



# Institute for Data Science



## New Center of AI at NJIT

The significant advances in artificial intelligence have been impacting every industry and people's daily lives and are expected to be the driving force of emerging technologies. The Center for AI Research aims to provide an intellectual environment and primary home for AI research initiatives at NJIT. Aiming to promote cutting-edge and high-quality research activities...

**continued on page 02**

## OVERVIEW:

- New Research Center
- New Department of Data Science
- Advisory Board
- Lecture Announcements
- Fall 2021 Seminar Series

Page 03 - SABOC Journal publication

Page 04 - Department of Data Science welcome letter

Page 07 - Advisory Board Members

Page 09 - Student Spotlight

Page 11 - Seminar Series

A graphic for the Center of Artificial Intelligence featuring a hand interacting with a digital interface, with binary code and circuitry in the background.

**Center of  
Artificial  
Intelligence**

A graphic for the Center for Big Data featuring a compass rose with an arrow pointing towards the right, with the words "BIG DATA" at the bottom.

**Center for Big  
Data**

A graphic for the Cybersecurity Research Center featuring a globe with various security icons like a padlock, a magnifying glass, and a shield, with binary code in the background.

**Cybersecurity  
Research Center**

A graphic for the Structural Analysis of Biomedical Ontologies Center (SABOC) featuring a group of people standing together, with a dark background and white text.

**Structural  
Analysis of  
Biomedical  
Ontologies  
Center (SABOC)**

# New Center of Artificial Intelligence

**Meet the Center Director:**  
**Grace Wang Ph.D.** currently a professor and the associate dean for research of Ying Wu College of Computing



The significant advances in artificial intelligence have been impacting every industry and people's daily lives and are expected to be the driving force of emerging technologies. The Center for AI Research aims to provide an intellectual environment and primary home for AI research initiatives at NJIT. It aims to promote cutting-edge and high-quality research activities and to cultivate faculty and student publications and patents in AI and machine learning.

It will foster collaborations and interactions between professors in fundamental AI research and professors who leverage AI methods to solve domain problems and to develop synergies among research groups across different departments and colleges. It plans to train postdoctoral researchers, graduate students, and undergraduate students in fundamental AI research as well as applied AI and prepare them to be a skilled and capable workforce in both academia and industry.

# Visual comprehension and orientation into the COVID-19 CIDO ontology

Journal of Biomedical Informatics  
Volume 120, August 2021, 103861

Ling Zheng, Yehoshua Perl, Yongqun He, Christopher Ochs, James Geller, Hao Liu, Vipina K. Keloth

[Click here to download the full article](#)

The current intensive research on potential remedies and vaccinations for COVID-19 would greatly benefit from an ontology of standardized COVID terms. The Coronavirus Infectious Disease Ontology (CIDO) is the largest among several COVID ontologies, and it keeps growing, but it is still a medium-sized ontology. Sophisticated CIDO users, who need more than searching for a specific concept, require orientation and comprehension of CIDO.

In previous research, we designed a summarization network called “partial-area taxonomy” to support comprehension of ontologies. The partial-area taxonomy for CIDO is of smaller magnitude than CIDO, but is still too large for comprehension. We present here the “weighted aggregate taxonomy” of CIDO, designed to provide compact views at various granularities of our partial-area taxonomy (and the CIDO ontology). Such a compact view provides a “big picture” of the content of an ontology. In previous work, in the visualization patterns used for partial-area taxonomies, the nodes were arranged in levels according to the numbers of relationships of their concepts. Applying this visualization pattern to CIDO's weighted aggregate taxonomy resulted in an overly long and narrow layout that does not support orientation and comprehension since the names of nodes are barely readable. Thus, we introduce in this paper an innovative visualization of the weighted aggregate taxonomy for better orientation and comprehension of CIDO (and other ontologies). A measure for the efficiency of a layout is introduced and is used to demonstrate the advantage of the new layout over the previous one. With this new visualization, the user can “see the forest for the trees” of the ontology. Benefits of this visualization in highlighting insights into CIDO's content are provided. The generality of the new layout is demonstrated.







