

# Role of Artificial Intelligence (AI) and Machine Learning (ML) in NCHS Programs

Board of Scientific Counselors

09/14/2023

Travis Hoppe, PhD, [qrd5@cdc.gov](mailto:qrd5@cdc.gov)

Associate Director of Data Science and Analytics, Office of the Director

National Center for Health Statistics,  
Centers for Disease Control,  
Health and Human Services (HHS/CDC/NCHS)

The findings and conclusions in this report are those of the author and do not necessarily represent the official position of the National Center for Health Statistics, Centers for Disease Control and Prevention

# NCHS AI activities, per EO 13960 (Trustworthy AI)



## Developed in house

- [SANDS](#): Semi-automated non-response detection, live model release
- [MedCoder](#): Coding cause of death information to ICD-10
- Detecting Stimulant and Opioid Misuse and Illicit Use in EHRs

## Commercial-off-the-shelf or open-source implementation

- Whisper: Speech to text (CCQDER) improvement, 20x improvement
- Private AI: PII detection model used for FOIA requests

## Governance, policy, and strategic planning

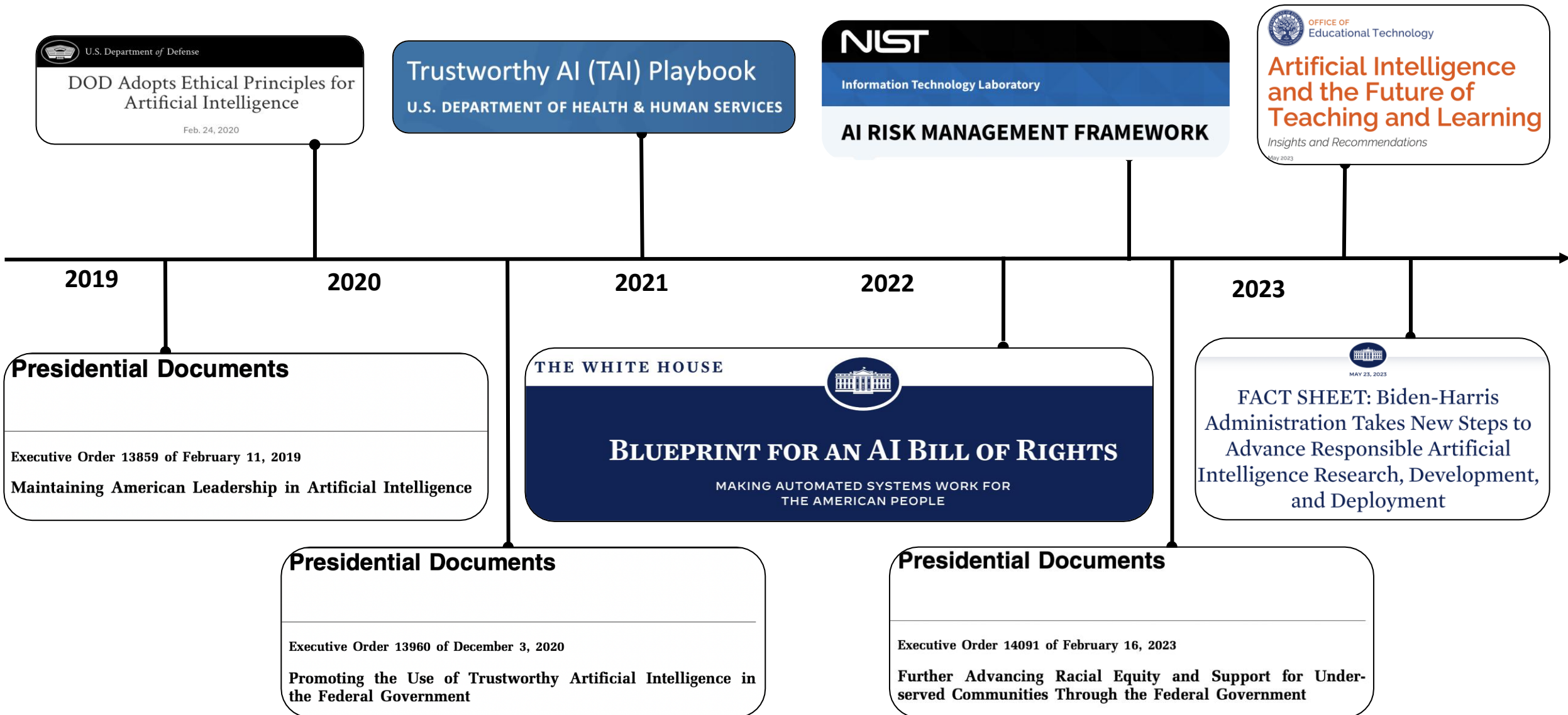
- Utility and Risks to CDC of Conversational Artificial Intelligence

