The Good, Bad, and Ugly of Real-World AI Use:

Practical applications of transformative impact of AI in healthcare, informatics, and the research community

September 18, 2024
1:00 – 2:00 PM

RTI INTERNATIONAL

Who We Are

RTI at a Glance

Worldwide Presence and Financial Strength

$1.2B
FY2023 Revenue

1,126
FY2023 Clients

3,832
FY2023 Projects

8
U.S. Offices

Research Triangle Park, NC
Ann Arbor, MI
Atlanta, GA
Berkeley, CA
Chicago, IL
Fort Collins, CO
Waltham, MA
Washington, DC

9
International Offices

Barcelona, Spain
Belfast, Northern Ireland
Jakarta, Indonesia
Ljungskile, Sweden
Manchester, United Kingdom
Lyon, France
Nairobi, Kenya
New Delhi, India
San Salvador, El Salvador

RTI Teams and Talent

- Analytics
- Communication
- Data
- Education
- Environment
- Health
- Justice
- Technology
- Transformative Research Unit for Equity

Over 550
IT specialists and data scientists

More than 1,600
staff holding advanced degrees in clinical, public health, and biomedical informatics.

Researchers, subject matter experts and technologists **collaborate on projects**

Working in many scientific domains, including public health, informatics, and data science

Center for Informatics at RTI International

Comprised of ~50 staff in three programs:

Bioinformatics: Leverages computational tools to **capture, curate, and analyze genetic, epidemiologic, and clinical data**, turning raw information into actionable insights for human health.

Environmental Informatics: Uses data and computational approaches to **transform environmental information into evidence-based** solutions and policies that promote human and environmental health

Health Informatics: Lead and support projects across a wide variety of health-related areas including **clinical and translational research, healthcare quality, and applied public health**.

Work with a wide variety of clients on large (**>$100M**) and small projects (**~$100K**)

Potential Uses for LLMs in Care Delivery and Health Research

Enhancing Dissemination

PDF Data Sources

Drug Discovery

Summarizing Clinical Notes

Medical Literature Reviews

Decision Support

Informatics, Decision Support and Artificial Intelligence

AI has the potential to impact many areas of informatics, including:

- shared decision making (SDM)
- shared care coordination and planning
- decision support (DS) solutions
- clinical decision support (CDS)
- patient-centered CDS (PC CDS)

Key areas of CDS impact include:

- **Automating processing**
- **Improving sharing and replication**
- **Accelerating technical development**
- **Supporting patient decision making**
- **Enhancing cognitive processing**
- **Increasing patient-provider interaction**

https://cdsic.ahrq.gov/cdsic/implementation-case-studies

Webinar Speakers

Laura Marcial, PhD, FAMIA (she/her): Senior Director, Center of Informatics, RTI International, lmarcial@rti.org

Jamie Pina, PhD, MSPH (he/him): Senior Director, Center for Data Modernization Solutions, RTI International, RTI International, jpina@rti.org

Daniel Brannock, M.S. (he/him): Research data scientist, RTI International, mbrannock@rti.org

Emily Hadley, M.S. (she/her): Research data scientist, RTI International, ehadley@rti.org

Background

Types of Artificial Intelligence

Narrow AI also known as “weak AI,” focuses on performing specific tasks within a limited domain

Generative AI (genAI) refers to artificial intelligence systems that can create new, original content, such as text, images, or music, by learning patterns and structures from existing data.

General AI refers to highly autonomous systems that feel like they possess human-level intelligence and can handle various cognitive tasks across different domains. Still considered theoretical.

