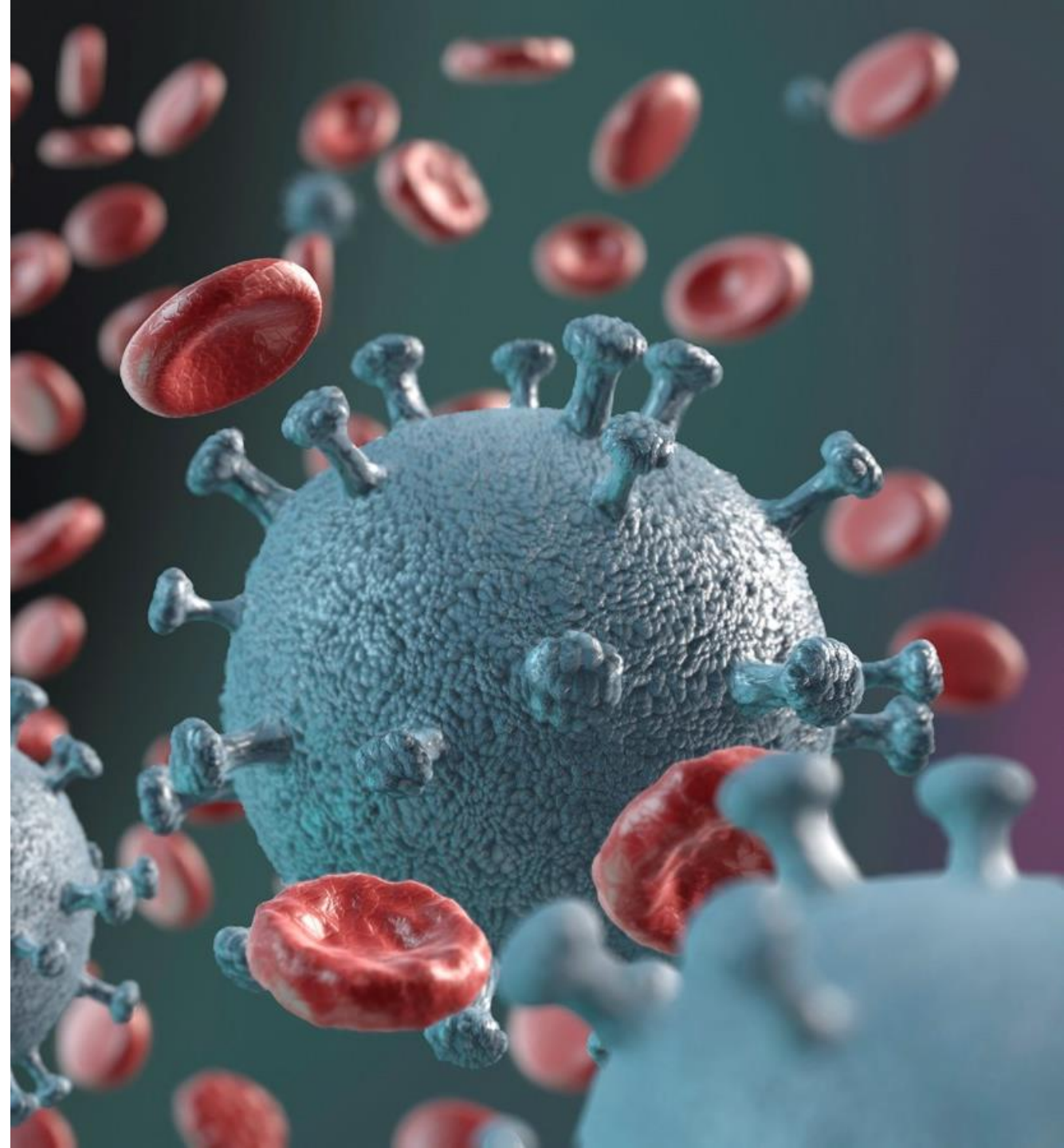
COVID-19 Response

Program

- \$20M investment in Covid19 research

Impact areas

- Data and insights
- Treatment and diagnostics
- Allocation of resources
- Dissemination of accurate information
- Basic scientific research



White House Covid19 HPC Consortium

Lessons in Collaboration

COVID-19 High Performance Computing Consortium



Advancing the pace of scientific discovery to stop the virus

- Private-public effort spearheaded by White House Office of Science and Technology Policy (OSTP)
- Provides researchers access to world's most powerful high-performance computing resources
- Microsoft is a founding member and provides grants through AI for Health