# (Generative) AI at across the Federal government

## NCHS → CDC → HHS → USG

- NCHS led the development of a generative AI whitepaper
- NCHS is helping lead a new CDC AI Community of Practice
- OPHDST is developing a FedRamp approved ChatGPT interface
- CDC is drafting an AI strategy, including generative AI
- HHS published the [2023 AI inventory](#)
- Guiding Frameworks: NIST [AI RMF](#), HHS [Trustworthy AI Playbook](#), and upcoming [OMB guidance on generative AI](#)

# Trustworthy AI initiatives in the federal government



# Potential benefits of generative AI for CDC

Potential benefit to CDC	Examples
<b>Save resources:</b> Automate or accelerate certain tasks and processes	<ul style="list-style-type: none"><li>• Draft an executive summary of a written report from talking points</li><li>• Draft a chart abstraction questionnaire</li><li>• Draft a scope of work for a contract</li><li>• Draft a position description for a vacant job</li></ul>
<b>Increase efficiency:</b> Accelerate manual or repetitive tasks that do not require a high level of expertise or using structured formats	<ul style="list-style-type: none"><li>• Draft or edit scientific products including abstracts and manuscripts</li><li>• Create resumes or other structured documents</li><li>• Create or edit code and documentation</li><li>• Translate code between programming languages</li></ul>
<b>Synthesize and summarize information</b>	<ul style="list-style-type: none"><li>• Summarize written material</li><li>• Describe methodological options, with references, for conducting multiple imputation at scale</li></ul>
<b>Communicate:</b> Adapting text to various reading levels, platforms, audiences	<ul style="list-style-type: none"><li>• Draft communications messages</li><li>• Translate websites or other written materials between spoken languages</li><li>• Draft tweets from a list of talking points.</li></ul>
<b>Personalize responses</b>	<ul style="list-style-type: none"><li>• Draft responses to public inquiries with references to relevant CDC guidance</li></ul>

# Potential risks of generative AI for CDC

Potential harm to CDC	Examples
<b>Failure to meet ethical or regulatory standards</b>	<ul style="list-style-type: none"><li>Relevant regulations and policies that could be threatened by AI use include those addressing data accuracy, data security and privacy/confidentiality, intellectual property, health equity, discrimination, and scientific integrity.</li></ul>
<b>Fabrication</b>	<ul style="list-style-type: none"><li>Conversational AI may “hallucinate” or fabricate information including data and references has been reported widely. Fabrication, falsification, and plagiarism are included in the federal definition of reportable research misconduct.</li></ul>
<b>Bias and discrimination</b>	<ul style="list-style-type: none"><li>Conversational AI technology is trained on vast amounts of data from undisclosed sources, and tone and content may include biases, misconceptions, or outdated information regarding specific populations</li></ul>
<b>Lack of privacy and data security</b>	<ul style="list-style-type: none"><li>Cloud-based models rely on external infrastructure, which may not always be compliant with the strict security standards required in the healthcare sector or by federal assurances.</li></ul>
<b>Other legal or ethical risks, including risks to agency credibility</b>	<ul style="list-style-type: none"><li>Transparency, oversight, and accountability for AI-generated material can be unclear.</li></ul>