Significant Events in Artificial Intelligence

- **1950** — Turing Test
- **2011** — IBM Watson
- **2017** — “Attention is All You Need" paper by Vaswani et al. in 2017
- **2020** — OpenAI releases GPT-3
- **2023** — Proliferation of generative AI
- **2024+** — Direct Industry Application

Transformer Attention Mechanisms

Transformer attention mechanisms enable models to dynamically focus on different parts of the input data, improving understanding and processing of natural language

- **Contextual Understanding**
- **Parallel Processing**
- **Scalability**
- **Flexible**

arXiv:1706.03762v5 [cs.CL] 6 Dec 2017

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

1 Introduction

Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

AI Readiness in AMIA's Scope

Advanced Data Analytics

AI-driven algorithms are increasingly capable of analyzing vast amounts of healthcare data, leading to more accurate diagnostics and personalized treatment plans.

Regulatory and Ethical Frameworks

Ongoing development of guidelines and standards ensures safe, ethical, and responsible AI integration in healthcare practices.

Interoperability and Infrastructure

Improved healthcare IT infrastructure is facilitating seamless integration of AI tools, enhancing clinical decision-making and operational efficiency.

Objectives

Understand the Responsible Use of AI in Healthcare through Real-world AI Applications

Explore Applications of AI in Large-scale EHR Data Repositories

Discover the Capabilities of Large Language Models (LLMs) for Novel Approaches to Text Analysis

Applying AI in Large-scale EHR Data Repositories

Daniel Brannock

Massive (EHR) Databases

The Good

- Huge data enabling powerful AI
- Power for studying underserved groups (e.g., rare diseases)
- Shared costs and infrastructure

The Bad

- Esoteric data structures
- Steep learning curve
- Missing data abounds; positive, unlabeled data

The Ugly

- Bias driven by access and utilization
- Observational: causal analysis is challenging

The Data, They Proliferate

NIH *All of Us*

- 400k people
- EHRs, genetics, questionnaires

NCATS National COVID Cohort Collaborative (N3C)

- 23M people, 9M with confirmed COVID
- EHRs

UK Biobank

- 500K people
- EHRs, genetics, imaging, questionnaires

Managed Platforms

Shared access facilitates collaboration

Managed distributed cloud compute

Common Data Models

RECOVER—N3C

Purpose

- Study Long COVID

Platform Details

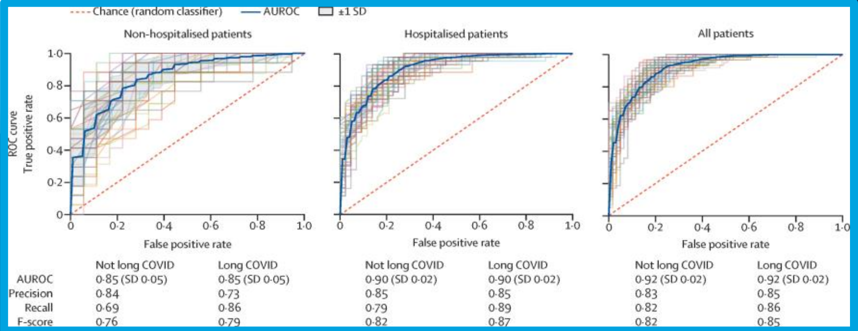
- Palantir Foundry, excellent computational support
- Python, R, Spark SQL
- Assumes knowledge of OMOP data structures

