



History

In 2013, New York University's Center for Data Science was established to fulfill an ambitious mission: to create a world-leading Data Science training and research facility and to arm researchers and professionals with the tools to harness the power of Big Data. With tremendous support and commitment of resources and faculty lines, in particular through the initiative of then-Vice Provost for Science and Engineering Gerard Ben-Arous and under the leadership of inaugural CDS Director Yann LeCun, CDS started its operation in a compact space at 726 Broadway in Manhattan. CDS also began offering a master's degree program, one of the first degree-granting Data Science programs in the US, enrolling an initial cohort of 29 students.

NYU was selected as one of only three recipients of Moore Sloan Data Science Environment grant in 2013. The funding of \$12.6 million would provide support to mobilize the potential of Data Scientists and big data for basic research and scientific discoveries. The initiative has been instrumental to consolidating the involvement of scholars from a wide range of NYU schools. It has played a critical role in encouraging talented early-career Data Scientists to pursue original research and in invigorating the newly established CDS research environment.

In 2016, CDS moved to the 6th and 7th floors of the historic 60 5th Avenue building to accommodate the expansion, and is now home to 17 joint faculty, 31 PhD students, and approximately 300 master's students. In Fall 2020 CDS will occupy an extra floor in our 60 5th Ave building, expanding to three floors.

About CDS

CDS' interdisciplinary faculty and research scientists have emerged as leading experts in the field of Data Science, straddling methodologies and domains. On the methodologies side, they cover a multitude of areas such as machine learning, optimization and large-scale computation, natural language processing, computational statistics, quantitative methods, responsible data use, and system design for large-scale data. On the domains side they provide expertise in neuroscience, psychology, linguistics, politics, music technology, physics, business analytics, biology, medicine and healthcare.

CDS space is a hotbed of interdisciplinary interaction and innovations. It is a place of lively intellectual exchange and offers widely attended workshops, and recurring seminar series such as "Math and Data," "Text as Data," and weekly lunch seminars. In addition to research, CDS offers rigorous Data Science education programs to PhD, Master's, and Undergraduate students. In 2017, CDS launched one of the first Data Science PhD Programs, and in Fall 2019, introduced innovative undergraduate Data Science major and minor programs. These educational programs aim to equip our students with a critical understanding of both applicable and ethical principles.

Recent Accomplishments

- Received \$3 million NSF Research Traineeship (NRT) award for our PhD program
- CDS Founding Director Yann LeCun received the ACM A.M. Turing Award in 2019
- Welcomed majority female (53%) master's cohort in Fall 2019
- Began the undergraduate Data Science for Everyone (DS4E) course in AY 19-20 and enrolled first cohort of undergraduate majors and minors in spring 2020
- Launched DS3 initiative as a central service for NYU researchers to access expertise in Data Science and statistical methodology

Challenges & Solutions

The center continues to grow and attract new talent, and CDS is proud to be a leader in Data Science at NYU and beyond. Yet with this growth, along with the extraordinary circumstances our community and country has faced this year due to the ongoing COVID-19 outbreak, we would be remiss if we did not acknowledge the challenges we face. Like every burgeoning Data Science institute, our challenges are many and diverse. We identify just a few here which we believe are particularly relevant to the broader academic Data Science community, and the efforts we have undertaken to confront these challenges:

| *Defining and distinguishing the Data Science PhD from the domain PhD*

The interdisciplinary nature of Data Science as a field is our core advantage in distinguishing the Data Science PhD from more traditional domains. Yet many of our students struggle to carve their niche between domains, to define themselves as Data Scientists rather than statisticians, engineers, sociologists, etc. What many of us find to be a challenge could thus be conceived as an opportunity. We **1) emphasize the unique benefits in Data Science** in bringing together diverse disciplines to data-intensive research and work, providing clarity and focus to the rife potential for miscommunication between domains; and **2) hone in on the technical methodology aspect** to Data Science which allows us to exist as a distinct field. We find this adjustment in our messaging to be effective in empowering our students to define and distinguish their training as both necessary and necessarily separate from the domain PhD.

| *Sustaining interest within the wider university community*

With the advent of our new field and centralized academic home, we have welcomed an intense interest from university leadership and from researchers across domains. Yet how are we to translate this initial flighting fancy to a permanent marriage of equals? We are challenged to sustain this interest, harnessing it into long-term support and engagement with our Center, by **1) maintaining consistent dialogue** and **2) offering abundant opportunities** for research support. To that end we offer several opportunities per year for researcher networking, and maintain active and frequent communication channels. We additionally provide direct research support through our capstone and summer research internship programs, and have recently launched the Data Science & Software Services (DS3) initiative to match software engineer and student expertise to research projects. These efforts, among others, have helped us sustain existing enthusiasm and consistently generate successive interest each academic cycle.

| *Managing engagement and balancing efforts of joint faculty*

The CDS model of joint faculty appointments enables us to create a robust web of networks across the university community. But we are cognizant of the increased burden and potential for exhaustion our faculty face, particularly junior faculty. Along with the benefits of this model, we must balance **1) flexibility in our departmental requirements and responsibilities** of faculty and **2) strong administrative relationships** with joint appointment departments in order to more easily facilitate and negotiate shared responsibilities. In this way we effectively communicate to our faculty a support system and safety net as we collaboratively navigate this appointment model.

CDS is the focal point for NYU's universitywide efforts in Data Science