

JAVIER DUARTE
JULY 13, 2023



NSF HDR POSTBAC WORKSHOP 2023 SUMMARY

- ▶ Webpage: <https://indico.cern.ch/e/hdrpostbac2023>
- ▶ Workshop June 20-21, 2023 located at UCSD
- ▶ Goals: to bring together trainees, including postbaccalaureate fellows, graduate students, and postdoctoral fellows, from all NSF HDR institutes to
 - ▶ Present & discuss their research
 - ▶ Network
 - ▶ Provide feedback on the programs
 - ▶ Learn about cutting-edge data science research
 - ▶ Participate in professional development
 - ▶ Prepare for the next phase of their education/training/employment



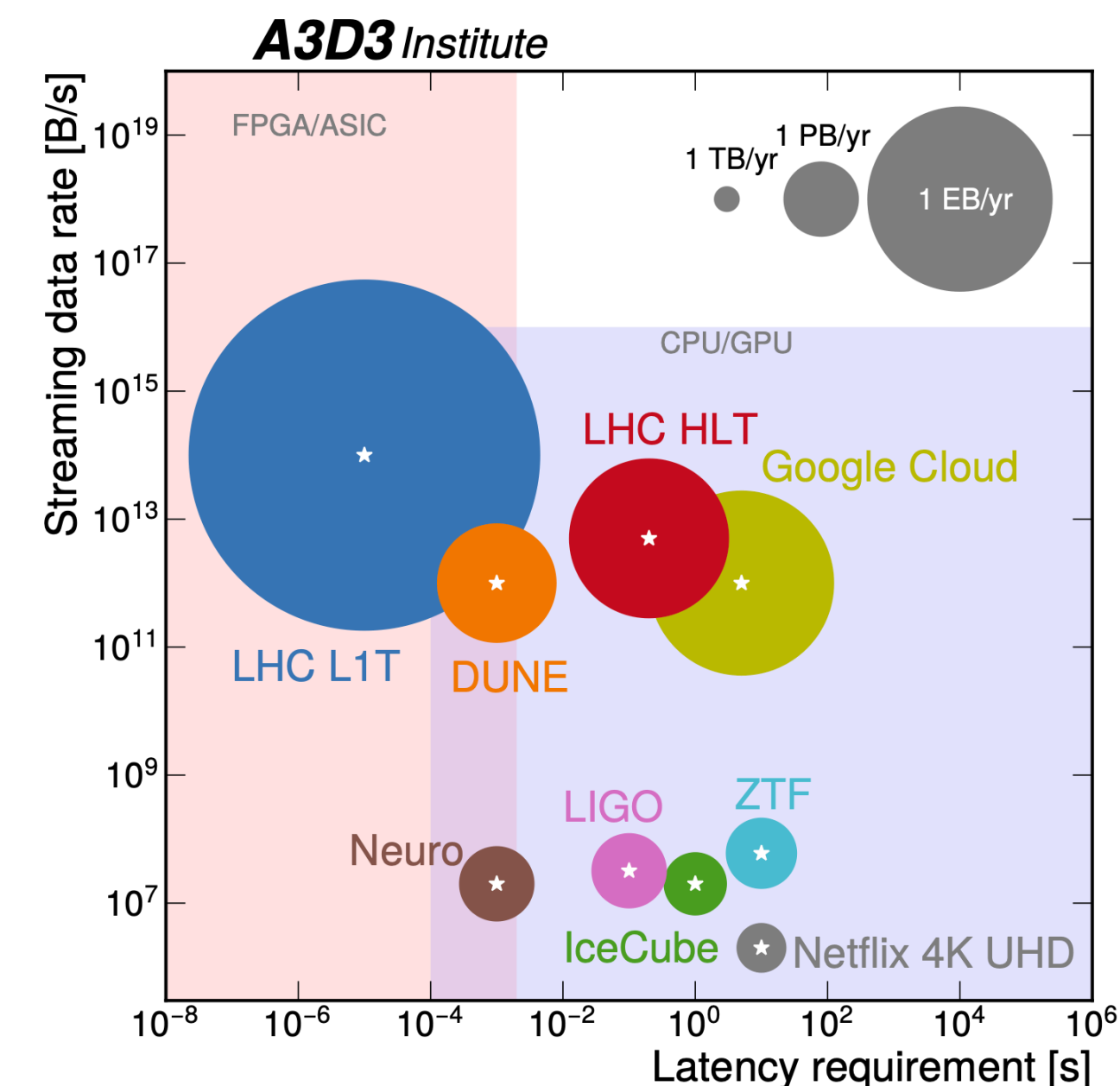
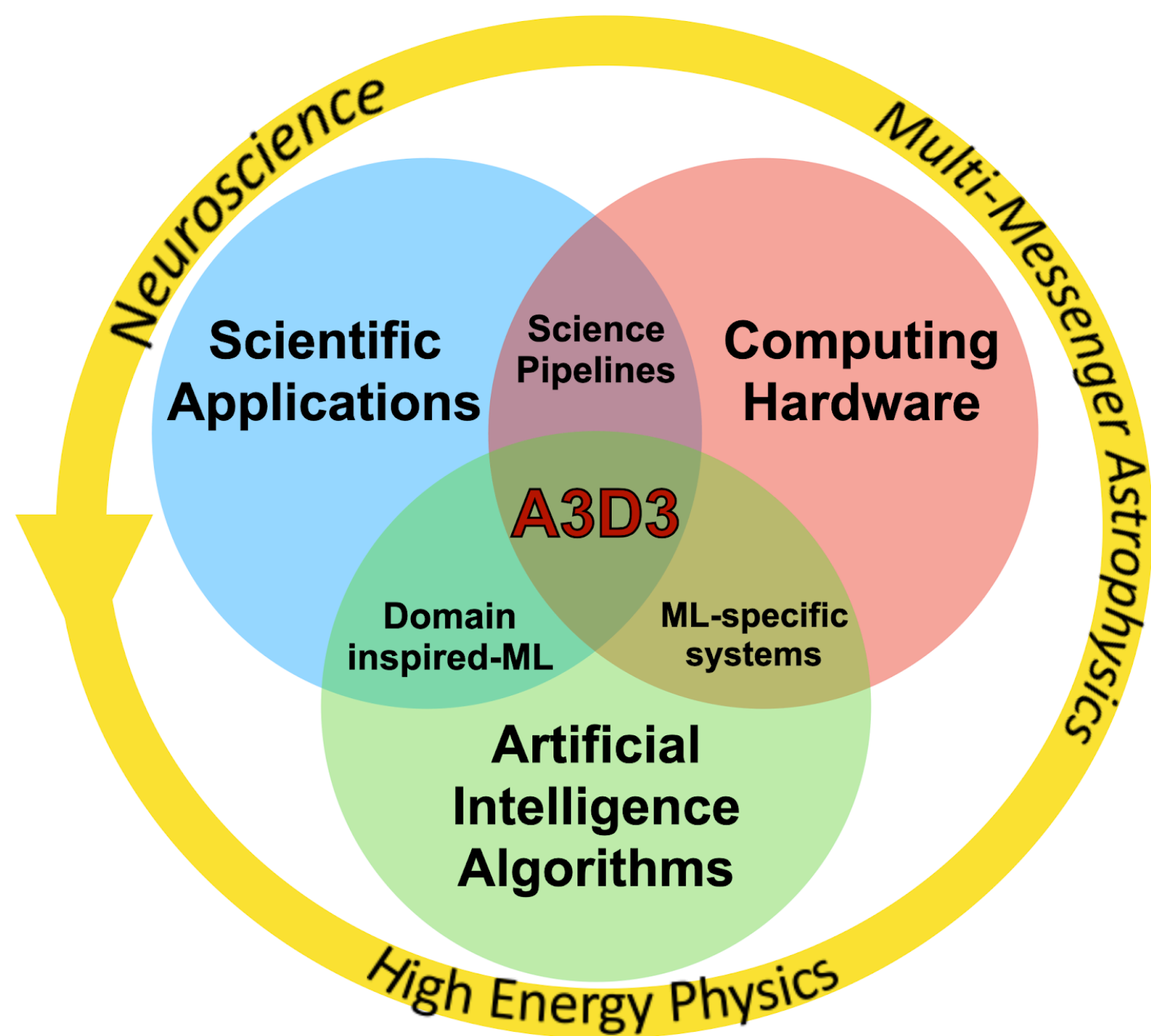
- ▶ 14 attendees (postbacs, grad students, postdocs, PIs)
 - ▶ 4 Y1 A3D3 postbacs + 2 Y2 A3D3 postbacs
 - ▶ 1 ID4 postbac
 - ▶ 2 I-GUIDE climbers
 - ▶ + several postdoc/grad student facilitators



