What do Data Scientists do?

Impact from Analyses:

- **Strategic Understands** - In-depth analyses on complex, ambiguous, and very important problems
- **Ad-hoc Analyses** - Brief analyses on simpler, less complex problems
- **Influencing & Implementation** - Communicating with human beings in person and in writing so they LISTEN TO THE DATA

Impact from Execution:

- **Experiments** - Consult on design, interpretation, launch decisions & debugging strange results
- **Metrics Monitoring** - Know the data, the ups & downs
- **Data Engineering** - Logging, tables, dashboards
INSTITUTE NEWS

In 2018-19, the Institute for Applied Computational Science expanded as it welcomed its inaugural cohort of master’s students in Data Science, admitted its sixth cohort of students to the master’s in Computational Science and Engineering program, launched a new core course, *Critical Thinking in Data Science*, and strengthened industry and academic connections. The vision of its founders – to be an intellectual home for faculty and students applying computational methods to major challenges in science and the world – remains central to the institute’s activities.

ENGAGING WITH INDUSTRY

IACS held its bi-annual advisory board meeting in 2019 with faculty and representatives from leading companies and national laboratories who reviewed curriculum, met with students, and shared perspectives on emerging trends and training needs at the master’s level at Harvard. In addition to the advisory board, many other companies and organizations (including Google, TripAdvisor, BASF, the MBTA, Spotify, the Associated Press) are involved with IACS through participation in the capstone research project course, and through talks and a regular IACS seminar series. This year, partners NVIDIA, IBM Research AI, Google, and Microsoft-Azure provided training at the annual IACS Computefest workshops, focusing on Deep Learning. IACS also launched a new education fund for student support at Computefest funded by sponsors and workshop attendees.

COLLABORATIONS AND EVENTS

IACS continued its partnership with MIT’s Institute for Data, Systems, and Society, Stanford University, and Microsoft Research New England, hosting a day-long *Women in Data Science* (WiDS) conference in Kendall Square. More than 250 academic leaders, industry professionals and students attended the event, which was one of more than 150 WiDS events held simultaneously around the world.

The 2019 Computefest Symposium brought together leading researchers doing machine learning at the frontier of computational work in the physical sciences. IACS continues to collaborate with the Harvard Data Science Initiative, the Center for Research in Computation and Society at the John A. Paulson School of Engineering and Applied Sciences, and with departments and schools across Harvard.

SUPPORT FOR PUBLIC SERVICE

IACS helped develop the Public Service Data Science Graduate Fellowship, offered to master’s students in the data science programs at Harvard University and sponsored by the Harvard Data Science Initiative. This fellowship is designed for those who want to explore career paths at not-for-profit and public sector organizations during the summer before their final semesters in the master’s program.

A GROWING COMMUNITY

With nearly 300 alumni of IACS master’s programs as of May 2019, IACS celebrated the fifth reunion of the first class by hosting alumni events in key cities where alumni have clustered: Boston, New York, and San Francisco. Alumni work in technology, finance, data-driven consumer companies, and startups and a few have gone on to doctoral programs.

Catherine A. Chute
Executive Director, IACS
SEAS Assistant Dean for Professional Programs

Efthimios Kaxiras
Faculty Director, IACS
John Hasbrouck Van Vleck Professor of Pure and Applied Physics
NUMBER OF MASTER’S STUDENTS IN IACS PROGRAMS

Applications vs. Enrollment

- CSE Applications
- CSE Offered Admission
- CSE Enrolled
- Data Science Applications
- Data Science Offered Admission
- Data Science Enrolled
WHERE STUDENTS GO AFTER GRADUATION

CSE graduates go on to work across a variety of industry sectors or choose to pursue further studies at leading graduate business or doctoral programs. IACS collaborates with the Harvard Office for Career Services to expose students to a diverse range of companies — through the Data Analytics and Technology Fair (attended by nearly 100 employers), and individual company tech talks or tech treks.

INFORMATION FROM THE GRADUATING CLASSES OF 2013-2018

TECHNOLOGY

61 STUDENTS

Afiniti · Akamai · Alibaba · Amazon · Bose · Facebook · GCP Applied Technology · Google · Houzz · Hubspot · Hudl · Lyft · Microsoft · PlusAI · QR · Salesforce · SquareSpace · Tencent · TripAdvisor · Uber · Wayfair · Yelp

INVESTMENT / FINANCE

31 STUDENTS

Acacia Global Investors · Accadian Asset Management · Arrowstreet Capital · Axovant Sciences · Balyasny Asset Management · Bank of America Merrill Lynch · Bloomberg · Bridgewater Associates · Capital One · China Asset Management · Citadel · Goldman Sachs · JP Morgan · Man Numeric · Maquarie Group · Moneyion · Morgan Stanley · OnDeck · The Thasos Group · Vitu Financial · Weiss Asset Management · Wellington Management