# CDC should anticipate malicious use by external actors

Potential action	Examples
<b>Generating misinformation</b>	<ul style="list-style-type: none"><li>• External actors might generate large volumes of text-based communications that include unintentional errors or propagate misinformation.</li><li>• Misinformation may use various strategies to appear credible, such as using scientific language or citing references.</li></ul>
<b>Impersonating CDC</b>	<ul style="list-style-type: none"><li>• External actors might use ChatGPT to imitate CDC or other public health authorities (“deepfakes”) or hallucinated written material falsely attributed to CDC.</li><li>• CDC may need to expend resources explaining or debunking information falsely attributed to CDC, including requests to verify statements, publications, or guidance.</li></ul>
<b>Spamming CDC</b>	<ul style="list-style-type: none"><li>• In platforms for public communication where CDC is required or obligated to respond, external actors might spam CDC with large volumes of credible inquiries.</li><li>• This includes public comments (including on government documents or at open government meetings), public inquiries to CDC (including CDC-Info and NIP-Info), social media comments and inquiries, and media or other requests for CDC response.</li><li>• Recently, volume of public comments at a CDC ACIP in late 2022 increased from an anticipated 600 comments to 127,000 comments (unknown proportion AI).</li></ul>

# State of AI: How we got here and current capabilities

Board of Scientific Counselors

09/14/2023

Ben Rogers, MS, [qtw4@cdc.gov](mailto:qtw4@cdc.gov)

Data Scientist, Division of Research and Methodology

National Center for Health Statistics,  
Centers for Disease Control,  
Health and Human Services (HHS/CDC/NCHS)

The findings and conclusions in this report are those of the author and do not necessarily represent the official position of the National Center for Health Statistics, Centers for Disease Control and Prevention



## Considerations for the BSC:

1. How can NCHS and other statistical agencies benefit from AI/ML?
2. What role should federal agencies take in the AI/ML research space given the rapid innovation?

## State of AI key take aways:

- Recent innovations in technology and data have spurred massive growth in the capabilities of AI
- Existing text/audio/images workflows can be augmented with AI
- Advances in AI/Machine Learning are not slowing down
  - Emergent capabilities in larger models not present in smaller models
- Open source models and data help AI adoption and innovation