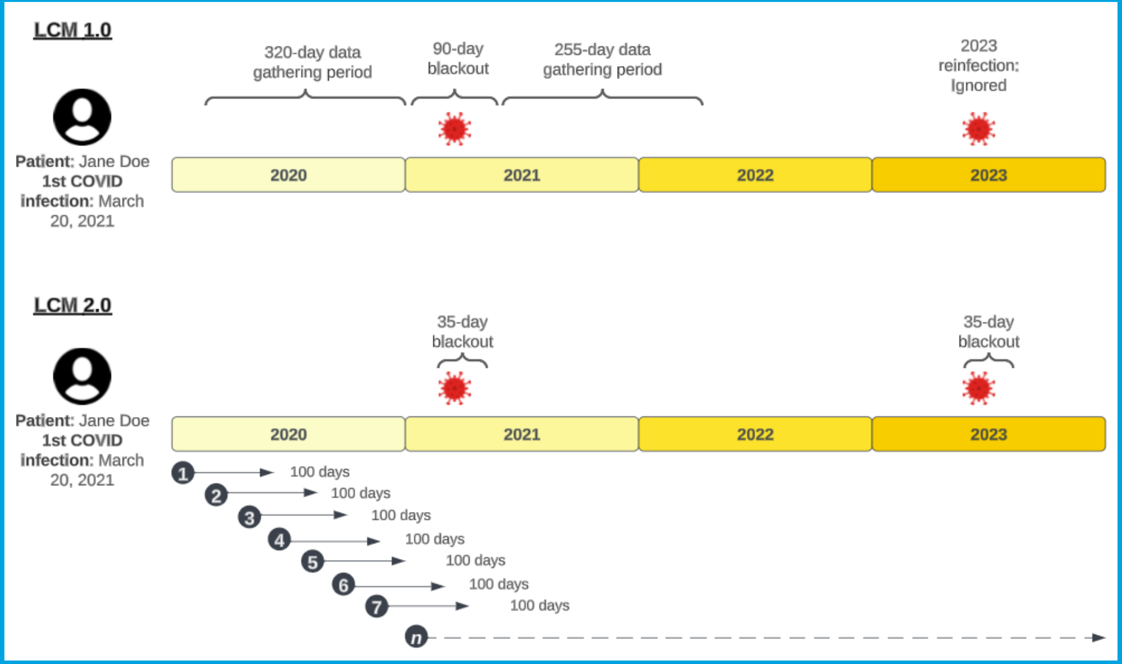
Multi-Disciplinary Team

- Bioinformaticians
- Clinicians
- Patients

RECOVER—Characterizing Long COVID

- Identifying **long COVID** and its onset
- **XGBoost model** using clinically diagnosed cases
- **Monitoring and retraining** are critical even in research





RECOVER—Other Example Projects

Does Paxlovid prevent hospitalization and/or long COVID?

Target trial emulation

Does vaccination prevent long COVID?

CEM, IPTW

Characterizing reinfections and effect on long COVID

Building the right cohort: who is the best control?

Table 3 Odds of hospitalization in Paxlovid-treated vs. Non-Paxlovid-treated patients

	Dependent variable: Hospitalization		
	Unadjusted (1) OR (95% CI)	Adjusted (2) OR (95% CI)	Vaccine-adjusted (3) OR (95% CI)
Paxlovid Treatment			
No (control)	*ref*	*ref*	*ref*
Yes (treatment)	0.33 (0.24-0.45)***	0.35 (0.29-0.42)***	0.32 (0.24-0.42)***
Vaccination Status[1]			
Unvaccinated	-	-	*ref*
Vaccinated	-	-	0.49 (0.41-0.58)***
Vaccination∗Paxlovid[2]	-	-	1.08 (0.71-1.64)
Observations	410,642	410,642	136,815
Log Likelihood	-31590.1	-25466.1	-6551.1
Akaike Inf. Crit.	63184.0	51000.2	13172.3

[1]Unvaccinated refers to patients who received 0 doses at index, Vaccinated refers to patients who recieved atleast 2 doses at least 14 days prior to index.

[2]Interaction term

Note: Model (2) adjusts for sex, age, race and ethnicity, Charlson Comorbidity Index, Community-well being index, data partners, and month of COVID-19 Index date. Model (3) additionally adjusts for the main and interaction of Vaccination.

$p < .05^{*}$, $p < .01^{**}$, $p < .001^{***}$

Analysis	**Treatment Effect (95% CI)**
N3C PASC Primary Outcome[a]	0.985 (0.956-1.014)

RECOVER—Other Example Projects

Does Paxlovid prevent hospitalization and/or long COVID?