FURTHER GRADUATE STUDIES

17 STUDENTS

Applied Physics PhD · Architecture Technology PhD · Biostatistics PhD · Computer Science PhD · MBA · MD/PhD · Mechanical Engineering PhD · Statistics PhD

ACADEMIA

7 STUDENTS

Harvard Ash Center · Harvard Center for Astrophysics · Harvard Center for Clinical Data Science · Harvard IACS · Johns Hopkins Applied Physics Lab · MIT research · National Cheung Kung University

GOVT / MILITARY / NATIONAL LABS

6 STUDENTS

GovTech Singapore · MIT Lincoln Laboratory · Singapore Military · U.S. Coast Guard · U.S. Navy

ADVERTISING / MARKETING

6 STUDENTS

Accenture Interactive · Pypestream · Tribe Dynamics · Yieldmo

OTHER

15 STUDENTS

BCF Associates d’Affairs · BookXChange · Legendary Entertainment · McKinsey & Company

TYPICAL JOB TITLES: DATA SCIENTIST · DATA ENGINEER · SOFTWARE ENGINEER · PRODUCT MANAGER · DATA ANALYST · QUANTITATIVE STRATEGIST

STUDENTS REPORTED STARTING SALARIES BETWEEN $80,000 – $140,000.
Open to Ph.D. students in the Graduate School of Arts and Sciences, the Computational Science and Engineering (CSE) Secondary Field (Harvard’s term for a minor) equips students across disciplines with an understanding of rigorous computational methods for approaching scientific questions. In 2019 six Ph.D. students graduated with their Secondary Field in CSE. Starting in the fall, IACS will offer a Secondary Field in Data Science.

**SECONDARY FIELD PROGRAM ATTRACTS STUDENTS FROM A WIDE VARIETY OF DISCIPLINES**

MAJOR FIELD OF STUDY FOR CSE SECONDARY FIELD STUDENTS

**Statistics | 5.5%**
Biostatistics • Population Health Sciences

**Social Sciences | 12.5%**
Economics • Health Policy • Sociology

**SEAS | 32%**
Applied Math • Applied Physics • Engineering Sciences • Environmental Science and Engineering

**Sciences | 50%**
Astronomy • Biological and Biomedical Sciences • Biophysics • Chemical Physics • Chemistry and Chemical Biology • Earth and Planetary Sciences • Molecular and Cellular Biology • Organismic and Evolutionary Biology • Physics • Systems Biology

Total number of secondary field students = 56
Offered for the first time this past Spring, Critical Thinking in Data Science examines the wide-ranging impact that data science has on the world and teaches students to think critically about issues of fairness, privacy, ethics, and bias. The curriculum includes topics on algorithmic fairness, data privacy and security, ethical frameworks, and experimental and product design, and guest lectures are given by speakers from Computer Science and Statistics as well as the Kennedy School, Law School, and Philosophy department. At the completion of the course, students will be prepared to recognize the ethical considerations they will encounter as data scientists and have a better understanding of the obligations to themselves and society at large when addressing them.
Students in the capstone course apply skills such as machine learning, statistics, data management, and visualization to solve real-world problems. Now part of the core research requirement for the data science master’s program, the capstone research project course has been developed, managed and taught during the last four years by IACS scientific program director Pavlos Protopapas.

**Alpitour - Predicting Travel Trends From Italy to Cuba**

**Students:** Yuting Kou, Stephen Slater, Eleonora Cappuccio, and Tommaso Scarlatti

Italian tourism company **Alpitour** is a large international travel group. Students in the Capstone course were tasked with predicting the demand for travel from Italy to Cuba for each week in the following year. The group developed and delivered a software pipeline to be used by **Alpitour** software engineers and data scientists.

**Austin Pets Alive! - Tinder for Dogs**

**Students:** William Burke, Jaemin Cheun, Eric Choi, Rachel Moon, and Will Claybaugh

**Austin Pets Alive! (APA!)** is an Austin, Texas-based animal shelter that has led the movement to make Austin the largest no-kill city in the country. Students in the Capstone course built a Tinder for Dogs web application for potential dog adopters that recommends dogs based on biographical information entered as well as preferences as determined by swipes.
Thanks to the generosity of an anonymous donor, IACS has been able to offer fellowships to a small group of students in the Computational Science and Engineering (CSE) and Data Science programs. For academic year 2018–2019, two students were offered scholarships.

**Camilo Fosco, M.E. ’19**
While at IACS, Camilo’s research focused on the guiding processes that affect our perception of the visual world. His thesis explored the subfields of human perception and cognition and provided insights on the connection between attentive processes and human cognition within the context of vision. In the fall, Camilo will begin a Ph.D. program at MIT.