- ▶ HDR: Institutes for Data-Intensive Research in Science and Engineering (I-DIRSE) seeks to create an integrated fabric of interrelated institutes that can accelerate discovery and innovation in multiple areas of data-intensive science and engineering
- ▶ Ecosystem of 5 institutes: nsfhdr.org
 - ▶ Accelerated AI Algorithms for Data-Driven Discovery (A3D3)
 - ▶ Institute for Geospatial Understanding through an Integrative Discovery Environment (I-GUIDE)
 - ▶ Institute for Data Driven Dynamical Design (ID4)
 - ▶ Institute for Harnessing Data and Model Revolution in the Polar Regions (iHARP)
 - ▶ Imageomics Institute

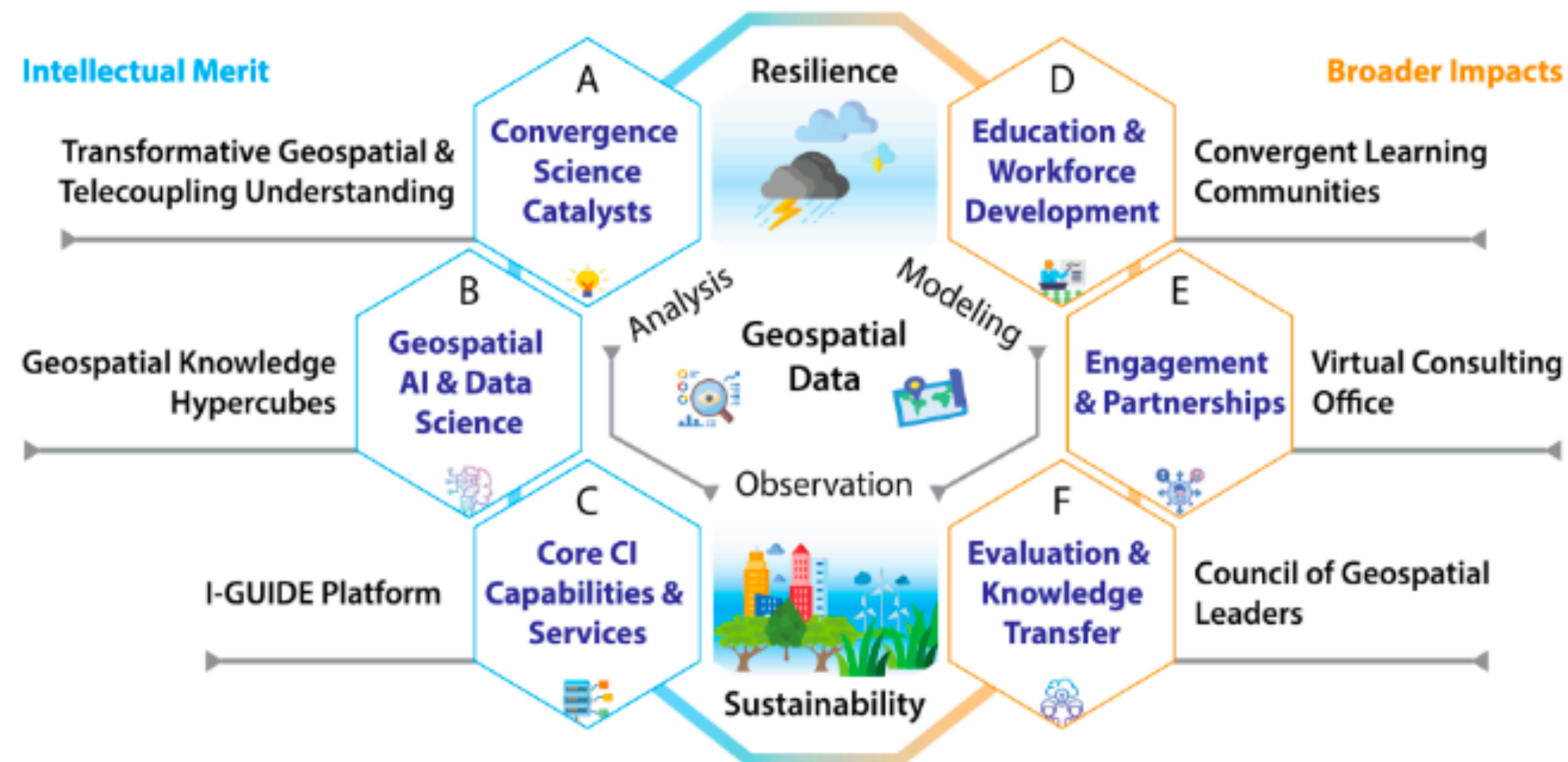


- ▶ Multi-disciplinary and geographically distributed entity with the primary mission to lead a paradigm shift in the application of real-time artificial intelligence (AI) at scale to advance scientific knowledge and accelerate discovery
- ▶ Webpage: a3d3.ai, award: [OAC-2117997](https://www.oac-senate.org/awards/2021-2022)