# DEPARTMENT OF DATA SCIENCE

## The Department of Data Science is Up and Running!

by: James Geller

What do amoebae and university departments have in common? On the face of it, nothing. Amoebae are among the simplest monocellular life forms. University Departments are complex ecosystems of research and education, with professors at four ranks, students in BS, MS, or Ph.D. degree programs, lecturers, staff members, professors of practice, visiting scholars, offices, laboratories, etc. Yet, these two “life forms” have one thing in common: They “multiply” by cell division, they make two out of one.

Up to the 19th-century Natural Philosophy spanned all investigations of nature, without distinctions between subject areas. Isaac Newton's book on the laws of motion and gravity was called "Mathematical Principles of Natural Philosophy" (1687) [1,2]. Work by Lord Kelvin and Peter Guthrie Tait, which helped define much of modern physics, was titled “Treatise on Natural Philosophy” (1867) [1]. With the subsequent increase in knowledge, physics was refined into several subfields, such as quantum mechanics, nuclear physics, electromagnetism, optics, cosmology, etc. [3]. Chemistry was divided into five major branches, starting with the distinction between organic and inorganic chemistry [7], and biology was refined into at least 25 subfields [4].

Many times when a field of study became too large and a natural split emerged, the university departments teaching this subject were also divided. Thus, major universities have distinct departments of nuclear physics, organic chemistry, molecular biology, etc. The impetus for the conceptualization of a new subfield of science is often the development of a distinct new set of research methods. Naturally, there is a delay between the conceptualization of a new field of scholarship and the creation of a new department for it. Administrators have to be convinced that a new department is necessary.



## The Department of Data Science is Up and Running! Continued...

The development of computer science is somewhat more complicated in that it came about in two different ways. On one hand, there were the electrical engineering departments that pioneered this nifty new device called a computer. On the other hand, computer science owes a lot to mathematics as a progenitor, and until today there are universities where mathematics and computer science cohabitate in one department.

The first computer science degree program was offered at the Computer Laboratory of Cambridge University in 1953 [5]. The first computer science department in the United States was formed at Purdue University in 1962 [5]. It was housed in the Mathematical Sciences building [6]. In the early days, the Purdue computer science department provided computing services, besides research and education. In 1968 the department and the Computing Center were separated [6]. Cell division.

When I arrived at NJIT in 1988, I joined the Computer Information Sciences Department (CIS). Within a few years of my joining, the Information Systems faculty in the department developed a scholarly profile that was notably distinct from computer science and valuable in its own right. They adopted, and in many cases pioneered, methods for analyzing the interaction between computers and groups of users. By the turn of the millennium, it was clear that the Information Systems faculty needed and deserved their own department, leaving behind the Department of Computer Science. Cell division.

Around fifteen years later, the Department of Computer Science reached a size and diversity of research and teaching topics that made it unwieldy to administer. I have the right to say this, I was its department chair from 2012 to 2015. Meanwhile the development of new hardware, new data-centric computing methods, and data sets of unprecedented and previously unimaginable sizes, many of them created by social media companies, advanced at breakneck speed. It became clear that a distinguishable new field of study had arisen out of computer science that is now called Data Science. It was time for a new cell division.

At NJIT the result of this latest split is the new Department of Data Science and I am happy and proud to be the interim founding chair of the Department. Just because these divisions of departments have happened over and over again in history, and just because they usually make sense after the fact does not mean that the actual process of creating a new department is trivial. It was not easy, but we did it.

# The Department of Data Science is Up and Running!

## Continued...

The following faculty members have joined me at the Department of Data Science as founding members: Dr. David Bader and Dr. Chase Wu, formerly of the Computer Science Department, as well as Dr. Aritra Dasgupta and Dr. Hai Phan, formerly of the Informatics Department.

So what is Data Science? Anybody with a Google search bar can find a surprising variety of definitions of Data Science, of which I like the following best:

“Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data” [8]. To this should be added that Data Sciences relies heavily on a combination of methods from Machine Learning and Applied Statistics. Once again, it is the set of methods that makes the field.

I would like to close by sharing with you the Vision Statement of the new department:

The data revolution has created novel challenges and unprecedented opportunities. The vision of the Department of Data Science is to create a first-class academic department that trains the next generation of students as data scientists who will solve these grand challenges and innovate through world-class research to take advantage of these opportunities.

We would like to thank Dean Gotsman for his leadership in creating the Department of Data Science and Provost Deek for his unwavering support.