# Using definitions of AI and Machine Learning from NIST:

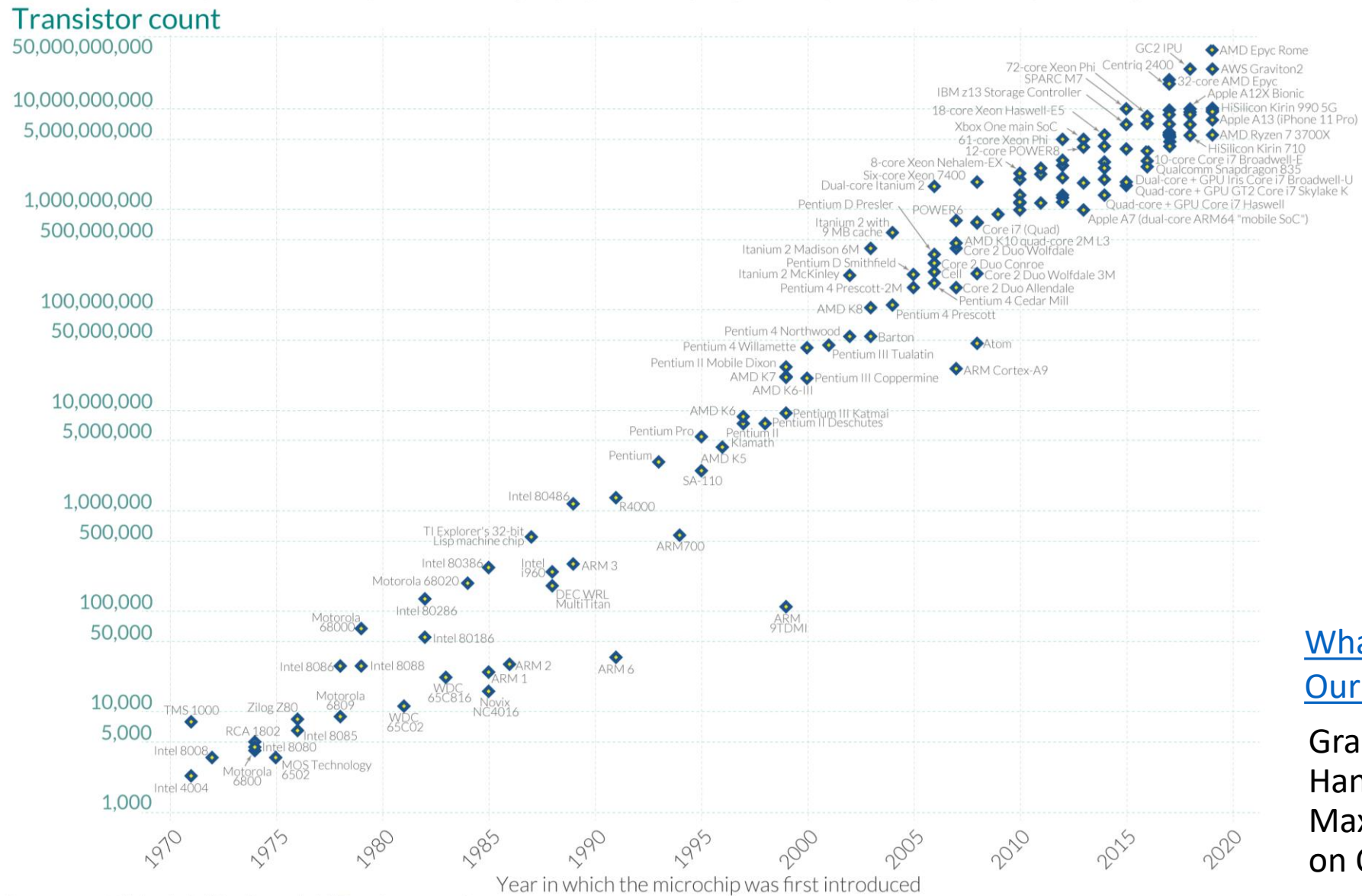
AI: Interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning. \*

Machine Learning: The study of computer algorithms that improve automatically through experience.\*

**AI, How did we get here?**



# It starts with Moore's Law



[What is Moore's Law? - Our World in Data](#)