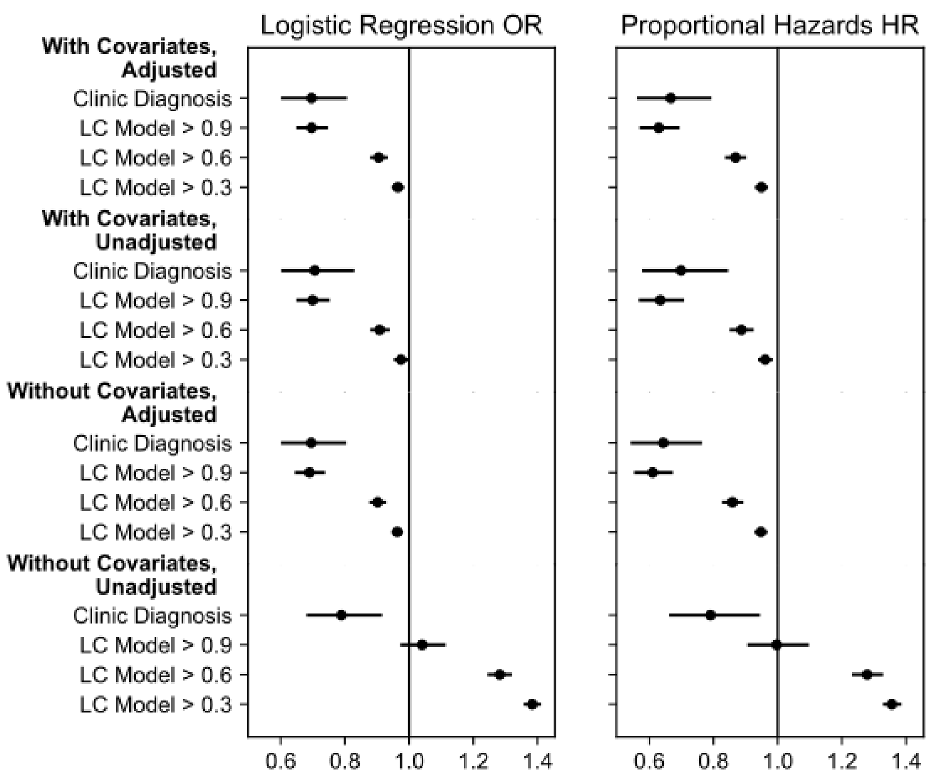
Target trial emulation

Does vaccination prevent long COVID?

CEM, IPTW

Characterizing reinfections and effect on long COVID

Building the right cohort: who is the best control?



AIM AHEAD—*All of Us*

Purpose

- Advance equity and researcher diversity in AI/ML research

Platform Details

- All of Us Researcher Workbench, built on Google Cloud
- Python (Jupyter Notebooks), R (RStudio), SAS
- Good exploratory GUI for beginners, still requires OMOP expertise to conduct robust research

