

October 27, 2020

PROFESSOR RAJESH GUPTA
Halicioğlu Data Science Institute

SUBJECT: Proposed MS in Data Science

At its October 12, 2020 meeting, the Graduate Council approved the proposal to establish a graduate degree program of study leading to a Master of Science (M.S.) in Data Science. The Graduate Council will forward the proposal for placement on an upcoming Representative Assembly agenda.

Please note that proposers may not accept applications to the program or admit students until systemwide review of the proposal is complete and the UC Office of the President has issued a final outcome.

Sincerely,

Lynn Russell, Chair
Graduate Council

cc: M. Allen
J. Antony
S. Constable
R. Continetti
B. Cowan
T. Javidi
C. Lyons
J. Morgan
K. Ng
R. Rodriguez
D. Salmon
Y. Wollman

A Proposal for a Program of Graduate Study in Data Science leading to a degree in **Master of Science in Data Science (MS/DS)**

By:

MS Program Committee, HDSI

Yoav Freund, Shankar Subramaniam, Virginia de Sa, Ery Arias-Castro, Armin Schwartzman, Dimitris Politis, Ilkay Altintas, Rayan Saab.

Contacts:

Academic:

Rajesh K. Gupta

Director, Halicioğlu Data Science Institute (HDSI)

(858) 822-4391

rgupta@ucsd.edu

Administrative:

Yvonne Wollman

Student Affairs Manager

Halicioğlu Data Science Institute (HDSI)

(858) 246-5427

ywollmann@ucsd.edu

Version History:

March 6, 2020: Version 1.0 submitted for review by graduate division.

April 12, 2020: Version 2.0 submitted to academic senate and revised administrative review.

May 22, 2020: Version 3.0 submitted to academic senate after administrative revisions.

Sep 21, 2020: Version 4.0 submitted to academic senate graduate council review

This version: Sep 21, 2020.

Online Link: <https://bit.ly/HDSI-MS>

A Proposal for a Program of Graduate Study in Data Science leading to a degree in Master of Science in Data Science (MS/DS)

Executive Summary

The Halicioğlu Data Science Institute, an academic unit at UC San Diego, proposes a new master's degree program in "Data Science" (MS/DS). HDSI's first graduate degree program aims to serve the need for advanced graduate studies in the area of Data Science, a field in which HDSI currently offers a well-received Bachelor of Science degree as a part of its academic mission "to promote a unified campus-wide approach to research and teaching in Data Science". The field of Data Science spans mathematical models, computational methods and analysis tools for navigating and understanding data and applying these skills to a broad and emerging range of application domains. A whole range of industries – from drug discovery to healthcare management, from manufacturing to enterprise business processes as well as government organizations – are creating demand for data scientists with a skill set that enables them to create mathematical models of data, identify trends and patterns using suitable algorithms and present the results in effective manners. The target systems can be, for example, biological (e.g., clinical data from cancer patients), physical (e.g., transportation networks), social (e.g., social networks) or cyber-physical (e.g., smart grids). In all these cases, there is a combination of core knowledge in information processing coupled with the skills to abstract, build and test predictive and descriptive models that must be taught and learnt in the context of an application domain. These application areas are in many domains served by Engineering, Physical Sciences, Social Sciences, Health & Life Sciences, and Arts & Humanities. Thus, the goal of the program is to teach students knowledge and skills required to be successful at performing data driven tasks, and lay the foundation for future researchers who can expand the boundaries of knowledge in Data Science itself.

To achieve these goals, the graduate program is structured as a total of 12 courses consisting of 4 *foundational* courses (to be drawn from an offering of six courses to account for diversity of background of students entering the graduate program), 5 *core* courses and 3 *domain-elective* courses. Successful completion of the MS program requires either a *thesis*, or a *course-directed comprehensive examination* option. The proposal describes in depth both options. The implementation plan is designed to launch the program in AY 2021 with an initial enrollment of 15-20 students. The program structure allows for multiple specialization track areas designed to catalyze and support corresponding specializations of existing MS programs in departments identified in this proposal. No new resources are requested by this proposal.