<https://covid19-hpc-consortium.org/>

Learnings:

- A shared experience lowers barriers
- Cross-sector, cross-discipline, cross-competitor collaboration IS possible
- Speed matters
- Operational agility is key
- Equilibrium is necessary in the value proposition
- There is an appetite for “open”
- Bias is real
- Attention to portfolio is necessary



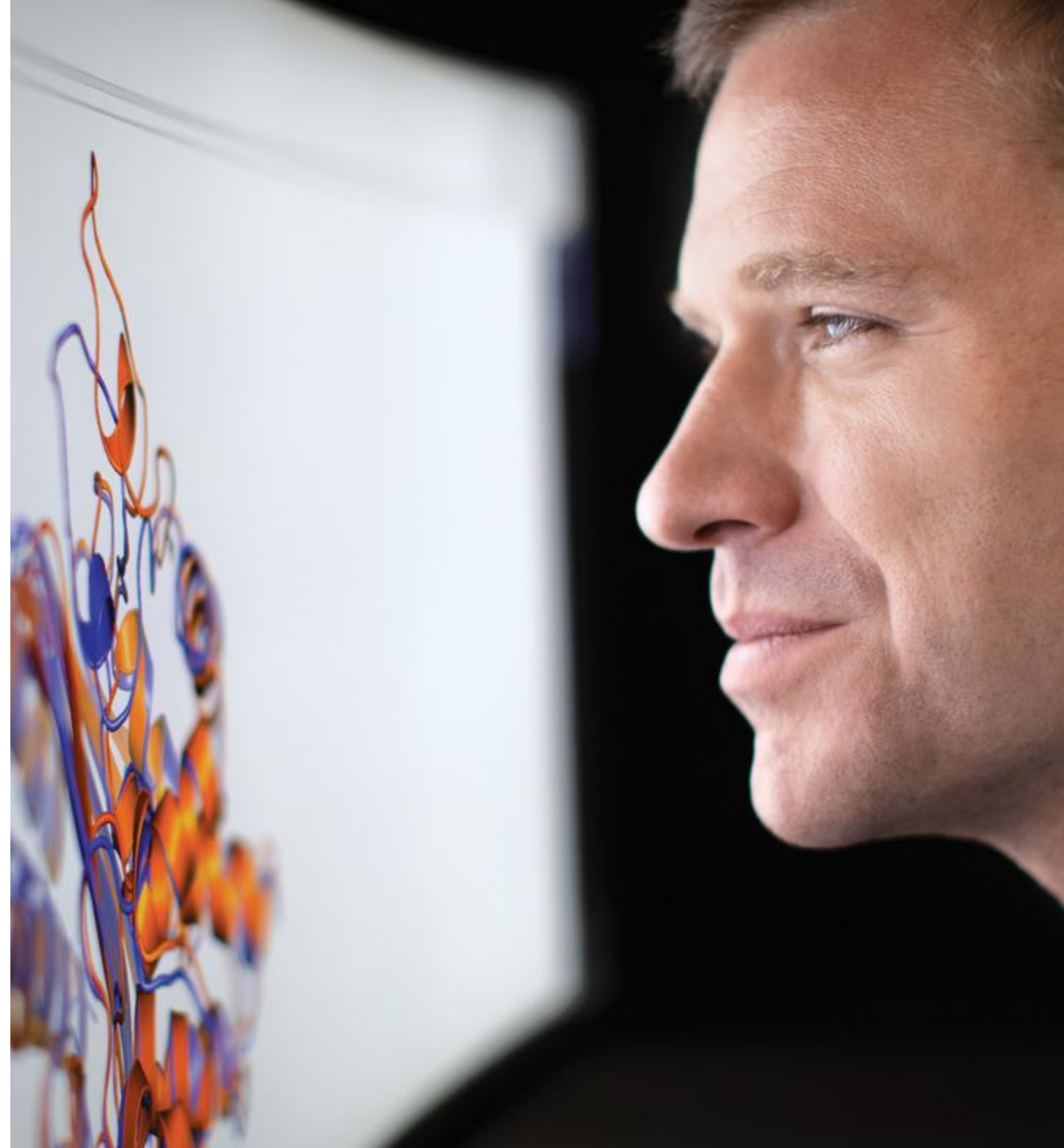
When days matter

Massive compute at your fingertips

Grantee: Folding@home

Harnessing a global computer to understand Covid19

- Distributed computing project at Washington University in St. Louis
- Organizing millions of volunteers to run COVID-19 protein simulations
- Accelerated by Azure to run simulations at speed and scale
- F@H is likely the most powerful supercomputer on the planet--a combined 470 PetaFLOPs (one quadrillion floating-point operations per second)