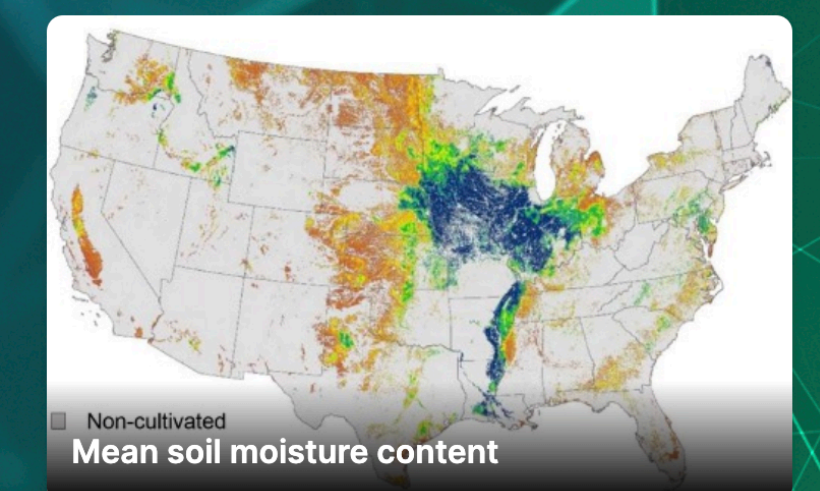
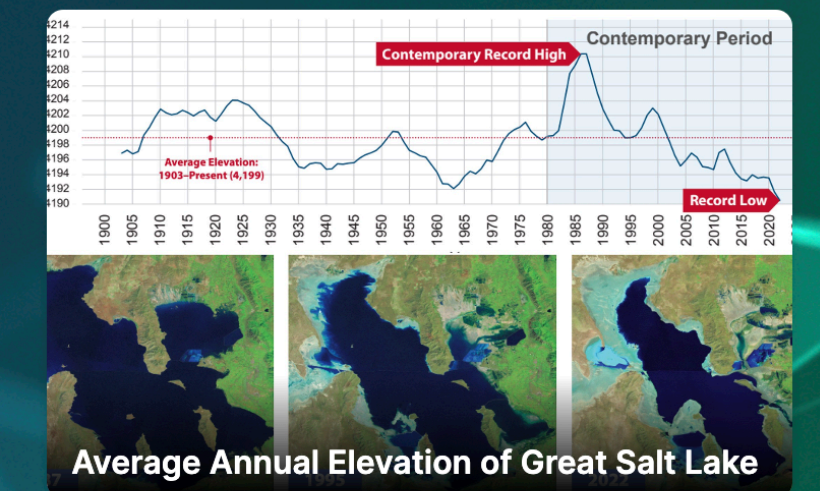
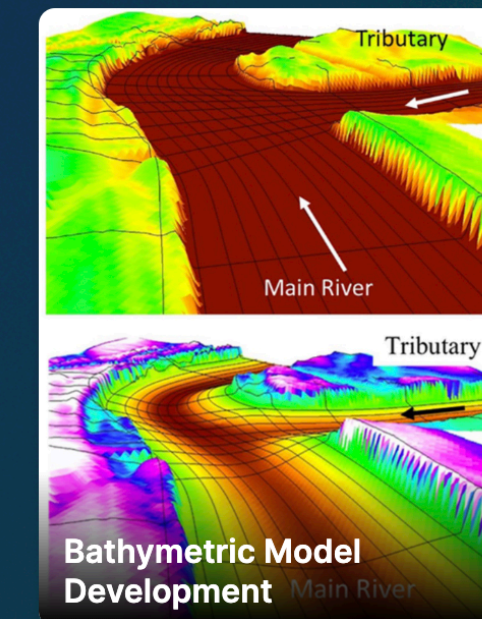
INSTITUTE FOR GEOSPATIAL UNDERSTANDING THROUGH AN INTEGRATIVE DISCOVERY ENVIRONMENT (I-GUIDE)⁶

- ▶ Vision: Digital discovery and innovation through harnessing the geospatial data revolution
- ▶ Mission: Transform convergence and geospatial sciences for holistic sustainability solutions
- ▶ Webpage: iguide.illinois.edu, award: [OAC-2118329](#)



Map.
Connect.
Discover.

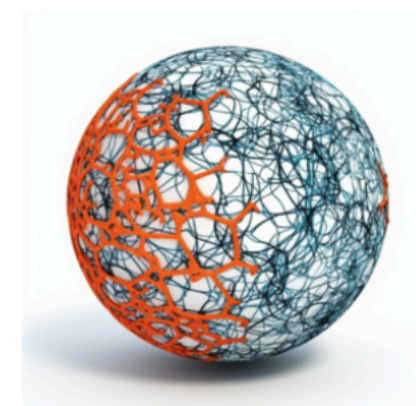
I-GUIDE is empowering diverse communities to produce data-intensive solutions to society's resilience and sustainability challenges.



- ▶ Transform how scientists & engineers harness data when designing materials and structures
- ▶ From chemistry to civil engineering, create platforms that accelerate the discovery of new mechanisms and dynamics through the union of human & machine intelligence
- ▶ Train next generation and engaging with the broader data-driven community
- ▶ Webpage: mines.edu/id4, award: [OAC-2118201](https://oac.mines.edu/award/OAC-2118201)