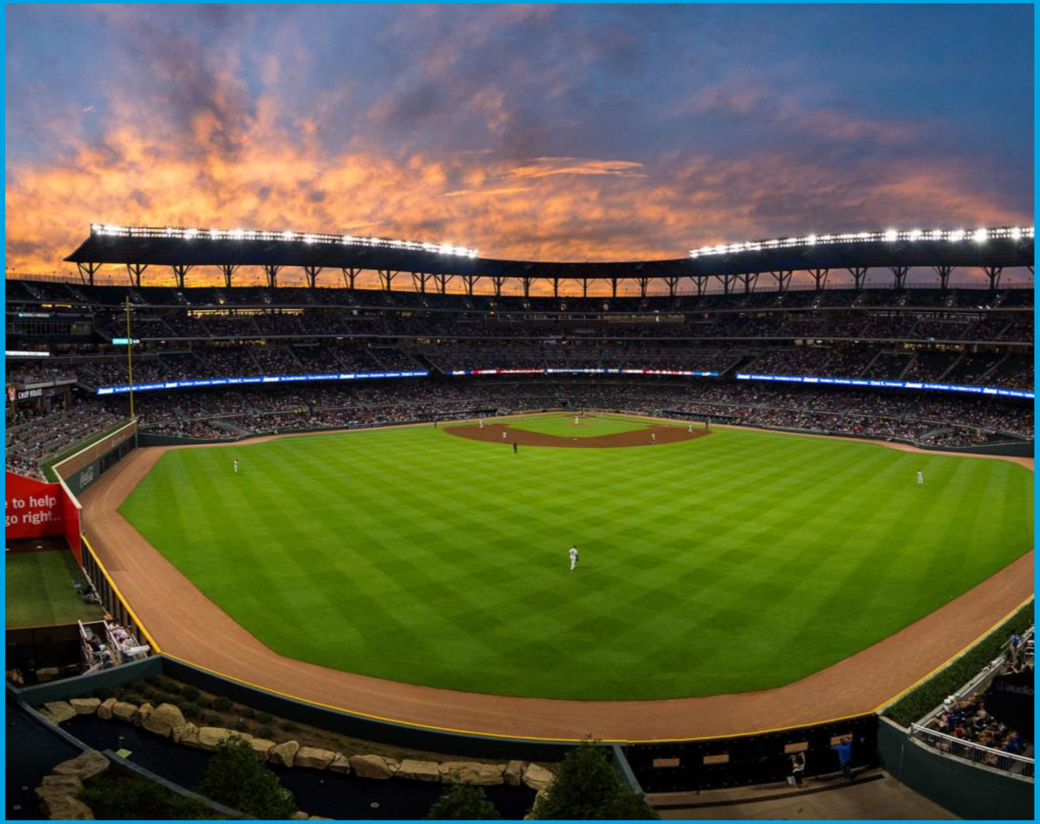
Multi-Disciplinary Team

- Network of mentors, mentees, training material

If We Build It, Please Come

Government has invested in these repositories

- NIH Strategic Plan for Data Science from 2018 calls out ***All of Us*** and the **Cancer Moonshot Initiative**



CONFIDENTIAL

Lessons Learned

Data and methodology are progressing quickly

Multidisciplinary collaboration is key

There are still profound limitations

- Observational
- Nonrandom missingness
- Positive and unlabeled data
- Biased populations

Applying NLP and LLMs in health research

Emily Hadley

Large Language Models

Gemini

ChatGPT

Claude

- Advanced computer program
- Trained on large amounts of text data
- Can understand and generate human-like language

Large Language Models

The Good

- Large-scale text generation and summarization
- Versatile applications for innovation and accessibility
- User-friendly

The Bad

- Substantial financial and environmental costs
- Highly dependent on quality of training data
- Can reflect and amplify biases in data

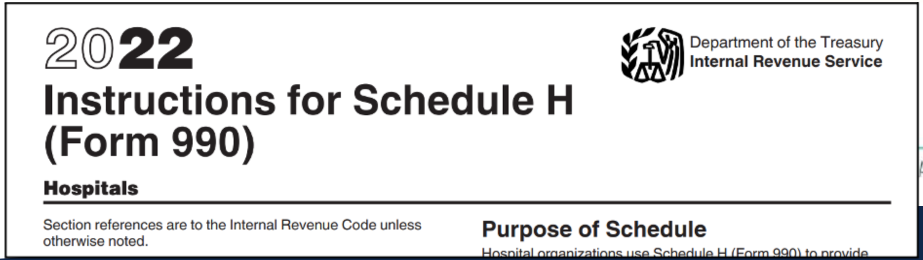
The Ugly

- Generate convincing misinformation, hallucinations, and false information
- Major concerns about authorship and data privacy

Why Use an LLM Approach?

- ☑ Research questions **not well addressed** with existing methods
- ☑ Publicly available data
- ☑ **No user interaction** or **decision**
- ☑ **Low cost** and **low risk**

Case Study: Community Benefits IRS Documents



2022

Instructions for Schedule H (Form 990)

Department of the Treasury
Internal Revenue Service

Hospitals

Section references are to the Internal Revenue Code unless otherwise noted.

