Data Science in Computer Science Education

CRA Snowbird Panel July 2022

David S. Ebert, University of Oklahoma
Michael Franklin, University of Chicago
Magda Bałazinska, University of Washington
Atul Prakash, University of Michigan
NSF Computing in Undergraduate Education (CUE): Revisioning broadening participation pathways

• Transformation
  • Addressing key challenges in undergraduate education (CS+X, updating courses, holistic degree support, inclusive online teaching)

• Pathways
  • Effective pathways involving 2-year colleges (entry points, transitions)

• Mobilize
  • Developing a national vision through convening the CISE community

Proposals due: September 19, 2022
Larger budget requests have a higher expectation for the breadth of impact
Contact: iuse-cue@nsf.gov
2016 CRA Report on Computing Research and the Emerging Field of Data Science

• It highlighted the fact that data science will drive fundamentally new research in computer science and that the computing community has the opportunity to shape the emerging field of data science.

• This panel:

How has the data science growth impacted computer science departments?
Each Panelist Will Answer

- What are you doing?
- What worked well, didn’t work well?
- What are gaps or issues?
Purdue University: What Worked Well

• Campus-wide minor in data science with different requirements per college
• DataMine – living-learning lab/dorm with over 1000 students across campus focusing on data science
Learning

Stackable Ones

One credit hour courses focusing on the three elements that can be "stacked" to create a custom data-science curriculum. These courses will be offered online too. Students can also satisfy the "Data Mind" requirement by taking existing courses in their major that cover any of the three elements. A full list of such courses will be announced soon. Additional opportunities are also being explored for undergraduate and graduate students who want to pursue more in-depth studies.

Machine Learning

We propose to develop a machine learning group in the College of Engineering. We believe that this is an urgent and critical move, for otherwise faculty and students will not see the presence of machine learning in the CoE. We have three objectives: 1) Create new graduate courses on machine learning; 2) Create and sustain a machine learning seminar series and a machine learning club that covers general machine learning topics at both undergraduate and graduate levels; and 3) Develop and maintain a group web page in machine learning in the CoE.

Development of two new graduate courses in Machine Learning is being funded by an Engineering Faculty seed grant to Stanley Chan and Ali el Gamal, assistant professors in Electrical and Computer Engineering. Machine Learning I, which is being first taught in Spring 2019, includes modules that review fundamentals in statistics, and optimization, introduce classification methods, summarize python computing techniques. Machine Learning II will be taught in Fall 2019 and will review advanced statistics techniques, graphical models, and Big Data computing.
University of Oklahoma – What isn’t working that well

- Separate undergrad minors or certificates in 3 colleges
- Coordination of course content and material and short data science intro (1CR) could be better
University of Oklahoma – What is working well!

• Graduate Program – MS and PhD in Data Science with students of all backgrounds
• Campus-wide coordination and growth of convergent data science research!
Why Were We Created?

STRATEGIC VERTICALS

AEROSPACE, DEFENSE AND GLOBAL SECURITY
- International Security Policy
- Advanced Radar & Applications
- Advanced Technologies
- Air Force Sustainment

ENVIRONMENT, ENERGY AND SUSTAINABILITY
- Preserving Biodiversity and Ecosystems to Maintain Resilient Ecosystem Services
- Transforming Energy Systems to Meet the Needs of a Growing Population, Enhance Sustainability, and Limit Environmental

THE FUTURE OF HEALTH
- Eliminating Health Disparities
- Molecules to Medicine
- Predicting, preventing, and responding to emergent pathogenic threats

SOCIETY AND COMMUNITY TRANSFORMATION
- Technology and Society
- Native Nations, Sovereignty, and Partnerships
- Equity and Opportunity

DISC: Advance data-enabled research and foundational data science
Lead: Ebert
Lead: Koch
Lead: Fagg
Lead: Maher

Enabling Core Capabilities
- College Specific Academic Disciplines
- Strategic Plans

Data Science and AI/ML

Outcomes
- New Knowledge
- Enterprise Development
- Impact on Society
- Workforce & Leaders of the Future
Impactful Data Science: Driven by Grand Challenges

Interdisciplinary & Integrative Approach to:

- **Research**
  - Support structured research efforts in Data Science foundations, applications, data-driven discovery

- **Integrated Gov/Business/Industry Teaming with Impact**
  - Research, delivery, and workforce

Co-mingle Researchers, Partners and Educational Ecosystem

- **Promote creative collaboration through co-location**

- **Education – through partnerships**
  - Growing inclusion across campus

---

**Data Science Fundamentals**

- Human-Guided, Trustable AI & ML
- Human-Computer Teaming
- Predictive Analytics
- Collaborative, Data-Driven Decision-Making
- Scalable, High-performance Software and Hardware Architectures

**30+ Faculty**

**Data Science Applications**

- Environmental Science, Digital Humanities, Resilient Communities,
- Advanced Manufacturing, Predictive Maintenance, Engineering,
- Energy, Health/Biomedical,
- 175+ Faculty
Data Science in Computer Science Education

CRA Snowbird Panel July 2022

David S. Ebert, University of Oklahoma
Michael Franklin, University of Chicago
Magda Bałazinska, University of Washington
Atul Prakash, University of Michigan
Where should data science live in the university since it is ubiquitous?
How do CS department chairs leverage the excitement into DS/AI?
Is data science a subset of computer science or a superset or a new field?
Is data science dead and taken over by AI? Ditch data science?
How do we create a shared understanding across campus of data science/AI?
What is the minimum delta between data science and computer science degrees?
Are we done with data science/CS discussion?
What is the role of non-STEM (social science and humanities) in data science?
Data science minor – CS+X similarity
Is retention of DS faculty a problem?