Navigating complex design spaces



Custom AI for domain-specific simulations



Knowledge distillation

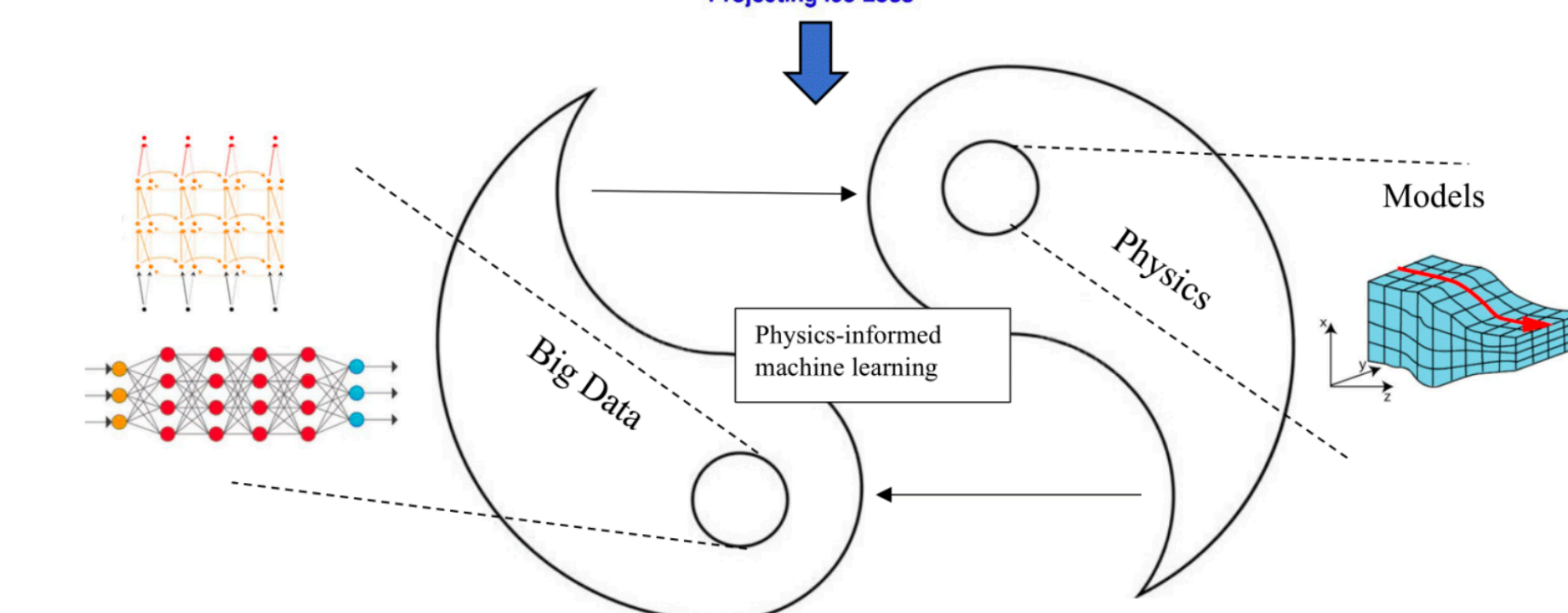
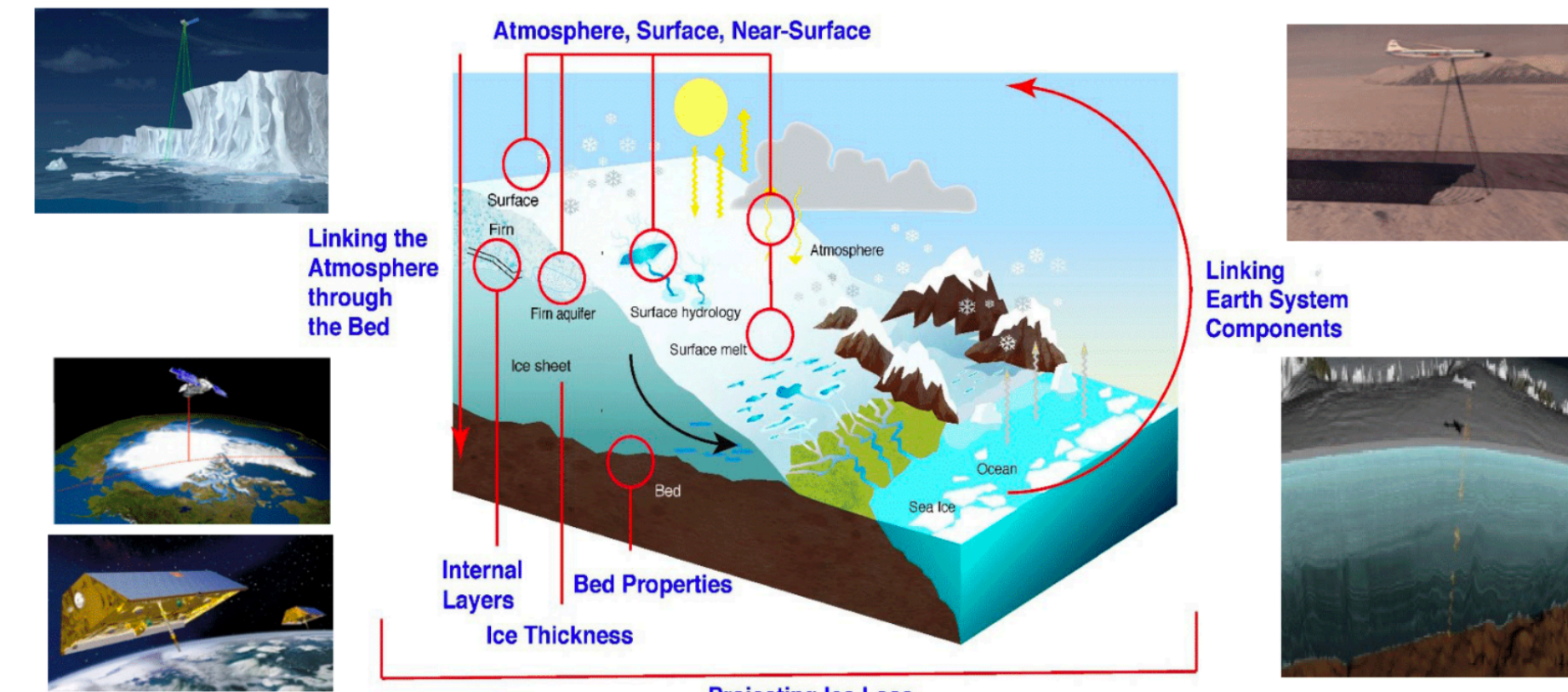


Human engagement and guidance

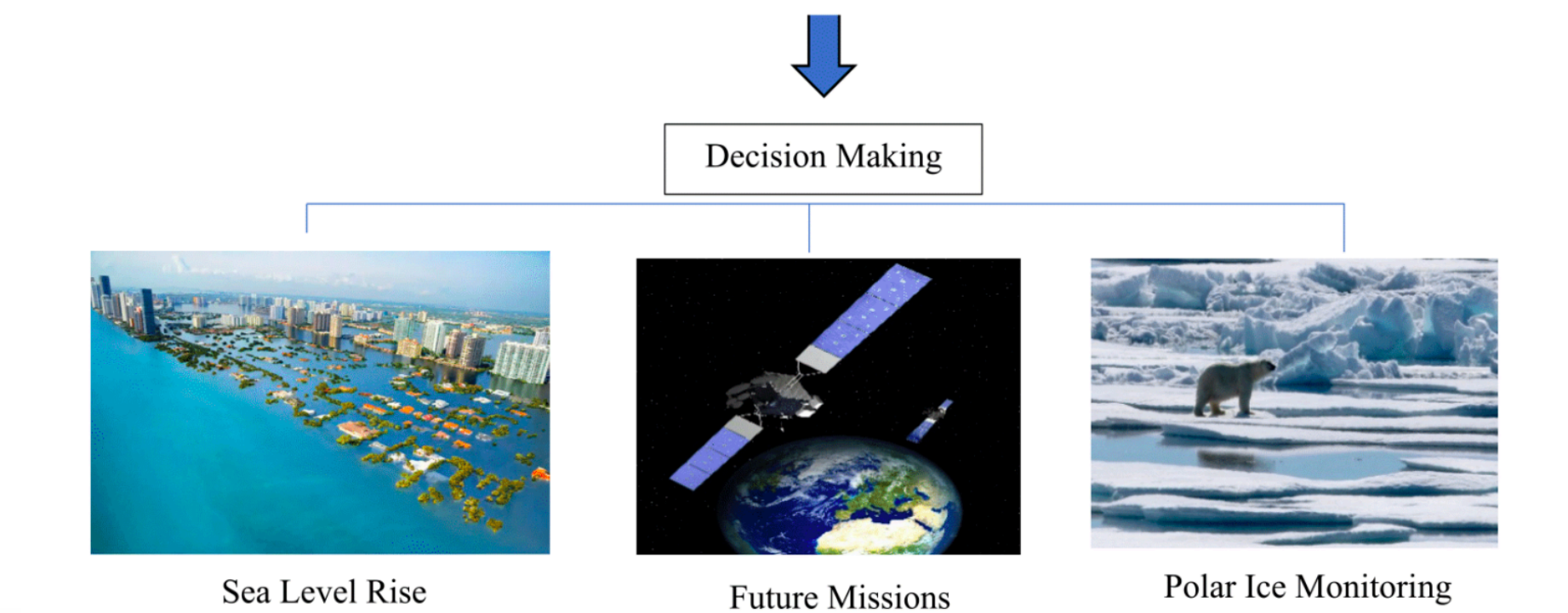


INSTITUTE FOR HARNESSING DATA AND MODEL REVOLUTION IN THE POLAR REGIONS (IHARP)

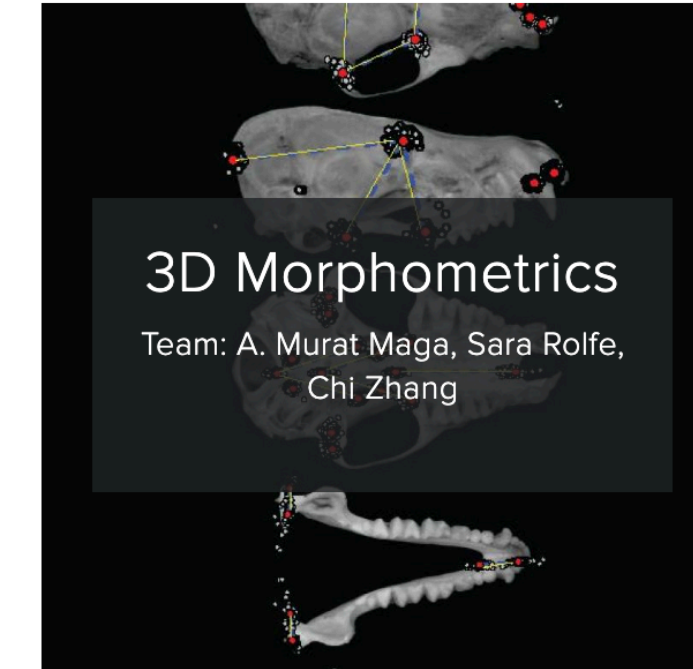
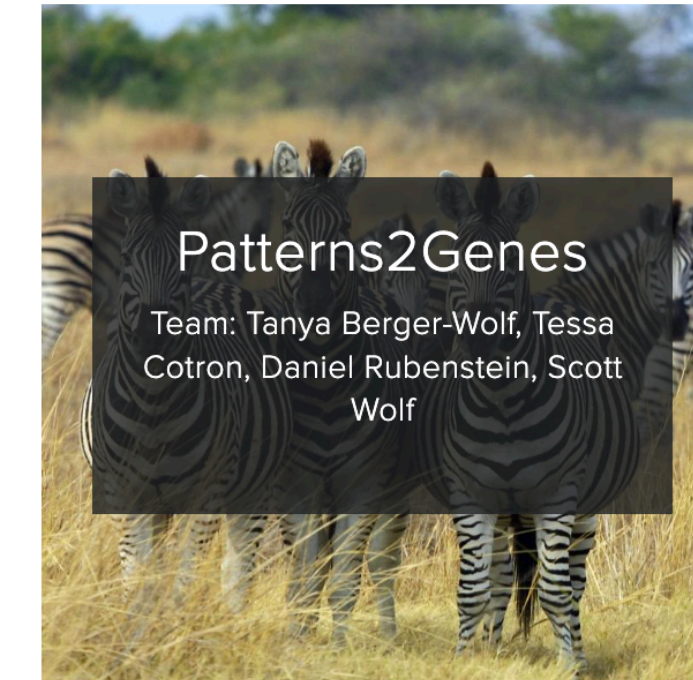
- ▶ Advances our understanding of the response of polar regions to climate change and its global impacts by deeply integrating data science and polar science to spur physics-informed, data-driven discoveries
- ▶ Webpage: iharp.umbc.edu, award: [OAC-2118285](https://www.nsf.gov/awardsearch/showAward?AWD_ID=OAC-2118285)



Knowledge discovery: Ice thickness, Ice dynamic, Surface mass balance, Subglacial topography, Spatial-temporal monitoring,....



- ▶ Mission: Create collaborative research, training, and community-facing environment for extracting existing and new biological traits from images of organisms, with infrastructure for cyber, information, and model development
- ▶ Vision: Establish new scientific field that harnesses data science and computing, and rapidly expanding collections of biological image data, in order to accelerate biological understanding of phenotypic traits extracted from images of organisms
- ▶ Webpage: imageomics.osu.edu, award: [OAC-2118240](#)



▶ Timetable:

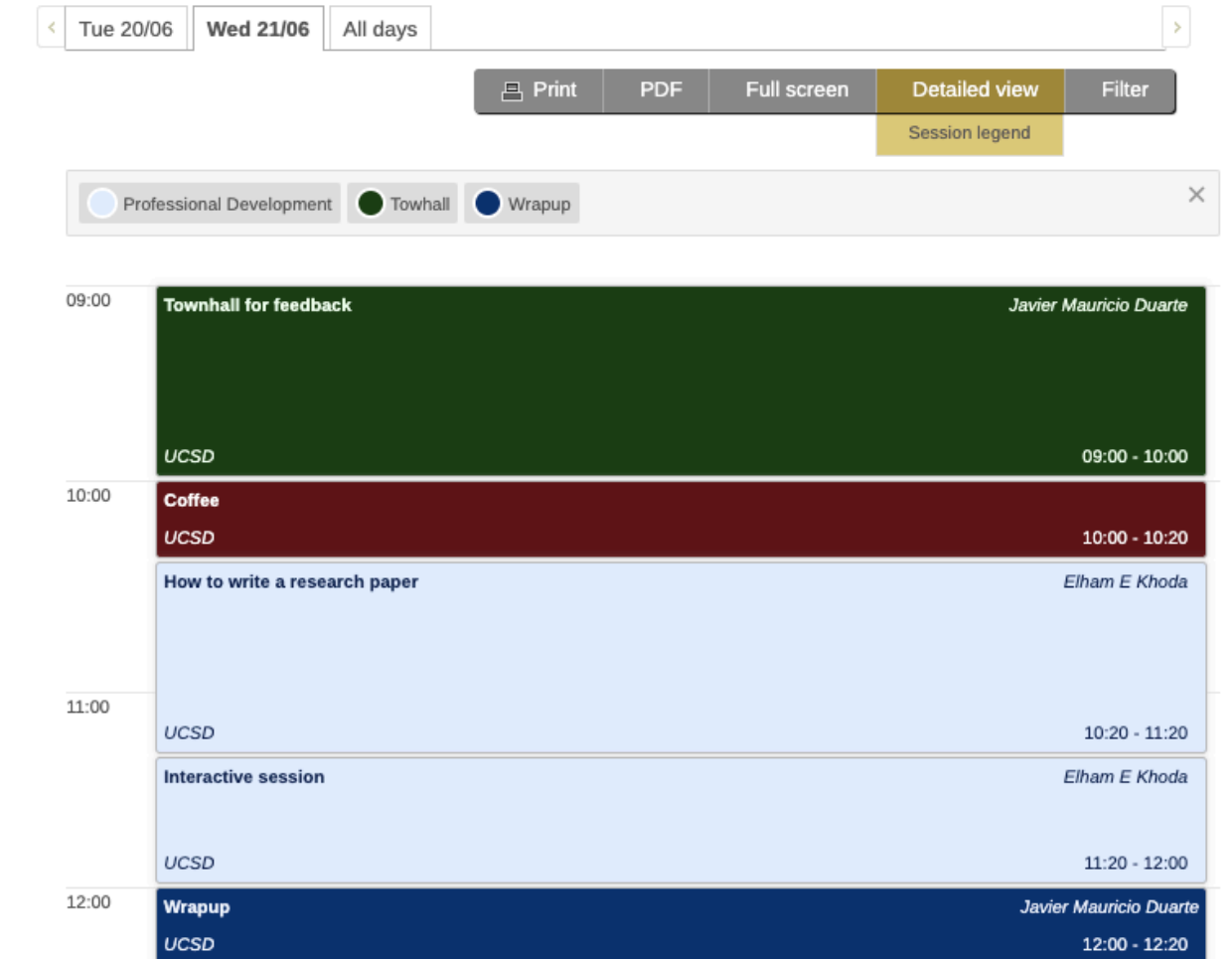
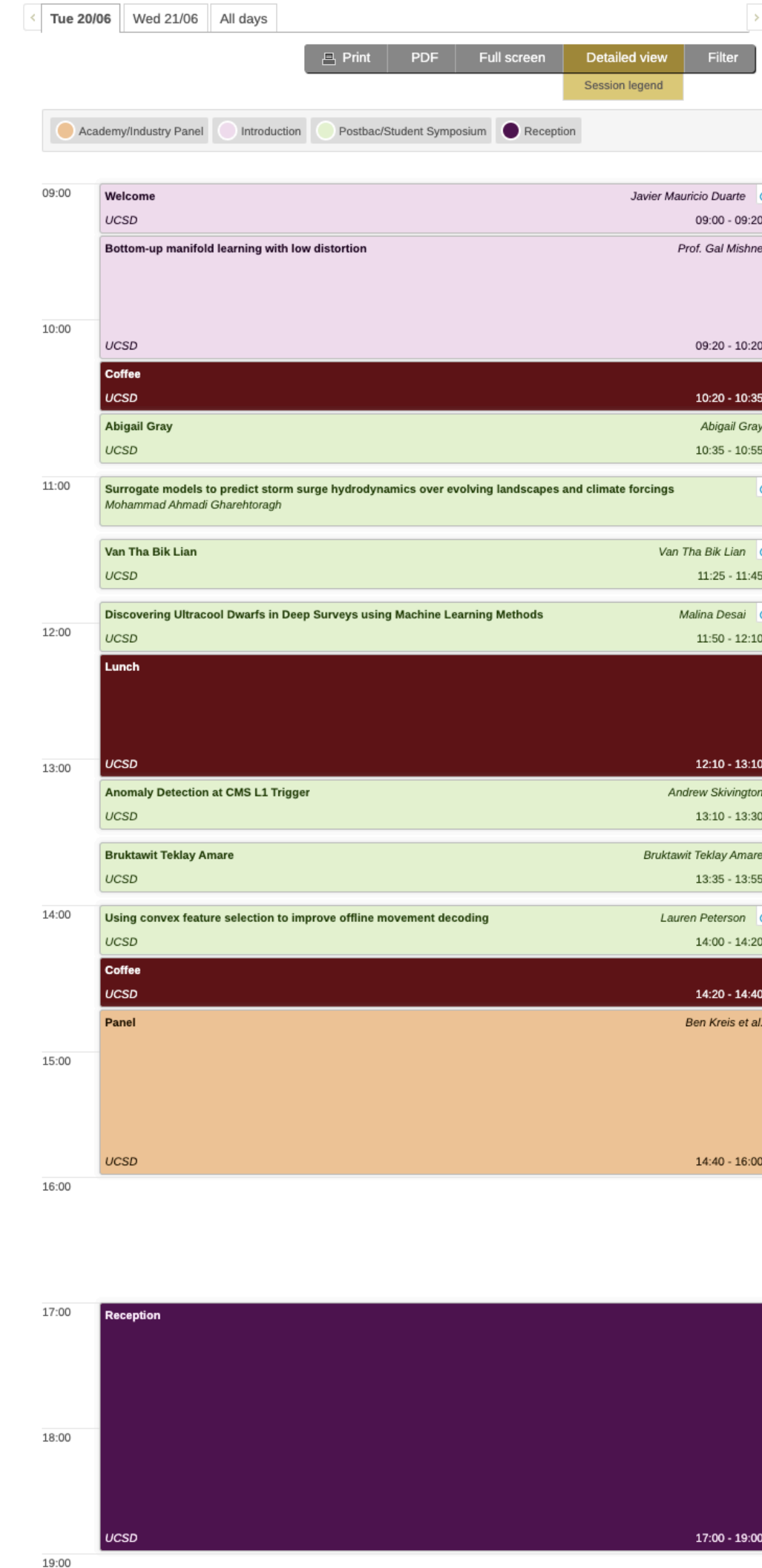
<https://indico.cern.ch/event/1253923/timetable>

▶ Day 1

- ▶ Keynote speaker Prof. Gal Mishne
- ▶ Postbac symposium to share research
- ▶ Industry/academia panel
- ▶ Reception

▶ Day 2

- ▶ Town hall for feedback
- ▶ Professional development: "How to write a research article 101" facilitated by Dr. Elham Khoda