Purpose of Schedule

Hospital organizations use Schedule H (Form 990) to provide

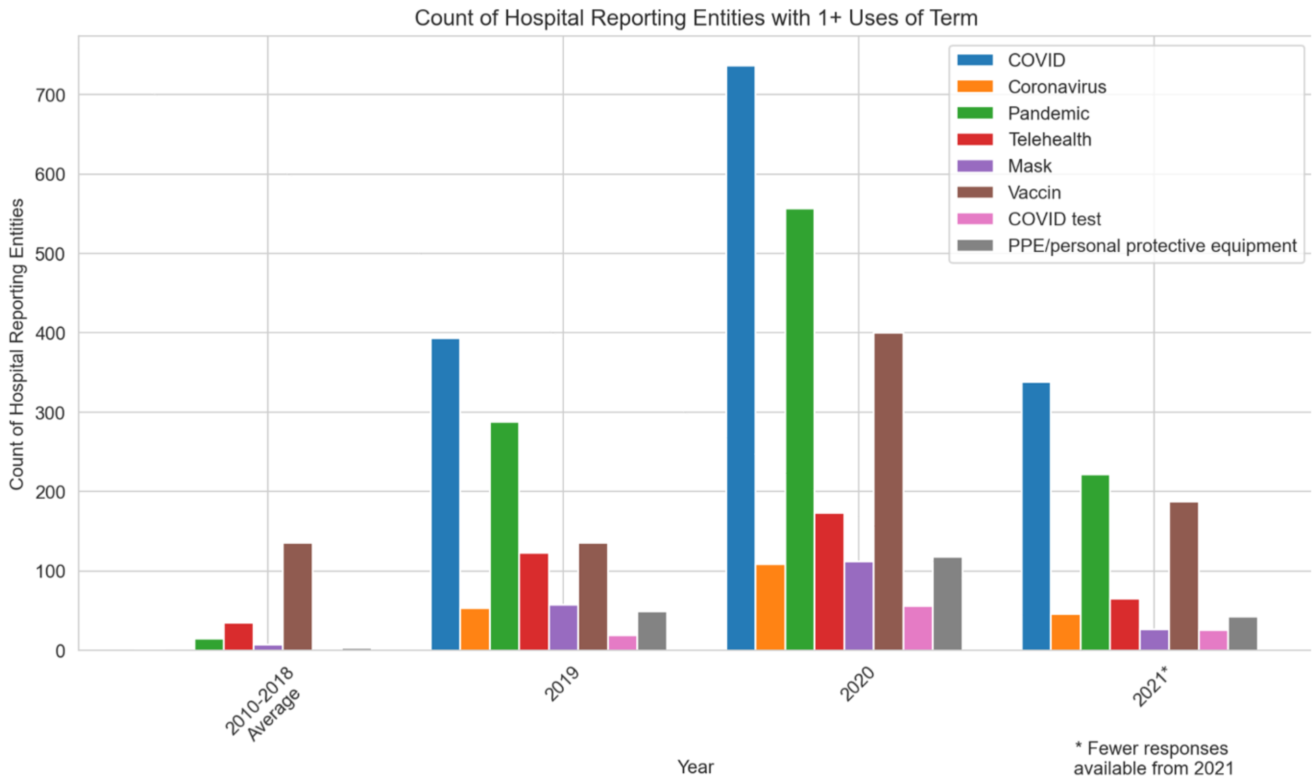
https://www.communitybenefitinsight.org

Narrative Section to Describe Community Benefits

PART II, COMMUNITY BUILDING ACTIVITIES:

health services' coalition building promotes the health of the communities it serves by networking with other community agencies to address the health and safety issues of the community. health services participates in the following state and local coalitions:(1) safe kids coalitions to promote awareness and use of child seat belts and bicycle safety;(2) statewide immunization collaborative;(3) alliance for the social determinants of health; (4) opioid community collaborative;(5) living well with chronic conditions statewide program with the utah department of health; (6) diabetes-related coalitions to help reduce the incidence of diabetes in children and adults;(7) multiple mental health collaborations and suicide prevention efforts; and(8) other coalitions that address healthcare issues in the community.two health services' hospitals provide space and maintenance for community gardens made available to community members to provide access to fresh, healthy food. health services' employees utilize their clinical expertise to collaborate with other community agencies and county and state health departments to provide education and other initiatives. health services also recruits physicians and mid-level providers to medically underserved areas to meet healthcare needs of residents, thereby helping reduce barriers to accessing care.