Table of Contents

Executive Summary	2
1. Introduction	4
A. Goals and Objectives of the Program	4
B. Historical Development of the Field	8
C. Rationale and Justification	9
D. Timetable for Development of the Program	10
E. Relationship of the Proposed Program to Existing Programs on Campus	11
F. Contributions to Diversity	12
G. Relationship of the Proposed Program with Other UC Campuses	13
H. Program Administration and Resource Planning	14
I. Plan for Evaluation of the Program	16
2. Program	18
A. Undergraduate Preparation for Admission into MS/DS	18
B. Admission Requirements	19
C. Foreign Language	19
D. Overview of the Proposed Masters Program	19
E. Plan	19
F. Unit Requirements	19
G. Structure of the Proposed Masters Program: Required and Recommended Courses	19
H. Thesis and Comprehensive Examination Requirements	28
I. Field Examinations	29
J. Qualifying Examinations	30
K. Thesis and/or Dissertation	30
L. Final Examination	30
M. Explanation of Special Requirements Over and Above Graduate Division Minimum Requirements	30
N. Relationship to Master's and Doctor's Programs	30
O. Special Preparation for Careers in Teaching	30
P. Sample Program	30
Q. Normative Time from Matriculation to Degree	31
3. Projected Need	32
A. Student Demand for the Program	32
B. Opportunities for Placement of Graduates	33
C. Importance to the Discipline	34
D. Ways in which the program will meet the needs of the society	34
E. Relationship of the program to research and/or professional interests of the faculty	34
F. Program Differentiation	34
4. Faculty	36
5. Courses	42
Appendix A: Listing of Research Areas	44
Appendix B: Letters of Support Received	46
Appendix C: Catalogue Copy Description [Preliminary Draft]	47
HDSI Bylaws	51
Faculty Biography	
eCourse Forms	

1. Introduction

A. Goals and Objectives of the Program

The proposed program is a brand-new degree program offered by the Halicioğlu Data Science Institute (HDSI) in collaboration with various academic units on the UC San Diego campus. This proposal is followed closely by another proposal on a doctoral degree program in Data Science being filed by HDSI. The goal of this Master's Program is to attract top students from around the world to the emerging field of data science to pursue the proposed M.S. degree that provides them with the necessary knowledge and skills to pursue a career in Data Science for industry, civil services or academia. To achieve this goal we outline here a broadly and equitably-accessible program of graduate study that clearly articulates core knowledge and skills to be expected from our graduates while ensuring success of students drawn from diverse educational backgrounds via well-articulated pathways through existing and new courses as well as on-ramp courses with financial support necessary to ensure a diverse talent pool consistent with aspirational goals of UC San Diego as an academic institution.

The educational objectives of this program are to teach students knowledge and skills that enable them to (a) collect raw data from various sources and convert this raw data into a curated form suitable for computational modeling and analysis (e.g., its use in designing experiments); (b) understand learning algorithms and how to appropriately use them in targeted domains such as in business, health etc. (e.g., in developing effective optimization methods); (c) interpret the results of these algorithms and iteratively drill down into the data, perform analysis, visualize results and carry out scientific enquiry appropriate for the targeted domains.¹

A successful execution of the proposed program also induces a number of imperatives for the Institute that are described later in this document: (a) engage partner academic units for maximum exposure of potential domain experts to graduate student training; (b) successfully recruit faculty into HDSI to create teaching capacity for timely graduation of students; (c) provide sufficient resources to ensure continued development of new and exciting (elective) courses that consistently draws new Data Science talent to HDSI and to the proposed program; (d) establish necessary advising and counseling services for the graduating students to appropriately guide them into post-graduation careers.