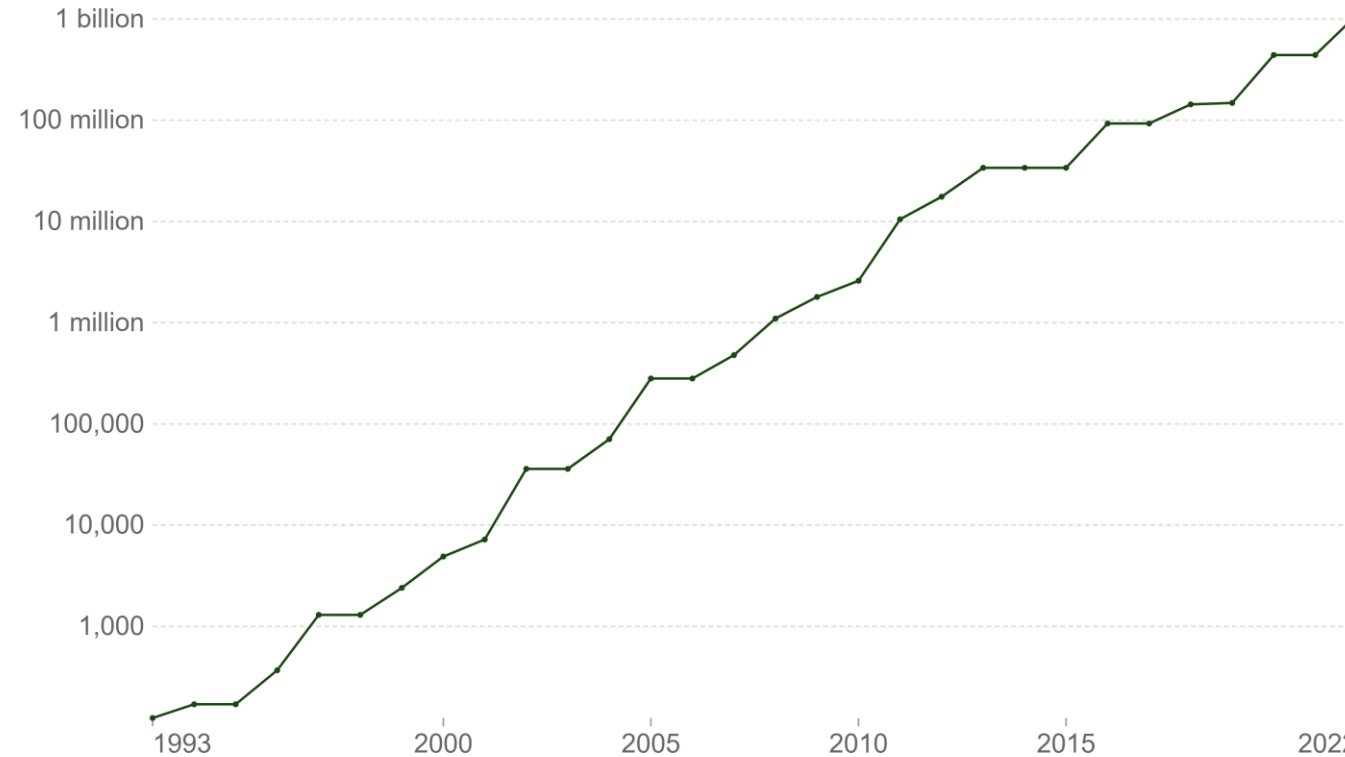
Graphic created by Hannah Ritchie and Max Roser available on Our World in Data.

# Implication: Exponential growth in computing power

## Computational capacity of the fastest supercomputers



The number of floating-point operations<sup>1</sup> carried out per second by the fastest supercomputer in any given year. This is expressed in gigaFLOPS, equivalent to  $10^9$  floating-point operations per second.



Source: TOP500 Supercomputer Database (2023)