Grantee: Duke University

Addressing short supply of ventilators

- COVID-19 HPC Consortium + AI for Health grantee
- Two-person ventilator splitting is possible and can help with supply issues, but only for patients with similar characteristics
- Over a single weekend, conducted millions of simulations in Azure to determine optimal volumes and pressures expected for ventilator splitting between two patients with different characteristics
- Submitted for emergency FDA approval



Grantee: UCB

The COVID-19 Moonshot Project

Using AI and computational design to accelerate drug discovery

- A grassroots employee volunteer effort to accelerate research to combat COVID-19
- Identified 150 new molecules that could potentially counteract replication of the SARS-CoV-2 virus and aid in drug therapies
- Utilized thousands of GPU cores in Azure datacenters across the world to do in three days what might otherwise have taken six months



Data Sharing

The Good, the Bad, the Ugly

The Good News: Data sharing IS happening with Covid19

- Research findings are available on pre-print servers before journal publication
- In most cases, peer-reviewed research publications relevant to the outbreak are either open access, or freely available for the duration of the outbreak
- Many consortiums, coalitions, and competitions/challenges encourage the sharing of covid19 data. A few examples:
 - The National COVID Cohort Collaborative (N3C) Data Enclave
 - The Trinity Challenge
 - CORD-19: COVID19 Open Dataset and AI Research Challenge
 - CDC's Covid Data Tracker: Cases, Data, and Surveillance
 - Stanford AIMI: COVID19 + Imaging AI Resources
 - Nextstrain COVID19 Genetic Epidemiology
- Data interoperability standards such as HL7 FHIR are gaining ground

The Bad News: Data sharing is <still> hard

- No common licensing framework or common set of data use agreements
 - Data sharing agreements are costly and take time to negotiate
 - Common licenses for open-source code, but nothing similar for data
- Regulations necessary to protect data privacy can impact data sharing
 - Regulations vary across the world, e.g. GDPR, HIPAA, CCPA, no regs
 - It's early for privacy preserving technologies such as differential privacy and synthetic data
- Data sovereignty rules restrict movement of data
- Data models/ontologies vary between research/clinical and across subdomains
 - Time consuming to normalize data
- Difficult to track origin and provenance of data
 - Examples of dissemination of inaccurate information via social media abound in Covid19
 - Field of archaeology distinguishes between provenience and provenance

The Bad News: Data sharing is <still> hard

- Cultural barriers discourage data sharing
 - Traditional models for advancement in academia encourage researchers to keep data private
 - Fear of big tech
- Data hoarding
 - Perception that data itself is what is valuable vs the outcomes produced using the data
 - “Delete never” mentality--companies often do not know what data they possess nor how to assess what needs to be retained
- Non-existent business model around data
 - Creation/collection of original data is costly/time consuming
 - Who pays to curate and host all of this data?
 - Exponential storage growth, little innovation

The Ugly News: We have gaps.

**The most important learning about data from Covid19:
It's is not about what data we have, rather it is about what
data we DON'T have.**

Examples of data gaps:

- We are concerned with issues of racial disparity in the prevalence of covid19. Yet, many states in the US do NOT report race when they report on testing, hospitalizations, and deaths. It is difficult to assess the true nature and severity of a problem when the data is incomplete.
- Governments around the world have implemented different policies in an attempt to reduce the number of covid19 cases. These policies vary by country and within countries, by region. To understand how the different lockdown policies impact the spread of covid19, we need to know what policies are in place in which regions over a given period of time. This data is not collected today.

AI and Data: The Challenge

AI will transform life and business.

AI is fueled by data.

Access to AI technologies and data could shape the distribution of geopolitical power and economic wealth.

We need to democratize both AI and the access to data on which it relies to ascertain that ALL communities and ALL countries benefit.

Thank you.

Geralyn Miller (geramill@microsoft.com)

Please Do Not Reproduce Without Permission