B. Historical Development of the Field

Scientific and engineering advances thus far have given us a better understanding of the physical world, material and structural properties and their use in accomplishing primarily physical work

¹ These conclusions have been arrived at through discussions within the HDSI faculty council informed by national debate on this subject organized in a series of meetings by the National Academy of Sciences, Division of Engineering and Physical Sciences under the "Roundtable on Data Science Post-Secondary Education", 2016-17.

more effectively and efficiently. These advances through new measurements and models have also given us insights into the living world and the processes of life, mind and the intellect. In doing so, these advances are incorporating human knowledge from social sciences, humanities, natural and life sciences into a greater understanding of us and the world around us.

Data – collected or synthesized – is the primary means of such knowledge exploration and integration. Historically, data analysis has been a domain of Statistics, a field that has reached across multiple centuries. Growth of scientific enquiry especially through the eighteenth century post-Napoleon period of quantitative scientific discovery relied upon calculus and probability to understand measurement data. Statistical analysis spread widely to many areas of human enquiry, in particular, areas of social sciences such as economics, clinical psychology etc. These efforts contributed to significant growth in statistics.

Statistics departments are now common in most universities. At UC San Diego, while there is no department of Statistics, statistics faculty are part of the Department of Mathematics on the General Campus, and also in the Division of Biostatistics in the School of Medicine. Many of these statisticians are also active faculty with partial appointments in HDSI. As Statistics matured with strong foundational results and practical methods influencing many application domains, more recent advances in computing hardware, software, engineering of sensory devices, etc. enabled not only volumetric growth in data but also in computational means to handle such data. Recent advances in algorithmic processing, machine learning, have significantly advanced computational means for data processing. Early efforts in defining computational means of handling large data sets and streams placed a new field of Data Science at the intersection of statistics and computer science² while others characterized it as a growth area of Statistics with strong applications focus.^{3,4}

While strong footprints of computer science, mathematics and statistics can be seen in its origins, Data Science has emerged as a new discipline in its own right at the intersection of multiple knowledge domains that together define either the core problems of the sciences and society, or fundamental theories and underlying methods and tools to solve these problems. Many of these problems concern reasoning, spanning intellect and knowledge domains that are assisted by computing machines, thus referred to as machine intelligence or artificial intelligence (AI). An independent and rich tradition in signal processing, information theory, detection and estimation theory from Electrical Engineering has contributed significantly to modern automation methods in AI. While AI has caught the imagination of computer scientists and mathematicians since the early years of computing machines nearly half a century ago, technological advances have only recently made it possible to realize answers to some of the pressing questions such as:

² David Blei, Padhraic Smyth, “Science and Data Science”, PNAS August 2017.

³ Bin Yu, “Let us own Data Science”, IMS Presidential Address, October 2014.

⁴ David Donoho, “50 Years of Data Science,” Journal of Computational and Graphical Statistics, Dec 2017

- *How do we relieve humans from tedious repetitive tasks without removing their agency?*
- *How can we incorporate machine intelligence into decision processes that are currently purely human, and thus transition from purely human decision making to combined human/machine decision making?*

As answers to such questions come to a realization in the form of new software and systems in specific application domains, entirely new sets of sophisticated questions are beginning to be asked. For instance,

- *How do we integrate algorithmic decision making into political, social, and economic institutions in a way that anticipates how the algorithm itself might change the incentives/behavior of individuals and cause negative and positive feedback loops?*
- *What are the ethical, moral and business considerations when an algorithm learns by observations and produces new products and services? Who are the ultimate beneficiaries of these intellectual or material products: for instance in a healthcare setting, is it the patient or the doctor being observed, the business creating new services or the machine itself?*