Traditional Keyword Frequency



Prompt: "Based on These Sentences, Make a List of Changes During the COVID-19 Pandemic"

Selected Topics	Selected Examples
Hospital Measures	Changes in visitor policies, creation of negative pressure rooms, renting additional hospital beds, establishment of fever clinics, vaccination requirements
Community Assistance	Food pantry assistance and food delivery services, holiday toy distribution, COVID-19 hotline, drive-through drug takeback events, distribute COVID-19 info in newspapers
COVID-19 Resources	Distribution of vaccines, masks, hand sanitizer, soap, face shields, and tests

Policy Approaches to Managing AI Risks

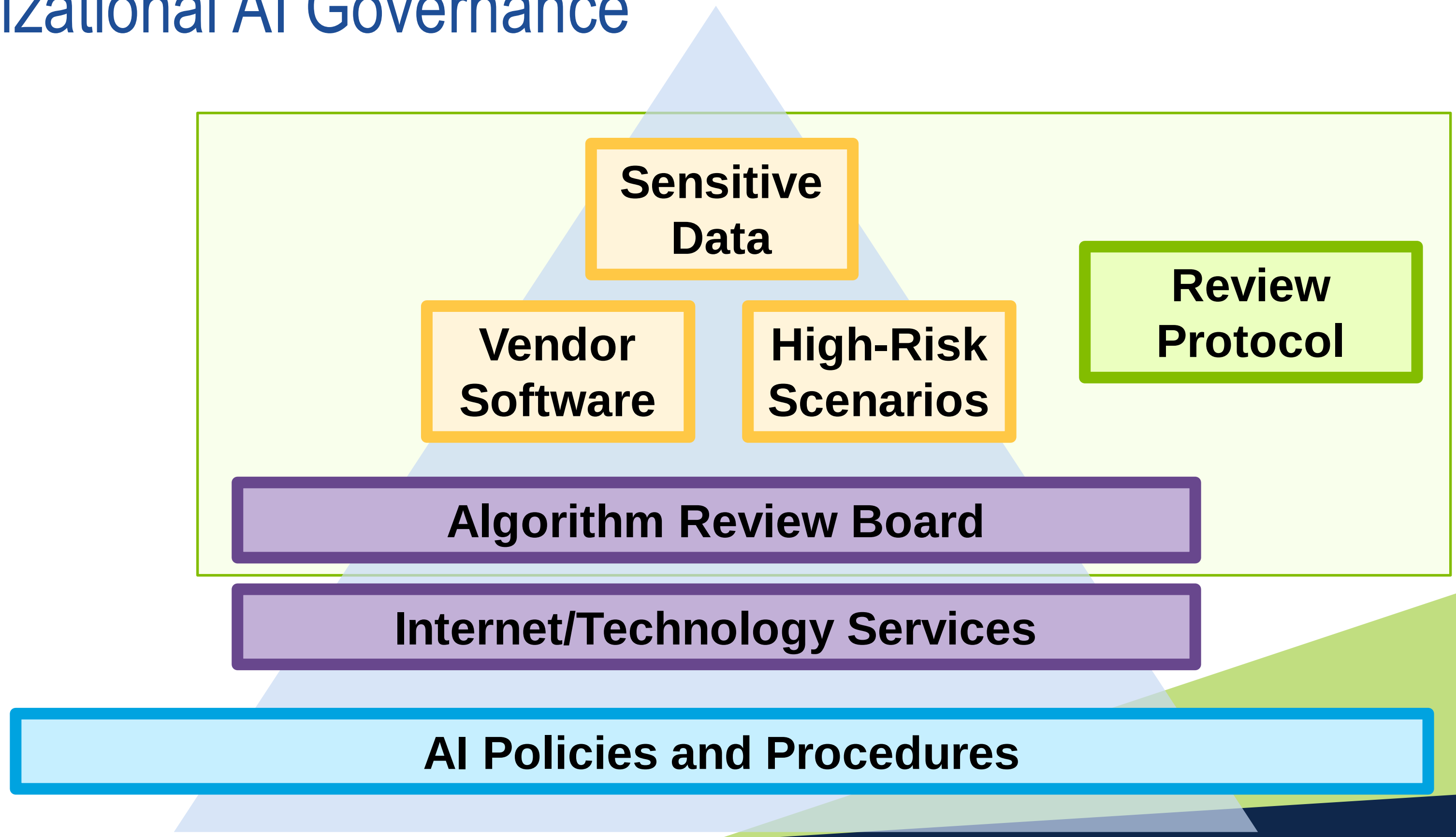
TABLE II. PRIORITIZATION RECOMMENDATIONS