OurWorldInData.org/technological-change • CC BY

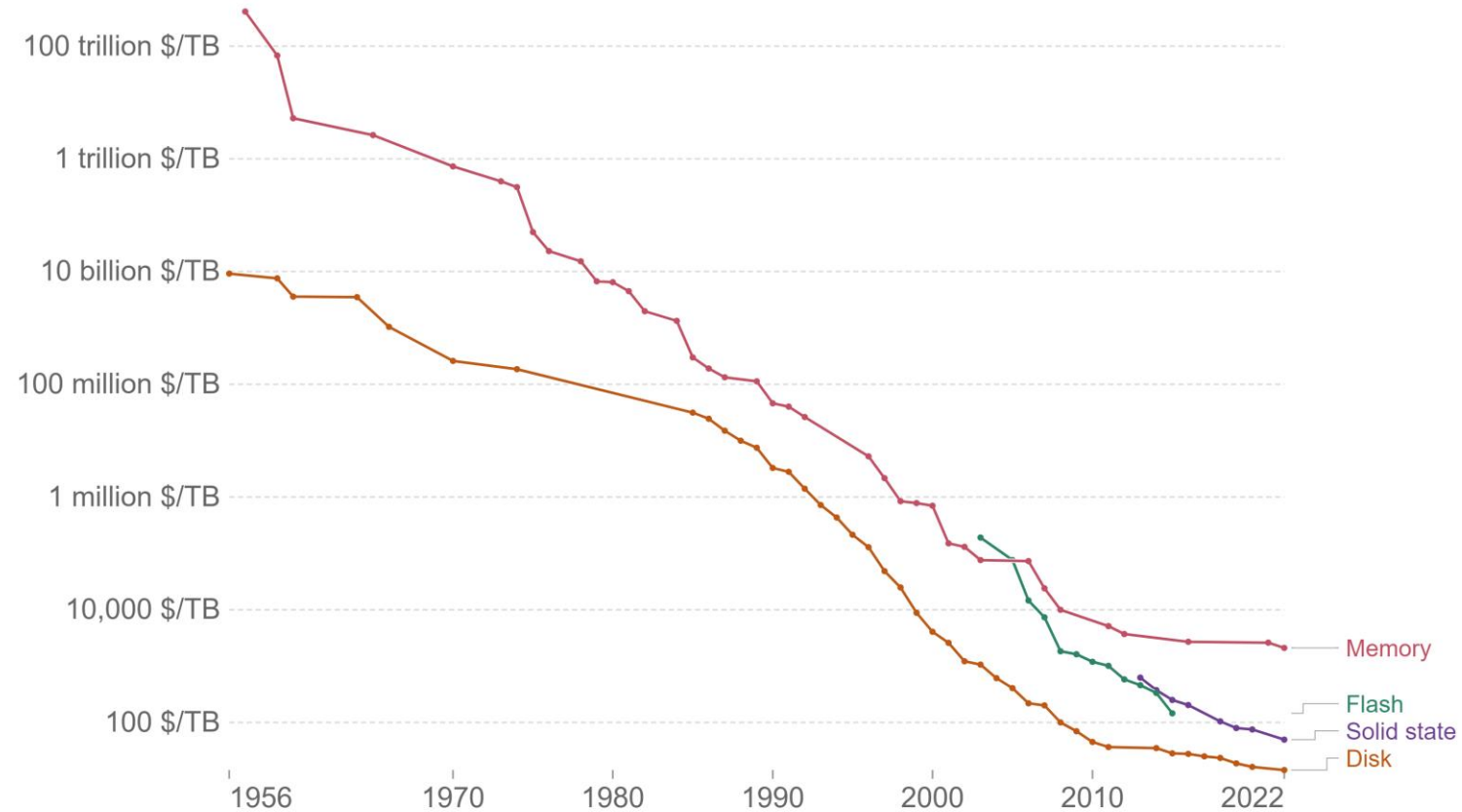
1. **Floating-point operation:** A floating-point operation (FLOP) is a type of computer operation. One FLOP is equivalent to one addition, subtraction, multiplication, or division of two decimal numbers.

[Computational capacity of the fastest supercomputers \(ourworldindata.org\)](https://ourworldindata.org/computational-capacity-of-the-fastest-supercomputers)

# Implication: Large availability of affordable data storage

## Historical cost of computer memory and storage

This data is expressed in US dollars per terabyte (TB). It is not adjusted for inflation.



Source: John C. McCallum (2022)

OurWorldInData.org/technological-change • CC BY

Note: For each year, the time series shows the cheapest historical price recorded until that year.

[Historical cost of computer memory and storage \(ourworldindata.org\)](https://ourworldindata.org/historical-cost-of-computer-memory-and-storage)

# Implication: Increased data availability

- Data are available through web-scraping, research article platforms, etc.
- This includes any text, image, audio, video in the public domain
  - Anything available on CDC.gov, NCHS's discontinued blog nchstats.com, etc.