The list of such questions is mind-boggling and touches pretty much every area of human enquiry. As an academic institution, the foremost questions we struggle with are how knowledge advances in the emerging domain will be achieved and how will we create a talent pool for the emerging area? As mentioned earlier, the two academic areas that have made the most early advances in methods, tools and systems to perform such data analysis are statistics and computer science, talent that is typically found in engineering departments of electrical engineering and computer science (and cognitive science in the context of UC San Diego), as well as mathematics, and statistics in natural sciences. New sensing/data-collection devices and computing/data systems have also brought together these domains, enabling practitioners to avoid making strong assumptions about the nature of the process that generates the data and use real-life datasets instead. Thus, analysis methods can be directly interfaced with real-life systems to actually capture and analyze real-data (and sometimes in real-time as well). Without the axioms underlying data generation processes, the mathematics becomes exceedingly complex to arrive at robust answers analytically. However, it is precisely in these circumstances where computing steps in with computational models and solutions to vexing mathematical problems. Yet, neither -- mathematical analysis or computational models -- can provide generic answers as problem solving methods (and tools) that applications can use, because in the absence of a broadly applicable axiomatic framework, the very success of these methods and tools depends upon the structure, dynamics and meaning of the actual data. In this environment, to make real progress and demonstrate impact, it is essential to work closely with the scientists, engineers and

social scientists -- the “domain experts” -- before the real problems are understood, articulated and solutions devised to make an impact.

Fortunately, UC San Diego provides a rich tapestry of domain experts starting with perhaps one of the most complex of application domains -- the human brain and mind -- and spread across the triumvirate of general campus, health and marine sciences. Over the past three years, starting one full year before the founding of the institute, Institute leadership has engaged deeply with a broad community of nearly 500 researchers across the campus through many meetings in small group settings. These efforts yielded a core group of 186 founding faculty who came together and organized their ideas in data science into 44 clusters⁵. We then matched these clusters against core advances being pursued by data science researchers across the nation. HDSI faculty council consists of 37 faculty members formally appointed into the HDSI including 15 new and 8 existing faculty who teach primarily in the classes offered by the Data Science major and minor (See Section 1-G). Over the past year, the HDSI faculty council worked closely together to nominate research challenges, grand-challenges that are central to Data Science as a field. A compilation of all these efforts reduced core areas of research into following seven research themes that form the scope of the HDSI academic programs and continues to drive our faculty recruiting strategy listed below. A listing of research areas within each theme is provided in Appendix A.

Data Science Core Theme	Brief Description
1. Artificial Intelligence	AI is about automating the decision processes, augmenting, complementing or amplifying decision making means. Challenging problem areas are related to finding and learning good representations of knowledge with connections to human cognition.
2. Machine Learning	Computing machines -- from neuromorphic to cloud processing -- have energized algorithms and CS theory that enable computing systems to learn from data, create examples, counter-examples. ML covers various learning methods (reinforcement, transfer, resource-constrained, deep learning), architectural acceleration and algorithms for game theoretic setups, natural language processing, etc.
3. Data Infrastructure	This theme covers the entire gamut of machines and systems that enable us to curate, organize, visualize and navigate large datasets, identify structure in such data; design, deployment and security of systems and their software stack including new programming paradigms, languages and methods.

⁵ <https://datascience.ucsd.edu/research/>

4. Mathematical Foundations	Mathematical foundations span areas of probability theory, statistics and computer science that are currently addressing some of the pressing challenges of data science such as causal inference, non-parametric data analysis, multiple hypotheses testing and submodular optimization etc.
5. Digital Humanities	Digital humanities is an umbrella term that in UCSD context spans both social sciences, arts and humanities. The research topics include privacy, public policy, ethics, computational social science, computational linguistics, and philosophy of information.
6. Systems and Applications	A large and heterogenous group of topics from algorithms and demonstrable systems, their use in specific domains from medical signal processing, economics, geospatial data systems to political and economic systems as well as applications in cyber-physical systems and robotics.
7. Scientific Discovery	Data-driven scientific advances through new instruments and analytics, especially for high-throughput biology and chemistry that enable us to understand the natural world as well as life processes.

C. Rationale and Justification

As a core subject area of the Halicioğlu Data Science Institute, the Data Science graduate program is a natural evolution of the undergraduate program in Data Science and a key vehicle for the academic research conducted by its full and jointly-appointed faculty members.

Besides being the academic home for Data Science at HDSI, a practical reason for launching of this program now by the HDSI is to satisfy the growing demand for the graduate program in data science both internally as well as externally: in a survey of our students graduating from Data Science major, roughly a third of them have indicated their interest in a graduate degree in Data Science program. In 2019, nearly a half (1571) of over 3000 applicants to various graduate degree programs at UC San Diego who indicated interest in Machine Learning and Artificial Intelligence related topics in Electrical Engineering, Computer Science and Cognitive Sciences, directly indicated their interest in Data Science programs at HDSI.

In short, the demand for a graduate training in Data Science is high and continues to grow. Further, the program not only serves the HDSI mission of educating talent in the area of Data Science, but also serves as a vehicle for continued engagement and proliferation of Data Science training across various graduate programs through new foundational, core and elective course offerings that engage domain experts into the field of Data Science. The program provides an excellent means to create new educational opportunities for students, especially for underserved

and economically-disadvantaged student populations who can benefit from graduate scholarships offered by HDSI as a part of its core endowment-supported activities. Among the core endowment-supported activities are: graduate and undergraduate scholarships, post-doctoral and visiting scholar programs, seminars and outreach programs. To be specific, the endowment supports a full-time academic coordinator professional who is tasked specifically with responsibilities to broaden participation of URM talent into HDSI’s educational programs (40%), as well as in creating pathways for successful and timely graduation of students drawn from diverse educational backgrounds including transfer students (40%). In addition, the endowment funds support dedicated professionals, computing and data science research specialists, with the responsibility to support HDSI faculty and researcher needs for optimized and easily accessible cloud computing and storage solutions.

D. Timetable for Development of the Program

The HDSI faculty council kicked-off discussion on the graduate program in Summer 2019 with a formal presentation and resolution at the two-day long Faculty Retreat on September 16-17, 2019. HDSI formed a faculty committee for defining and developing the MS program in Fall 2019 headed by Professor Yoav Freund and Dimitris Politis. The table below shows an expected timeline of the program review and approval process. A WSCUC screening request will be submitted after UCSD Graduate Council review and approval concurrent to its submission for the CCGA review. We plan to announce the program after approvals in early 2021. We anticipate launching this program beginning Fall 2021 starting with admission of internal Data Science Major applications (who normally face a later deadline in May), followed by general admission deadline in Fall 2021. Initial enrollment is estimated to be 15-20 students with approximately 50 new students per year at steady-state. We plan to conduct a review of the program and outcomes after the first three years of operation as a part of our re-assessment of capacity and any enrollment changes.

Needs assessment and faculty discussions in HDSI	Summer’19-Winter’20
Administrative Review and Routing	Spring 2020 (April)
Proposal submitted for UCSD Graduate Council	Spring 2020 (May)
Revised proposal submitted	Summer 2020 (September)
UCSD proposal submission to CCGA	Summer 2020 (October)
CCGA approval received	Early Fall 2020 (November)
UCOP approval received	Winter 2021 (January)

Program announced open for admissions (DSC Majors)	Winter 2021 (January)
Admission of first class announced	Spring 2021 (April)
Orientation and student advising	Summer 2021
Program offered and courses begin	Fall 2021

E. Relationship of the Proposed Program to Existing Programs on Campus

At its emergence, Data Science is seen as a *transdisciplinary field*, that is, a field with intellectually overlapping sub-fields in the departments of computer science, electrical engineering, mathematics and cognitive science at UC San Diego. As pointed out by the letters of support from the chairs of Computer Science and Engineering, Cognitive Science and Mathematics, the founding departments for the Data Science undergraduate program, this overlap is seen as a strength that makes the proposed degree program “*unique in bringing together the very best UCSD has to offer educationally..*” as articulated by Douglas Nitz, Chair of Cognitive Science in the Division of Social Science at UC San Diego.

Existing MS Degrees: There is a “Data Science and Machine Learning” concentration, one of twelve concentrations, in the Electrical Engineering MS program offered by the Electrical and Computer Engineering (ECE) Department. Among related programs that partially cover some of the subject areas of the MS/DS program, the Computer Science and Engineering (CSE) Department has a MS program in Computer Science which offers a concentration in machine learning as a depth area. The CSE department also has graphics and vision as a depth area in its graduate program as the areas of artificial intelligence, graphics and vision are among the fifteen research groups in CSE.⁶ Indeed, this overlap in research areas enabled the CSE department to help us put together a Data Science program in partnership with the departments of Mathematics and Cognitive Science, as noted in the support letter from Chair Dean Tullsen.

The Department of Mathematics offers an MS degree in Statistics with courses on Statistical learning and inference. Biostatistics in the School of Medicine has recently launched an MS in Biostatistics program, which includes required courses in statistical learning, including machine learning, inference and data analysis applied to biomedical and public health data. Finally, there is an MS program in Business Analytics from the Rady School of Business. As the attached letter of Dean Lisa Ordóñez exemplifies, the proposed MS/DS is complementary to these degree programs “by offering students in these programs new graduate courses that are created as part

⁶ <https://cse.ucsd.edu/faculty-research>

of this proposal.” Dr. Ordóñez identifies specifically our core courses (DSC 243 and DSC 250) that we believe will be of interest to other programs as well for similar reasons.

Existing Professional Masters of Advanced Studies: The CSE department in collaboration with the San Diego Supercomputer Center offers an on-site professional Masters of Advanced Study (MAS) program in Data Science Engineering. This is a small program of 30-40 students per year with a relatively high price-point designed for the working professional with classes held on Fridays and Saturdays. In a separate, concurrent proposal, we are also proposing to create a new Online Masters of Data Science (OMDS) program in cooperation with the Computer Science and Engineering department. OMDS is a professional degree program aimed at *working professionals* who want to continue their education. Thus the target audience is entirely different with little or no overlap in the set of students being recruited. On the other hand, we do see OMDS synergistic to HDSI efforts in that the OMDS can provide opportunities to students coming from very different backgrounds. To be specific, the existing and planned micromaster online courses in OMDS individually provide an alternative self-paced preparation pathway to students interested in the proposed MS in Data Science program.

The proposed program is directly and entirely dedicated to Data Sciences area and differs from existing programs into two material ways: in breadth of the student population it serves and in the scope of the transdisciplinary training it provides as discussed below:

(a) The proposed degree is envisioned as an advanced training for students drawn from a wide variety of backgrounds in their undergraduate education to serve a *different* and *diverse* group of learners interested in Data Science. This is in contrast to existing programs that either target a different population of students or focus on subject areas specific to their domain. For instance, the MAS/DSE program offered by JSOE is targeted for working engineers for weekend classes focused entirely on technical skills. The ML/DS specialization offered by ECE in JSOE is restricted to ECE MS students with a required Bachelor’s degree in Engineering, Sciences or Mathematics. The others are MS programs for Computer Science & Engineering, Statistics or Biostatistics restricted to students with respective background and training. For instance, the Biostatistics MS is targeted to biomedical and public health data.

(b) The courses offered by the MS/DS program will be available, and in fact, will be taught by faculty jointly appointed with other departments representing domain knowledge. For instance, a faculty member jointly appointed with HDSI and Bioengineering is planning to offer a graduate course in Biomedical Data Analysis. Such a course will constitute a core requirement in the Bioengineering MS program as well as a HDSI cross-listed elective course for students with background and interest in biology and engineering. Thus, by serving as a catalyst for creation of new courses and student support, HDSI seeks to enhance the overall capacity of UC San Diego

in serving educational and training needs of a growing population of students that can not be served by any of the existing programs.

Impact on Existing Programs: We do not anticipate any adverse material impact on the existing ECE specialization in machine learning/data science which is a part of a much larger MS in ECE degree. On the contrary, we expect a positive impact from increased participation of various academic units in Data Science related subject areas (such as Computational Biology, Computational Chemistry etc.).

MS degree programs both in CSE and ECE serve Computer Science and Electrical Engineering areas respectively (with a set of Machine Learning related courses). Both are rather large programs with growing student enrollments at roughly 500-600 students each selected from over 7,000 applicants annually. On the other hand, Mathematics' entire MS program has an enrollment of 20 students. The new MS Biostatistics program provides a specialized degree with a view towards creating a pathway to PhD degree in Biostatistics. We anticipate that some courses especially on Survival Analysis (Math 284) and other elective courses that will be taught by jointly appointed faculty and cross-listed to be accessible to students drawn from different departments.

Overall, the proposed MS/DS program expands the pool of applications by enabling students from diverse backgrounds such as Economics, Cognitive Science, Biology etc to consider a career in Data Analytics or its application to their own domains. For the selected specialization area courses, some of MS/DS students will increase enrollments in these classes most of which are cross-listed with DSC courses and/or taught for HDSI or jointly-appointed HDSI faculty in the Math, CSE, ECE departments. We expect to see a spread of MS/DS students across half a dozen or so available area specializations, thus causing a small (likely single-digit) increase in class enrollments with a resulting increase in corresponding class-based MS income to the partner departments. Similarly, academic units (such as Cognitive Science, Biology etc) that do not currently have an MS program can also participate in MS/DS program by creating/offering Data Science related 2-3 specialization courses in their areas for the graduate students, thus catalyzing growth of the graduate academic programs in Data Science.

F. Contributions to Diversity

Our vision for how the proposed program will advance UC's goals for diversity is informed by the founding document "HDSI Strategy for Inclusive Excellence". It is a living document available online at <https://bit.ly/HDSI-Diversity> that will be updated with information available from Diversity Dashboard and our surveys. The nascent nature of the HDSI organization provides us with additional flexibility to incorporate EDI goals into the DNA of the new institution, that is, embedded in all our processes and actions *ab initio* with these goals. In particular, the three core tenets of "access & success, climate and accountability" are vigorously pursued. The Institute

has taken two concrete steps towards EDI goals that will directly impact its programs -- including the proposed graduate program -- in the coming years. First, every faculty hire to HDSI has been provided with \$30K in support of EDI goals. These funds are held by the Institute and released for specific and approved activities that advance diversity goals. The faculty members are encouraged to pool these resources and seek additional matching resources from the Institute to launch substantive programmatic activities. Beyond this “access and success” part, the second element of HDSI strategy directly addresses climate and accountability. HDSI has proposed and eventually succeeded in using its endowment resources to identify and recruit a full-time coordinator for broadening participation (<https://bit.ly/HDSI-BPC>). While a permanent position is pending and yet to be created, working with the administration we have been able to recruit, Saura Naderi, to be tasked full-time in developing measures, establishing metrics and directing the activities related to EDI goals. The personnel and EDI share-pool mentioned earlier empowers HDSI faculty to put to action concrete plans -- including seminars, enrichment activities and additional counseling, etc that are implemented, tracked and accounted for. With increased direct attention to BPC (broadening participation in computing) plans in research proposals to agencies such as the NSF, HDSI’s broadening participation plan provides a sustainable institutional mechanism and support for the HDSI community.