Policy	U.S. Legislature	U.S. Government Agencies	U.S. States	Professional Societies	Organization
Licensure or Certification of AI Developers	⃠	✓	⃠	✓+	✓-
AI Ethics Statement	✓-	✓-	⃠	✓	✓-
Pre-Deployment Audits or Assessments	⃠	✓+	⃠	✓	✓+
Post-Deployment Accountability	✓+	✓	✓	⃠	✓
Database of AI Technologies or Incidents	✓	✓	⃠	⃠	✓-
Involvement of Community Stakeholders	⃠	✓	⃠	⃠	✓+
Policies That Support Responsible AI Education	✓	⃠	✓+	✓	✓
Policies That Support Responsible AI Research	✓	✓+	⃠	✓+	✓-
Policies That Support Diversity in AI Development	⃠	✓	⃠	✓	✓+

⃠ = No investment ✓- = Low Priority
✓ = Priority ✓+ = High Priority

Hadley, E. C. (2022). *Prioritizing policies for furthering responsible artificial intelligence in the United States*. Paper presented at 2022 IEEE Big Data Conference, Osaka, Japan. https://doi.org/10.1109/BigData55660.2022.10020551

Organizational AI Governance



Hadley, E., Blatecky, A., and Comfort, M. (2024). *Investigating Algorithm Review Boards for Organizational Responsible Artificial Intelligence Governance*. Pending review. Preprint available https://arxiv.org/abs/2402.01691

RTI Has Lots of Experience!

- Phenotype development (long COVID)
- Data harmonization (BDC)
- Knowledge graphs with NLP (HEAL)
- EHR deep learning
- ML CDS models (HIV)
- Taxonomy generation (with LLMs)
- GPT for concept classification
- NER to deidentify
- Text string auto-matching & auto-coding
- Fuzzy matches for search
- Similar concept search for metadata

National Clinical Cohort Collaborative (N3C)

RECOVER
Researching COVID to Enhance Recovery

NIH HEAL INITIATIVE

NIH National Heart, Lung, and Blood Institute | BioData CATALYST ®

Q&A

Jamie Pina

Resources

See More in our Newly Published Paper!

Journal of Data Science

Search within journ

Submit your article

Information

Article info

Related articles

Traditional and GenAI Text Analysis of COVID-19 Pandemic Trends in Hospital Community Benefits IRS Documentation

Emily Hadley Laura Marcial Wes Quattrone All authors (4)

https://doi.org/10.6339/24-JDS1144

Pub. online: 23 July 2024 **Type:** Data Science In Action Open Access

Responsible AI Resources

THE DIRECTOR

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

March 28, 2024

M-24-10

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Shalanda D. Young

SUBJECT: Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence