FARR:
FAIR in ML, AI Readiness, & Reproducibility Research Coordination Network

FARR
FAIR, AI Readiness & Reproducibility

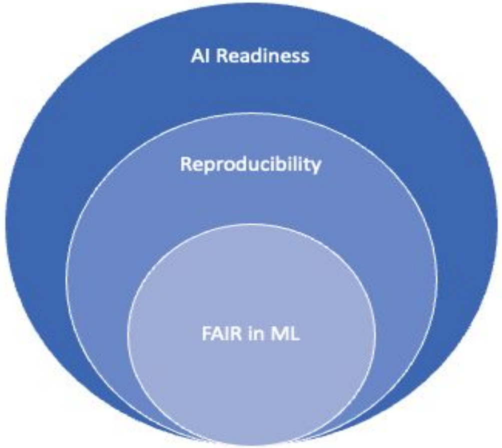
BRDI

April 20, 2023

SDSC SAN DIEGO SUPERCOMPUTER CENTER | NC STATE UNIVERSITY | ILLINOIS NCSA | National Center for Supercomputing Applications | UC San Diego | SCRIPPS INSTITUTION OF OCEANOGRAPHY

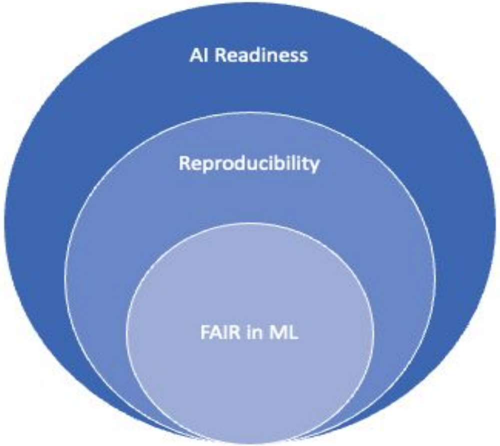
This work is supported through the NSF award #2226453.

Motivation

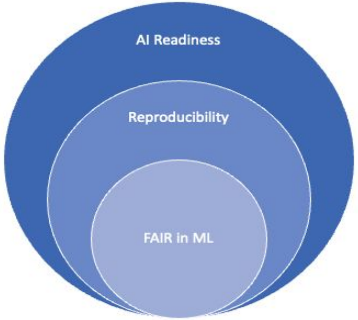


- If 80% of time spent with data is wrangling, can FAIR principles increase efficiency of people and machines?
- What does it mean for a repository or an organization to be AI ready?
  - What roles can repositories play in AI readiness?
- What are community practices in AI reproducibility?
  - What are the gaps in current knowledge?

AI Readiness



Literature Review of ‘AI Readiness’

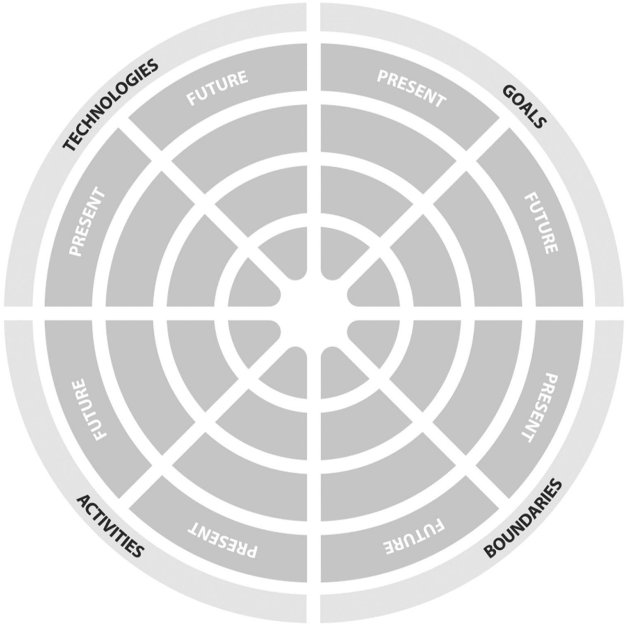


- Term used in computer science, healthcare, business, systems engineering
- Granularity varies: country, organizations/companies, data
- Definition examples
    - **Quantity and quality of AI-related research** in a country (Vuong)
    - **Preparedness** of organizations **to implement change** involving applications and **technology related to AI** (Najdawi)
    - Organization’s abilities to **deploy and use AI** in ways *that add value* to the organization (Holmström’s AI Readiness framework)
    - AI readiness process comprises a set of **strategic actions to be performed prior to AI adoption** (Issa)
    - **Data that can support AI solutions** should be known, understood, available, fit for purpose, secure (McKinsey’s 5 Steps to AI ready data)

AI Readiness Frameworks

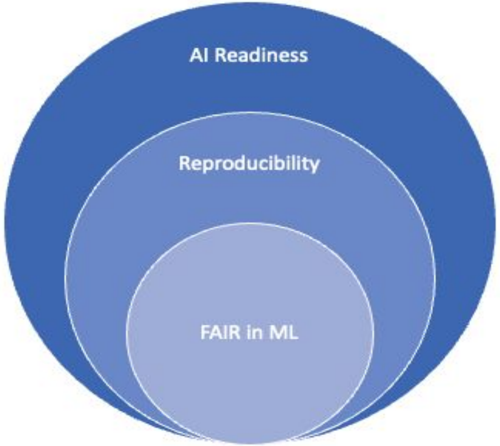
Table 2: Extended TOE framework for AI-Readiness at Business

Readiness Dimensions	Main Categories	Sub-Categories
Technology Readiness	Relative advantage	
	Compatibility	Business process*, Business case*
Organizational Readiness	Culture *	Top management support, change Management*, Innovative Culture*
	Organizational Size	
	Resources	Budget*, employees*, data* (availability, protection and quality)
	Organizational Structure*	
Environmental Readiness	Competitive pressure	
	Government regulatory issues	Employees'council, General Data Protection Regulation (GDPR)*
	Industry requirements *	
	Customer readiness *	



Najdawi, Anas. (2020). Assessing AI Readiness Across Organizations: The Case of UAE. 1-5. https://doi.org/10.1109/ICCCNT49239.2020.9225386 ; Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business Horizons*, *65*(3), 329–339. https://doi.org/10.1016/j.bushor.2021.03.006

AI Reproducibility



AI Results Can Vary (Software Bug)

Processing Unit:Software Environment	0	1	2	3	4
NVIDIA A40:tensorflow_22.03	0.86584	0.86584	0.86584	0.86584	0.86584
NVIDIA A100-40GB.tensorflow_22.03	0.86580	0.86580	0.86580	0.86580	0.86580
Tesla V100-16GB:tensorflow_22.03	0.85203	0.85203	0.85203	0.85203	0.85203
NVIDIA A40:tensorflow_22.06	0.86584	0.86584	0.86584	0.86584	0.86584
NVIDIA A100-40GB:tensorflow_22.06	0.86580	0.86580	0.86580	0.86580	0.86580
Tesla V100-16GB:tensorflow_22.06	0.85203	0.85203	0.85203	0.85203	0.85203
NVIDIA A40:tensorflow_2.8.0-gpu	0.86571	0.86555	0.85680	0.85927	0.86147
NVIDIA A100-40GB:tensorflow_2.8.0-gpu	0.86511	0.84516	0.86360	0.85895	0.86479
Tesla V100-32GB:tensorflow_2.8.0-gpu	0.85983	0.78835	0.85764	0.86031	0.86908
NVIDIA A40:tensorflow_2.9.1-gpu	0.83631	0.86287	0.83891	0.85724	0.85276
NVIDIA A100-40GB:tensorflow_2.9.1-gpu	0.85979	0.86619	0.83156	0.86412	0.84564
Tesla V100-16GB:tensorflow_2.9.1-gpu	0.84447	0.85320	0.85996	0.86511	0.86779
NVIDIA A40:tensorflow_2.9.1-gpu-cuda11.3-cudnn8.2	0.86184	0.86184	0.86184	0.86184	0.86184
NVIDIA A100-40GB:tensorflow_2.9.1-gpu-cuda11.3-cudnn8.2	0.84631	0.84631	0.84631	0.84631	0.84631

Slide source: Kevin Coakley (SDSC/NTNU)

Tolerance for Erroneous Conclusions: High

Image classification

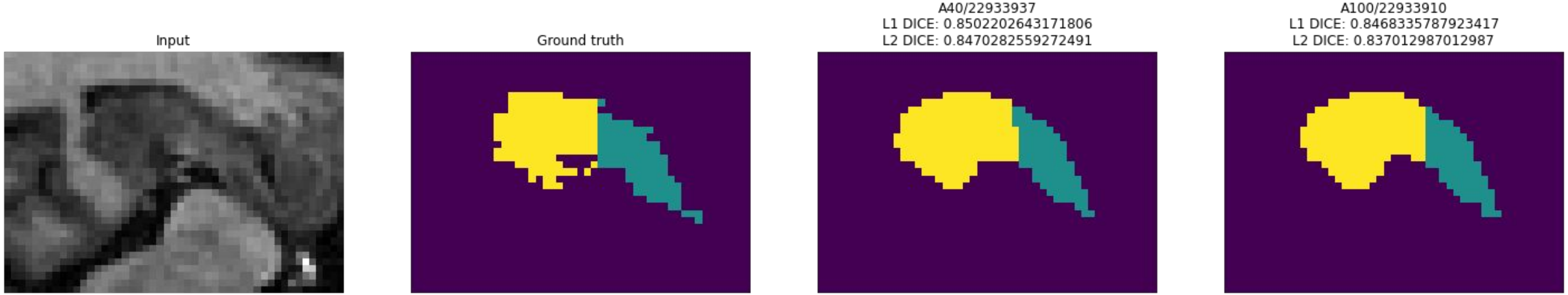
Find the four-leafed clover →

0187683105	0.8626000285148621	0.8573200106620789	0.8610799908638000	0.86492002
9915313721	0.8525599837303162	0.8678799867630005	0.8515599966049194	0.86799997
0276374817	0.8572800159454346	0.8664399981498718	0.8625199794769287	0.86211997
9826049805	0.8642799854278564	0.8643199801445007	0.8637199997901917	0.86291998
9908638000	0.8507199883460999	0.8378000259399414	0.8429200053215027	0.85740000
0108718872	0.8612400293350220	0.8637199997901917	0.8675600290298462	0.86783999
9786567688	0.8600000143051147	0.8575999736785889	0.8620799779891968	0.83524000
0290298462	0.8675600290298462	0.8675600290298462	0.8675600290298462	0.86756002
0204086304	0.8561199903488159	0.8648800253868103	0.8633199930191040	0.86211997
0290298462	0.8675600290298462	0.8675600290298462	0.8675600290298462	0.86756002
9830055237	0.8628799915313721	0.8389199972152710	0.8572400212287903	0.85276001
9857330322	0.8658800125122070	0.8632799983024597	0.8652399778366089	0.84384000
0012779236	0.8557199835777283	0.8619599938392639	0.8493199944496155	0.80923998
0214385986	0.8543199896812439	0.8555999994277954	0.8573200106620789	0.85979998
9811172485	0.8661999702453613	0.8315600156784058	0.8641200065612793	0.84564000
9716758728	0.8642399907112122	0.8579599857330322	0.8552399873733521	0.86268001
9885559082	0.8166800141334534	0.8648800253868103	0.8613200187683105	0.85868000
0199508667	0.8500400185585022	0.8637599945068359	0.8516399860382080	0.86103999
9983024597	0.8689600229263306	0.8511599898338318	0.8640800118446350	0.85535997

Slide source: Kevin Coakley (SDSC/NTNU)

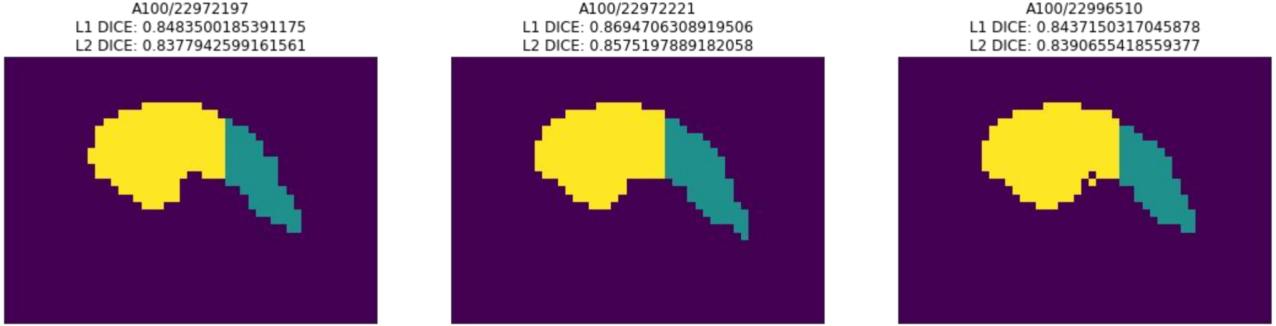
Tolerance for Erroneous Conclusions: Low

Running nnUNet with Medical Imaging



What amount of error is tolerable?

Need to quantify the variance



Slide source: Kevin Coakley (SDSC/NTNU)

FARR Goals and Activities

We welcome individual researchers, institutions/organizations, CI providers, repositories/facilities, and networks of facilities in Computer Science, Geosciences, and the ‘Research Data’ Community. FARR provides a neutral and novel meeting place for bridging multiple networks.

- Building communities to
  - promote better practices for AI
  - harness community efforts
  - improve efficiency and reproducibility
  - stimulate and enhance new research
- Activities will include
  - workshops
  - assessing community needs
  - fostering new collaborations (proposals)
  - setting research agendas
  - community-led reports

Incorporating EarthCube’s sustainability lessons from the beginning.

FARR Team

PI, Christine Kirkpatrick

Co-PI, Karen Stocks

Co-PI, Yuhan (Douglas) Rao

Co-PI, Daniel S. Katz

Project Director,
Lynne Schreiber

Sr. Personnel,
Kevin Coakley

Project Manager,
Julie Christopher

Communications,
Kim Mann Bruch

SDSC SAN DIEGO SUPERCOMPUTER CENTER

NC STATE UNIVERSITY

ILLINOIS NCSA | National Center for Supercomputing Applications

UC San Diego

SCRIPPS INSTITUTION OF OCEANOGRAPHY

FARR Partners

CERN

WEST BIG DATA INNOVATION HUB

Council of Data Facilities

RDA RESEARCH DATA ALLIANCE

GO FAIR Foundation

GO FAIR US

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION NOAA U.S. DEPARTMENT OF COMMERCE

MIDWEST BIG DATA HUB

itsc

Florida State University COAPS Center for Ocean-Atmospheric Prediction Studies

AI AFRICA CONSORTIUM

Radiant Earth Foundation
EARTH IMAGERY FOR IMPACT

TILOS

ARTIFICIAL INTELLIGENCE · ENVIRONMENTAL SCIENCE · RISK COMMUNICATION

CoMSES NETWORK

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

Lawrence Livermore National Laboratory

Argonne NATIONAL LABORATORY

ML Commons

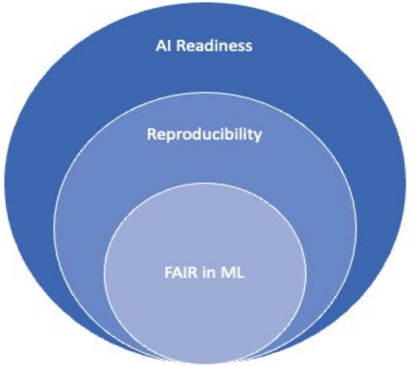
COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Data Discovery Studio

Early Signals

- (AI) Reproducibility is complex and resource intensive
  - Must prioritize key aspects to document
  - Lack of awareness of issue
  - Excellent library personnel who want to help
  - Papers >3 years old difficult to reproduce
    - *Even if containers are provided*
    - Suggests shifting priorities for curation teams
- AI Readiness
  - Definition evolving within communities
  - The other FAIR - Fully AI Ready
    - How do I get there?
    - Are we there yet?
    - How will I know when I'm FAIR?
- Listening Tour
  - Need for onramps:
    - 'Where do I start with AI?'
    - How do I choose a GPU?

Thank You



For more info, contact us at community@farr-rcn.org
To join our mailing list, visit http://farr-rcn.org