- ▶ Gal Mishne is an Assistant Professor in the [Halicioğlu Data Science Institute](#) (HDSI) at UC San Diego, and affiliated with the ECE department, the CSE department and the Neurosciences Graduate program
- ▶ Part of the [Neurotheory Network](#) and [Pathways 2 AI](#)
- ▶ Research interests:
 - ▶ Manifold learning
 - ▶ Diffusion geometry
 - ▶ Computational neuroscience
 - ▶ Image processing and biomedical signal processing
 - ▶ Applied harmonic analysis



Bottom-up manifold learning with low distortion

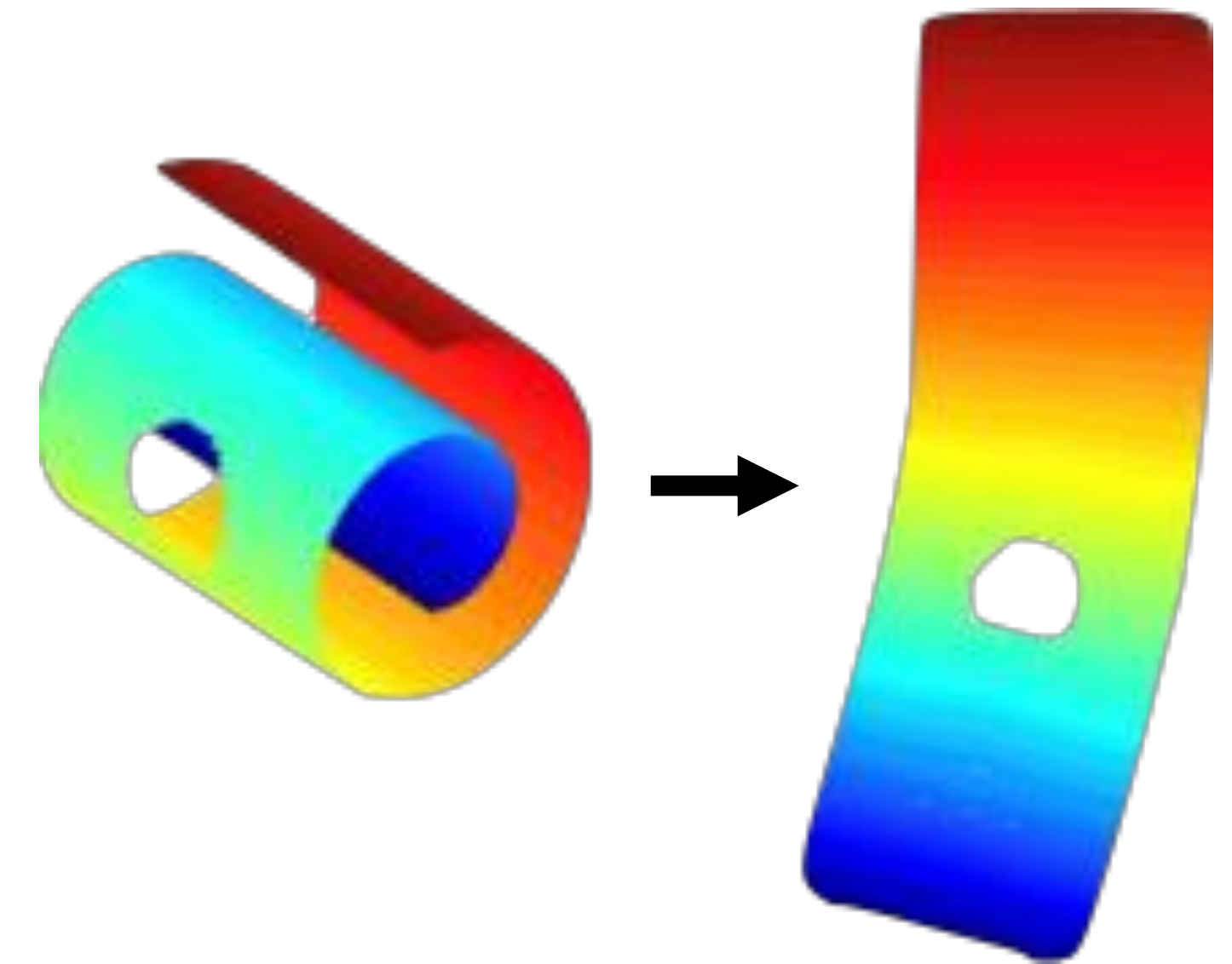
Gal Mishne
Halicioğlu Data Science Institute
UC San Diego

Joint work with Dhruv Kohli, Alex Cloninger,
Bas Nieuwenhuis and Devika Narain



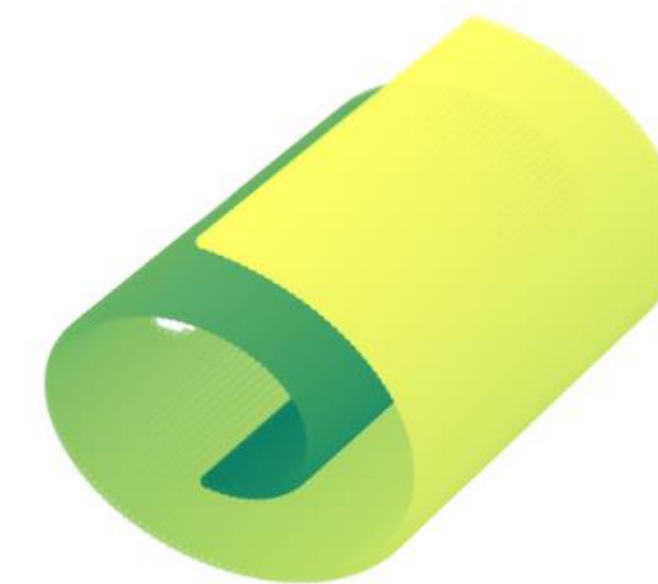
Manifold learning

- Learn manifold from data
- Non-linear **representation** of low-dimensional manifold.
- Preserve **geometric** properties
- Embedding with top eigenvectors of the
 - Covariance matrix (PCA)
 - Normalized graph Laplacian (Laplacian Eigenmaps)
 - Random-walk graph Laplacian (Diffusion maps)
 - ...



*[Tenenbaum et al., 2000,
Roweis and Saul, 2000,
Belkin and Niyogi, 2001,
Donoho and Grimes, 2002,
Coifman and Lafon, 2004,
van der Maaten and Hinton 2008,
McInnes et al. 2018, ...]*

Manifold learning frameworks



Top-down

- Initial global embedding of the data
- Optional: refine it iteratively by minimizing a measure of local distortion.
- ISOMAP, Laplacian Eigenmaps, t-SNE, UMAP...