**Srivatsan Srinivasan, S.M. ’19**
Srivatsan’s research focuses on reinforcement learning, imitation, transfer learning, and bayesian deep learning with applications to healthcare. Under the guidance of Professor Finale Doshi-Velez, Srivatsan’s research thesis was on reinforcement learning for healthcare. This summer he will join Google DeepMind as a Research Engineer.

**IACS Education Fund**
Established in 2018, the IACS Education Fund was created to support emerging leaders in the fields of Data Science and Computational Science and Engineering. Thanks to generous donations from Microsoft and Bertelsmann, an anonymous donor, and proceeds from ComputeFest workshop registrations, IACS will be able to award fellowships to four students for the 2019-2020 academic year.
**Chris Tanner**

In 2019, Chris completed his Ph.D. from Brown University, with a focus on Natural Language Processing. Specifically, he is interested in discourse and semantics, and he is always interested in collaborating with and mentoring students to tackle new, interesting projects. Chris teaches Data Science (CS109a and CS109b), along with the Capstone Project course (AC297r). Chris’s other passions include traveling, photography, sewing/designing backpacks, and going on challenging hikes.

**David Sondak**

David received his Ph.D. from Rensselaer Polytechnic Institute in Aeronautical Engineering in 2013. His research focuses on data-driven models and algorithms for physical systems with an emphasis on fluid mechanics. At IACS, David teaches CS207 (Systems Development for Computational Science), AC290r (Extreme Computing), and AC297r (the Capstone Project Course). He also supervises master’s students theses. In addition to his research, David is passionate about dogs and their welfare. In his free time he enjoys playing soccer and taking hikes with his dog and wife.

**Rahul Dave**

Rahul Dave completed his lecturership at IACS during the 2019-20 academic year, teaching AM 207 Advanced Scientific Computing: Stochastic Methods for Data Analysis, Inference and Optimization and contributing to the data science core curriculum. He previously taught CS207 (Systems Development for Computational Science) and co-taught CS109 (Data Science). He has been affiliated with the Harvard Smithsonian Center for Astrophysics since 2009 as a computational scientist. He holds a Ph.D. in Physics from the University of Pennsylvania and a Bachelor’s in Physics from the University of Bombay. In addition to lecturing, Rahul presents regularly on machine learning and systems development and enjoys the outdoors with his family.
IACS PUBLIC PROGRAMS

WOMEN IN DATA SCIENCE AND COMPUTEFEST
IACS is proud to have partnered with MIT, the Microsoft NERD Center, and Stanford University to bring the Women in Data Science (WiDS) Conference to Cambridge. Started at Stanford in 2015, WiDS is a global initiative designed to inspire and educate data scientists worldwide, regardless of gender, and support women in the field. This year’s conference took place on Monday, March 4th, and brought together more than 250 participants to hear inspiring female speakers from industry and academia share their breakthrough work.

WiDS Datathon
Hosted by Weiwei Pan, Harvard Institute for Applied Computational Science, and Anastasiya Belayeva, MIT Institute for Data Systems and Society, the WiDS conference was preceded by a datathon workshop where more than 50 participants, of all genders, were given a dataset of high-resolution satellite imagery and tasked with creating a model that could detect palm oil plantations. This predictive analysis challenge was designed to build awareness about oil palm plantations and deforestation. Participants formed teams of 3-4 people and competed with groups from around the world; the top four winners were announced at the conference.
Each January, IACS hosts ComputeFest, a four-day program of knowledge and skill-building activities in data science and computational science and engineering. This year’s event included an all-day symposium followed by three days of Deep Learning workshops.

**Symposium**
The 8th annual IACS symposium: *Data Science at the Frontier of Discovery: Machine Learning in the Physical World* welcomed nearly 300 attendees to hear speakers from SEAS, Cornell University, Cambridge University, and Google share their work on topics ranging from Molecular Machine Learning to Scalable Flood Forecasting, and Inverse Problems in Cosmology. This year’s event was organized by Efthimios Kaxiras, Michael Brenner, and Pavlos Protopapas.

**Workshops**
Taught by IACS students, alumni, and industry presenters: NVIDIA, Google What-If, IBM Research AI, and Microsoft Azure, this year’s workshops covered advanced topics on computer vision, interpretability and fairness in data science, and natural language processing.
Since the Board’s inception in 2010, members have played an important role in shaping IACS graduate programs. On May 16, 2019, the Board convened for the fifth time, beginning with a curricular overview highlighting project-based learning, new modules from leading Harvard groups and labs in the data science core courses, and industry involvement in the capstone project course.

Highlights of the program included a presentation by Jim Waldo of his new course, Critical Thinking in Data Science, a requirement for the master’s in data science program that integrates technology, policy, and ethics, building upon the Computer Science area’s new embedded ethics program; and a presentation by Salil Vadhan of his new Applied Differential Privacy elective course.