Come and join us!

#### References:

- [1] Natural philosophy. [https://en.wikipedia.org/wiki/Natural\\_philosophy](https://en.wikipedia.org/wiki/Natural_philosophy)
  - [2] Philosophiæ Naturalis Principia Mathematica [https://en.wikipedia.org/wiki/Philosophi%C3%A6\\_Naturalis\\_Principia\\_Mathematica](https://en.wikipedia.org/wiki/Philosophi%C3%A6_Naturalis_Principia_Mathematica)
  - [3] Branches of physics. [https://en.wikipedia.org/wiki/Branches\\_of\\_physics](https://en.wikipedia.org/wiki/Branches_of_physics)
  - [4] Branches of Biology. May 6, 2021. <https://leverageedu.com/blog/branches-of-biology/>
  - [5] Computer science. [https://en.wikipedia.org/wiki/Computer\\_science](https://en.wikipedia.org/wiki/Computer_science)
  - [6] Computer Science Department. Milestones and Principal Events from 1962-1996. Purdue University. <https://www.cs.purdue.edu/history/table1.html>
  - [7] The 5 Main Branches of Chemistry. The Thought Company. <https://www.thoughtco.com/the-5-branches-of-chemistry-603911>
  - [8] Data Science. [https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science) (citing Dhar, V. (2013). "Data science and prediction". Communications of the ACM. 56 (12): 64–73. doi:10.1145/2500499. S2CID 6107147. Archived from the original on 9 November 2014. Retrieved 2 September 2015. and Jeff Leek (12 December 2013). "The keyword in "Data Science" is not Data, it is Science". Simply Statistics. Archived from the original on 2 January 2014. Retrieved 1 January 2014.)
- All references were retrieved on August 4th, 2021.

# MEET OUR ADVISORY BOARD



**ELIZABETH B. BARAYUGA**

PRINCIPAL DATA SCIENTIST  
ADVANCED ANALYTICS  
GROUP - COE (IT)

UPS



**ANNE BAUER**

DIRECTOR OF DATA  
SCIENCE

THE NEW YORK  
TIMES



**RON BEKKERMAN**

CHIEF TECHNOLOGY  
OFFICER

CHERRE INC.



**JONATHAN BENTZ**

SR SA MANAGER FOR  
HIGHER EDUCATION

NVIDIA



**MUKESH DALAL**

CHIEF AI OFFICER

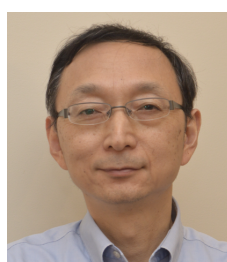
STANLEY BLACK  
& DECKER



**JOSH GREENBERG**

PROGRAM DIRECTOR

ALFRED P. SLOAN  
FOUNDATION



**YIFAN HU**

SENIOR DIRECTOR OF  
RESEARCH

YAHOO  
RESEARCH



**MANOJ KUMAR**

PROGRAM DIRECTOR,  
ANALYTICS SYSTEMS

IBM T.J. WATSON  
RESEARCH  
CENTER



**SHERRY MARCUS**

DIRECTOR, DATA SCIENCE,  
GLOBAL HEAD OF  
ANALYTICS AND INSIGHTS  
PROFESSIONAL SVCS

AMAZON



**KATHY MEIER-HELLSTERN**

PRINCIPAL ENGINEER

GOOGLE  
RESEARCH



**LISA O'CONNOR**

GLOBAL LEAD OF  
CYBERSECURITY RESEARCH  
AND DEVELOPMENT

ACCENTURE  
LABS



**NALINI POLAVARAPU**

VICE PRESIDENT,  
ENTERPRISE ANALYTICS  
AND DATA SCIENCE

BAYER  
CORPORATION



**RACHEL SCHUTT**

MANAGING DIRECTOR,  
HEAD OF AI LABS

BLACKROCK



**DANIEL SILVA**

DIRECTOR, ANALYTICS -  
CHIEF DATA OFFICE

PRUDENTIAL  
FINANCIAL



**KENJI YOSHIHIRA**

DIRECTOR OF US  
ECOSYSTEM INNOVATION

NEC  
LABORATORIES  
AMERICA, INC.





## Lecture Announcements

Institute director **David Bader Ph.D.** is invited to talk at the IEEE Computer Society as a distinguished lecturer on September 9th. Distinguished webinar talk: **Solving Global Grand Challenges with High-Performance Data Analytics**. IEEE Computer Society hosts free interactive webinars through their Distinguished Visitors Program. To register for this upcoming webinar please [click here](#).

For the 16th Conference on Computer Science and Intelligence Systems FedCSIS 2021, held from September 2nd - 5th, **David Bader Ph.D.** gave the keynote talk: **Solving Global Grand Challenges with High Performance Data Analytics**. To learn more about this talk [click here](#).



## Student Spotlight: Center for Big Data

Two Center for Big Data Ph.D. students, **Ms. Qianwen Ye and Mr. Wuji Liu**, presented our research on NoStop: A Novel Configuration Optimization Scheme for Spark Streaming, at the 50th International Conference on Parallel Processing, Argonne National Laboratory in Chicago, IL, USA, August 9-12, 2021 (ICPP21).

## Student Spotlight: IDS Summer 2021 Interns!



**First Name:** Anton

**School:** Eastlake High School (Sammamish, WA)

I've really enjoyed the internship! It gave me lots of opportunities to connect with students and academics, and I was able to work on relevant research with like-minded students. There was a very supportive environment, and even though I entered with limited prior experience, I felt able to pick up the tasks given and gained skills I will use later in my career. Overall, I found this to be a very useful opportunity and would love to continue working in the field.

# IDS Summer 2021 Interns

---



**First Name: Alissa**

**School: Montgomery High School (Montgomery, NJ)**

During my summer internship, I met many new people whose skills inspired me to expand my coding knowledge. I greatly enjoyed working with the other high school interns to brainstorm ideas and code algorithms. With the support of the people around me, I gained experience in coding graph algorithms in C++, which is a topic that interests me a lot. Overall, I really enjoyed this internship, and I look forward to the weeks to come.



**First Name: Vijay**

**School: Montclair High School (Montclair, NJ)**

I have really enjoyed how my summer internship has gone so far. I am working with a really awesome group of high school students to try and create algorithms for constructing Low-Stretch spanning trees in C++. It is something rather new for me and I have really enjoyed being able to learn about a new topic. I also really enjoy our weekly meeting as I get an overview of what some of the other students are researching and writing about.



**First Name: Pranhav**

**School: Churchill High School (Livonia, MI)**

My summer internship was extremely insightful. Before working with Dr. Bader, I had never experienced firsthand how research was conducted or how the process of research was supposed to occur. Yet, with the guidance of Professor Bader, I learned how to break down research papers and analyze the important information that pertained to me and my objectives. Putting all of this information together to create a program that would emulate multiple ideas was indeed a difficult task, however, this internship taught me the importance of perseverance and out-of-the-box thinking which will be of great use in my future. Thus, this virtual experience forced me to change my perspectives and methods multiple times, but in the journey, every task I encountered enabled me to become a better researcher and a better person.



# DATA SCIENCE SEMINAR SERIES

*New Jersey Institute of Technology's Institute for Data Science Fall 2021 will commence soon! Join us for upcoming seminars held Wednesdays at 4 PM Eastern Time. The series includes data science thought leaders from academia and industry.*

<https://datascience.njit.edu/events/>

*Follow us on Eventbrite:*

<https://njit-institute-for-data-science.eventbrite.com/>

*Subscribe to our YouTube channel:*

<http://www.youtube.com/c/NJITInstituteforDataScience>



**September 15th** - Costas Bekas, Citadel Securities



**October 6th** - Gordon Bell, Researcher Emeritus, Microsoft Research Silicon Valley Laboratory



**October 20th** - Jack Dongarra, University of Tennessee



**October 27th** - Larry Smarr, University of California San Diego



**November 3rd** - Laura Hass, University of Massachusetts Amherst

# INSTITUTE FOR DATA SCIENCE DIRECTOR'S OFFICE

---



## **DAVID BADER**

Institute Director  
david.bader@njit.edu



## **SELENNY FABRE**

Business Manager  
selenny.m.fabre@njit.edu



---

[About Us](#) | [Contact Us](#) | [Subscribe](#)

---