Bottom-up

- Calculate local views with low distortion
- Solve alignment problem to register views to global embedding
- LTSA, LDLE, RATS



Bottom-up manifold learning

Bottom-up manifold learning:

- Local neighborhoods in the data have their own parameterization (local views) with low distortion
- Local views are aligned to obtain a global embedding



- ▶ 7 talks from A3D3, I-GUIDE, and ID4 trainees

A3D3 Post-Bacc Final Presentation/Research Update



Abby Gray

PI: Prof. Michael Coughlin
University of Minnesota
Multi-Messenger Astrophysics



Denoising Autoencoder for raw, wireplane, waveforms in DUNE's LArTPC detector

June 20, 2023

Van Tha Bik Lian, Kate Scholberg, Mike Wang, Benjamin Hawks, Janina Hackenmueller, Tejin Cai, Jovan Mitrevski, Tingjun Yang, Thomas Junk, Maira Khan, Jennifer Ngadiuba



SURROGATE MODELS TO PREDICT STORM SURGE HYDRODYNAMICS OVER EVOLVING LANDSCAPES AND CLIMATE FORCINGS

Mohammad Ahmadi
NSF-Iguide



6/20/23 | 1

Discovering Ultracool Dwarfs in Deep Surveys using Machine Learning Methods

Malina Desai

UCSan Diego

Using convex feature selection to improve offline movement decoding

Lauren Peterson, Si Jia Lee, Leo Scholl, Amy Orsborn



Anomaly Detection at CMS L1 Trigger

Graph Neural Networks, Variational Autoencoders, and Interpreting Anomalous Latent Spaces

Andrew Skivington, A3D3 Postbac, 20/07/2023

Visual Alchemist:
Visualization tool for data-driven materials science

Shayan Monadjemi
Danni Liu
Bruktawit Amare

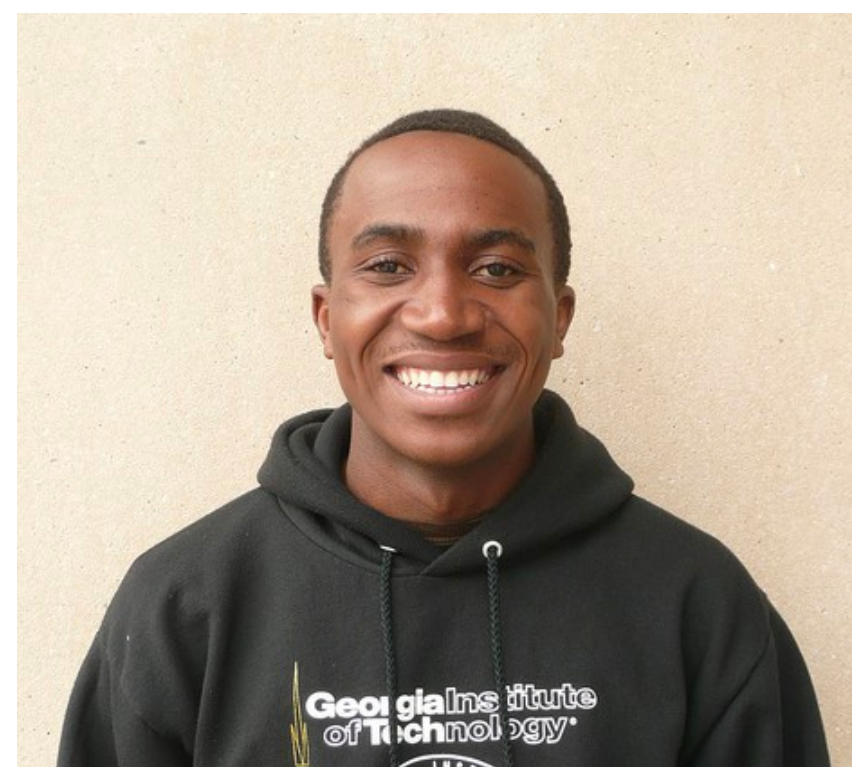




- ▶ Ben Kries is a Senior Data Scientist at Apple. Previously he was a postdoctoral research associate at Fermilab. He received his Ph.D. in physics at Cornell.
- ▶ Philip Hebda is Senior Data Scientist at Roblox. He was previously at Pinterest and Netflix. He received a Ph.D. in physics at Princeton.



- ▶ Rushil Roy is System Engineer at Qualcomm. He received his bachelor's degree in Electrical Engineering with a specialization in Machine Learning and Controls at UCSD
- ▶ Olivia Weng is a Ph.D. student in Computer Science and Engineering at UCSD



- ▶ Christian Aganze is a Ph.D. student in Astrophysics at UCSD
- ▶ Anthony Aportela is a Ph.D. student in Physics at UCSD

- ▶ What kinds of positions did you consider in your job search and why?
- ▶ What are the most useful skills in your current job that you learned during your undergraduate career?
- ▶ If you conducted research during your undergraduate career, how useful was that experience for your current position?
- ▶ What do you wish you knew before you started applying to your current career?
- ▶ Are there any skills you wished you had developed more before your current position?
- ▶ How has work/life balance impacted your current career and career progression?
- ▶ Where do you see your career progressing in the next 5 years?
- ▶ What made you decide to go to graduate school?
- ▶ What about graduate school is different from your expectations as an undergraduate?
- ▶ What made you decide to go into industry rather than academia?
- ▶ What surprises you most about working in industry?
- ▶ What backgrounds (majors, previous careers, etc.) or skills does your company look for in hiring?

- ▶ More structured initial few months would have added value
- ▶ Initial in-person meet with fellow postbacs
- ▶ ML Intro in each domain, intentional exposure to important research questions, literature review
- ▶ GitHub and similar organizational resources
- ▶ Clear goals/expectations
- ▶ Helping with picking a project (given 1 year timeline)
- ▶ Induction (1st month): really get to know your domain; talks from domain scientists
- ▶ Onboarding project (with PI)
- ▶ I-GUIDE: weekly 1-hour meeting with 20-25 participants
 - ▶ Every month, different leader (good for students, etc. to learn leadership skills)
 - ▶ ~2 research presentations - and how it is related to I-GUIDE goals

How to write a research paper

Elham E Khoda, Javier Duarte

NSF HDR Postbaccalaureate Workshop 2023
June 21, 2023

1

Concise Writing - Examples

Example 2

Identify the unnecessary words in the following sentences:

1. Professor Smith was picked by each **and every** person on the committee.
2. The student needs **to obtain** high marks in science **in order** to study medicine.
3. The manager will, **insofar as is possible**, make sure that the information is true and accurate.

21

Purpose of Science Communication

- Document your research.
- Transfer knowledge to future employees.
- Inform stakeholders (e.g., supervisors and industry leaders) of scientific achievements so they can make well-founded decisions.
- Advance your career.
- Attract new funding.
- Inspire new scientists.
- Educate the public of scientific achievements that can directly impact the quality of their lives.

The social contract isn't complete until scientific results are communicated, particularly when the research was publicly funded

3

Anatomy of a Research Paper

- Abstract
- Introduction
- Related Work (Literature review)
- Methodology
 - Model Description
 - Dataset Description
 -
- Experiment, Results, Discussion
- Conclusion
- References

30