Collaborative AI Research and Development in Government

Dr. Gil Alterovitz Director of Artificial Intelligence



U.S. Department of Veterans Affairs

leterans Health Administration Office of Research & Development

Sept. 26, 2019



DISCOVERY \star INNOVATION 🛧 ADVANCEMENT

VA/Deep Mind Collaboration Published in Nature

- Predict Acute Kidney Injury (AKI)
 48 hours in advance
- Early warning enables time to take preemptive action
- Helps prevent kidney injury



Letter Published: 31 July 2019

A clinically applicable approach to continuous prediction of future acute kidney injury



Incentivizing a Bidirectional Data Link







DISCOVERY **★** INNOVATION **★** ADVANCEMENT

We can make "AI-able Data"

- We can design data formats to get the most out of AI.
- We can create an ecosystem of datasets, platforms, and tools to enable standardized AI piloting, deployment, testing.
- We can leverage iterative release of datasets for training AI models, validation, and testing
- Develop honest broker with testing platform and/or standardized agreements for testing protocols.





Al-able Data Ecosystem Piloted in Health Tech Sprint and Al Tech Sprint

- Artificial Intelligence (AI) for Novel Therapeutics Accessibility (NTA)
- An AI ecosystem around three datasets built for Health Tech Sprint around clinical trials in structured, machine readable information
- Resources/personnel across dozen agencies
- Participation included federal and other companies/organizations:
 Flatiron Health, Microsoft Healthcare, Oracle, Philips (Research),
 Rush Medical, and TrialX (in collaboration with the Global Lyme Alliance).
- AI Tech Sprint (led by VA)- piloting standardized data use agreement, AI-able data ecosystem and recognition.





AI-able Data Ecosystem Framework

- The Opportunity Project (TOP)
 - (1) Use existing federal open data to make new digital tools
- AI-Able Data Ecosystem Framework
 - Data needed for AI is often new, voluminous, and "messy."
 - (2) New approach: Create novel datasets/data policies specifically for the tech sprint and iterate on data/tools to make it usable for AI.
 - An AI ecosystem enables a bidirectional data and results sharing between government and industry
 - (3) New approach: Create voluntary framework for incentivizing federal agencies to release data that is useful and usable by other sectors, like industry
 - (4) New approach: Create voluntary framework for incentivizing industry to make AI/data results usable by others, including federal agencies, so that they leverage industry tools.





DISCOVERY ★ INNOVATION ★ ADVANCEMENT

Draft

* Or other other entities

Data's Choice and Al's Choice as a Catalyst for Incentivizing a Bidirectional Data Link





DISCOVERY **★** INNOVATION **★** ADVANCEMENT

Data's Choice

Diamond Level: Al-centric data sets and tools Gold Level: -----Stakeholder feedback / Iterative data release Silver Level: Documentation: Provenance / Data quality metrics **Bronze Level:** Machine Readable

Al-able data- data sets useful for training/testing Al models in industry

Data useful for specific industry use cases

Data is de-risked for industry to evaluate overall quality

Data can be analyzed in more efficient manner by industry



DISCOVERY ★ INNOVATION ★ ADVANCEMENT

AI's Choice

Diamond Level: Giving Back to Al Ecosystem/ Community

Promotes AI ecosystem and reuse of usable solutions

Gold Level: Independent Use/Validation Demonstrates usability of AI solution by other parties

Silver Level: Al Ecosystem Test Data Analysis

> Bronze Level: Al Ecosystem Application

Promotes trust of AI solution through testing

Leveraging open data,

•---• promoting transparency and reproducibility.



DISCOVERY \star INNOVATION 🛧 ADVANCEMENT

Partnerships: Blind Date and Marriage?





No prior knowledge

High expectations



DISCOVERY **★** INNOVATION **★** ADVANCEMENT

Parallel Pathways for Potential Partnerships (PPPP)

- Parallel
 - Enables AI/ML collaborations faster, and parallel to any work on establishing full partnerships regarding larger datasets.
- Pathways
 - Fast-track pathways to data by *leveraging sprints* with small *data sets in format of actual data* so that AI/ML tools can be built which are reusable when potential agreement on larger/full dataset is made.
- Potential
 - Collaboration around datasets- without any expectations on federal/non-federal sides on outcome or future partnership
- Partnerships
 - Signed research agreements around a particular AI/ML problem/task with potentially mutual expectations





Questions?

• **Contact:** gil.alterovitz@va.gov



DISCOVERY * INNOVATION * ADVANCEMENT