## The websites in Google's C4 dataset

67 domains begin with "nchs"

RANK	DOMAIN	TOKENS	PERCENT OF ALL TOKENS
162,801	nchs-health.org	130k	0.00009%
209,009	nchstats.com	110k	0.00007%

Google's Colossal Clean Crawled Corpus introduced in <https://arxiv.org/abs/1910.10683v3>

<https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>



# Application: Foundational AI models

- Example: LLaMa 2 family of Large Language Models
  - Trained on Meta's Research SuperCluster
  - Estimated cost of computing with Microsoft Azure ~\$4.5 Million\*\*
  - Chat GPT and other famous Conversational AI models assumed to be on the same magnitude of computational cost
- Training foundational AI models at this scale requires thousands of data-center grade GPUs which are rare and in high demand internationally

[\\*\[2307.09288\] Llama 2: Open Foundation and Fine-Tuned Chat Models \(arxiv.org\)](#)

\*\*Using cost of \$1.3631 per hour of A100 GPU on Azure platform where LLaMa2 used 3,311,616 GPU hours

# Application: open source foundational models available for data scientists

- Foundational models are available to fine tune either through open-source (LLaMa2\*) or paid access (Chat GPT\*\*)
  - Vastly cheaper to fine tune these pre-trained models
  - OpenAI FedRAMP authorized through Microsoft Azure\*\*\*
- Zero-shot/Few-shot classification from transfer learning enable high-performance on classification tasks with off-the-shelf models

# Potential AI tasks considered by NCHS staff

- Text:
  - Translation, code generation, summarization, conversational text generation, classification, named entity recognition, PII detection, topic modeling
- Vision:
  - Classification, generation, segmentation
- Audio:
  - Transcription, translation, generation, speaker identification

# Current CDC/NCHS projects leveraging AI



# Use of machine learning for “nowcasting” suicide fatalities in the U.S.

- CDC’s National Center for Injury Prevention and Control has leveraged machine learning and artificial neural networks to integrate several streams of information to estimate weekly suicide fatalities in the U.S. in near real time.
- The ensemble machine learning framework reduces the error for suicide rate estimation, establishes a novel approach for tracking suicide fatalities in near real time, and provides the potential for an effective public health response and interventions to prevent suicides in the U.S.

[Development of a Machine Learning Model Using Multiple, Heterogeneous Data Sources to Estimate Weekly US Suicide Fatalities - PubMed \(nih.gov\)](#)

# HaMLET: Harnessing Machine Learning to Eliminate Tuberculosis

- At CDC, data scientists are exploring ways to use machine learning to leverage chest x-ray and tuberculosis data from overseas immigration visa screening examinations of immigrants and refugees to enhance programs to detect and treat tuberculosis (TB) to prevent TB importation.
- The computer vision model detects abnormal findings suggestive of TB in chest x-rays and can identify results that are discordant with the radiologist's findings



<https://github.com/scotthlee/hamlet>

# Transcribing cognitive interviews

- Transcriptions have had a 69.8% decrease in errors in 2 years
  - Previous CCQDER Transcription software has a word error rate of 68.4 on Cognitive interviews
  - \*2020 state-of-the-art transcription model Wav2Vec2-large-960h had a word error rate on FLEURS.en\_us is 14.6
  - \*2022 state-of-the-art transcription model Whisper word error rate on FLEURS.en\_us is 4.4

Transcription errors dropped from 1 in 2 words to 1 in 25 when applied to cognitive interviews conducted by NCHS staff

\* [Robust Speech Recognition via Large-Scale Weak Supervision](#)

# Semi-Automated Non-response Detection for Surveys

- SANDS is a non-response detection AI model designed to be used for data cleaning for open-ended survey text in conjunction with human reviewers.

Response	Gibberish	Uncertainty	Refusal	High-risk	Valid
sdfsdfa	0.998	0.000	0.000	0.000	0.000
idkkkkk	0.002	0.995	0.001	0.001	0.001
Because you asked	0.001	0.001	0.976	0.006	0.014
I am a cucumber	0.001	0.001	0.002	0.797	0.178
My job went remote and I needed to take care of my kids	0.000	0.000	0.000	0.000	1.000

*Outputs of various text inputs to SANDS model*