During breakout sessions, Advisory Board members suggested using the project-based curriculum to teach communications and leadership skills, and to incorporate best practices in diversity and inclusion. Student panelists discussed the program’s strengths: team project work, the diversity of student backgrounds, the value of the capstone project course, and the program’s overall rigor and high academic quality.

Overall, a few key themes emerged:

- Engaging in real world experience is essential for the students — through capstone projects, an entrepreneur-in-residence program, and connecting with practitioners.
- The small group project work is fundamental to developing technology leadership skills and incorporating diversity and inclusion.
- Financial aid/scholarships can attract diverse students and encourage government service.
**SEMINAR SERIES**

The IACS seminar series is a forum for thought leaders from academia, industry, and government to share their research on innovative computational and data science topics and methodologies.

### 2018-2019 Seminars

<table>
<thead>
<tr>
<th>Topic</th>
<th>Speaker</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning to Rank an Assortment of Products</td>
<td>Kris Ferreira</td>
<td>Harvard Business School</td>
</tr>
<tr>
<td>Fluid Mechanics with Turbulence, Reduced Models, and Machine Learning</td>
<td>David Sondak</td>
<td>Harvard, IACS</td>
</tr>
<tr>
<td>Computational Perception with Applications to Graphic Design</td>
<td>Zoya Bylinskii</td>
<td>Adobe</td>
</tr>
<tr>
<td>Data Science for Game Development</td>
<td>Dean Wyattte</td>
<td>Activision</td>
</tr>
<tr>
<td>Machine Learning in the Healthcare Enterprise</td>
<td>Mark H. Michalski</td>
<td>MGH &amp; BWH Center for Clinical Data Science</td>
</tr>
<tr>
<td>Bottlenecks, Representations, and Fairness: Information-Theoretic Tools For Machine Learning</td>
<td>Flavio P. Calmon</td>
<td>Harvard University</td>
</tr>
<tr>
<td>Machine Learning for Materials Discovery</td>
<td>Julia Ling</td>
<td>Citrine Informatics</td>
</tr>
<tr>
<td>Predictive Modeling of Aperiodic Astrophysical Behavior</td>
<td>Matthew Graham</td>
<td>CalTech</td>
</tr>
<tr>
<td>The Dedalus Project: A Flexible Approach To Accurately Solving Pdes, With Applications In Stellar Astrophysics</td>
<td>Benjamin Brown</td>
<td>University of Colorado</td>
</tr>
<tr>
<td>Forecasting Airport Transfer Passenger Flow Using Real-time Data And Machine Learning</td>
<td>Yael Grushka-Cockayne</td>
<td>Harvard Business School</td>
</tr>
<tr>
<td>Neural Modeling And Differential Equations</td>
<td>Isaac Lagaris</td>
<td>University of Ioannina</td>
</tr>
</tbody>
</table>


ADMINISTRATIVE AND TEACHING STAFF  Efthimios Kaxiras John Hasbrouck Van Vleck Professor of Pure and Applied Physics  Cathy Chute Executive Director  Pavlos Protopapas Scientific Program Director and Lecturer  Daniel Weinstock Associate Director of Graduate Studies and Lecturer  Lori Ray Program Manager  Natasha Baker Administrative Coordinator  Ignacio Lorente Visiting Professor  Sauro Succi Visiting Professor  David Sondak Lecturer  Cecilia Garraffo Research Associate  Marios Matthaiakis Research Associate  Weiewei Pan Postdoctoral Researcher 

RESEARCH STUDENTS AND VISITORS  Nicolás Astorga Student Researcher  Alessandro Bianchi Student Researcher  Marco DiGiovanni Student Researcher  Brian Hayes Associate  Javier Machin Matos Student Researcher  Simone Melchionna Visiting Scholar  Georgios Neofotistos Visiting Scholar  Manuel Pérez Student Researcher  Karim Pichara Associate  Rosalind Reid Fellow  Mauricio Santillana Associate  Isaac Slavitt Fellow  Georgios Tritsaris Visiting Scholar  Moreno Vendra Student Researcher  Alexander Wissner-Gross Associate
IACS TEAM

Tim Kaxiras
Faculty Director

Cathy Chute
Executive Director

Pavlos Protopapas
Lecturer and Scientific Program Director

Daniel Weinstock
Associate Director of Graduate Studies in Applied Computation

Lori Ray
Program Manager

Natasha Baker
Administrative Coordinator

Cecilia Garraffo
Research Associate
Cecilia’s work focuses on using deep learning techniques to address astrophysical problems.

Marios Matthaiakis
Research Associate
Marios’ research involves designing deep neural network architectures for implementation in Physics.

Weiwei Pan
Postdoctoral Researcher
Weiwei works on building machine learning models, particularly deep Bayesian models that satisfy domain-specific desiderata with a focus on health-care related applications.

IACS INCOMING CLASS OF 2018-2019