Beyond the two elements of HDSI EDI strategy mentioned earlier, the proposed graduate program will also provide an excellent vehicle for deploying our fellowship support to encourage URM participation as well as make the program accessible for academically strong but economically disadvantaged students to ensure the program provides an affordable pathway for a broad and diverse student population. Among the programs we have already devised and launched⁷ are scholarships for graduate students (a commitment of \$600K for the current year, and likely to rise in coming years), as well as access to learning outside the classroom through mechanisms such as EdX micromasters programs where the Institute offers financial support to all students interested in taking *on-ramp* classes at their own pace. This on-ramping is a critical element of our strategy to leverage the existing micromasters program (see Section 1.5 on Page 12 of the accompanying OMDS program proposal). It enables students from different backgrounds, who may otherwise be rejected from the MS/DS program, to demonstrate that they can do well in the program at low or no cost through our scholarship to undergraduate students across the campus.

Beyond strategic decisions and choices to appoint personnel, devote resources, we are cultivating a climate for faculty to conceive of new ideas that directly contribute to inclusive excellence. Among the measures that we seek to improve is participation rate of women and underrepresented minorities in our classes and degree programs, retention rate, progress towards

⁷ Please see page 60-61 of our annual report at <http://bit.ly/HDSIfirstyear>.

graduation and placement results. Institutional role models and mentoring are key means that are already implemented in HDSI foundations and shall remain a cornerstone of HDSI faculty recruiting and leadership advancement. Among concrete measures we have already taken are ensuring training every year by the *entire* faculty council for elimination of bias in faculty recruiting. Additional measures will surely be implemented as the coordinator for broaden participation starts her work in the coming months.

G. Relationship of the Proposed Program with Other UC Campuses

Among UC campuses, Data Science programs are offered by applied departments and schools as a part of professional training programs. Prominent among these is a Master of Information and Data Science (MIDS) offered by the School of Information (now part of Berkeley Division on Data Sciences) at UC Berkeley. This is a self-funded professional masters degree. All courses are offered online with a total of about 600 students. Another related online program from UC Berkeley is Master of Information and Cybersecurity (MICS). Unlike Berkeley, HDSI's MS program is an on-campus residential program. UC Davis offers masters programs in Business Analytics by the Graduate School of Management and Health Informatics by the School of Medicine. Similarly, UCLA and UCI offer masters in Business Analytics. UC Riverside offers a Data Science specialization of the MS in Engineering program by the Bourns College of Engineering.

These programs are different in curricular goals, student background and training and program logistics from the envisioned residential MS program. To a first order, none of these programs on other campuses prepare students for a research career in Data Science or Data Science applications in selected domain areas, an important objective of the proposed graduate program. More importantly, no program provides the wide accessibility to students from diverse educational backgrounds to Data Science and its applications. Indeed, the structure of the proposed MS/DS program consisting of *foundation*, *core* and *specialization* tracks makes it possible for HDSI to offer a single program that admits students with training as broad as engineering, sciences, social sciences, business and humanities and produces students with a Master's Degree in Data Science with well-defined specializations that make it easier for them to pursue different career paths in targeted domains.

H. Program Administration and Resource Planning

The proposed program will be offered by the Halicioğlu Data Science Institute (HDSI), established as an academic unit by the UC Academic Senate in June 2018 under a divisional budget model on the UC San Diego campus. The Institute also carries a \$75M founding endowment with an annual payout that is expressly dedicated to support the mission of HDSI in training and preparation of Data Science talent by the Institute activities and programs. Due to its

founding commitments, the proposal does not require a separate new infusion of campus resources.

The Institute faculty, and members of its faculty council, are listed below with their annual teaching workload in HDSI. Graduate Admissions and Graduate Program committees will be formed as a standing committee of HDSI. These two committees will also be supported by a full time academic coordinator currently under recruitment to ensure program operation and academic advising of the graduate students.

Faculty Members with Teaching Responsibilities in HDSI Programs			
<i>Faculty Group</i>	<i>Names and Title</i>	<i>Appointments</i>	<i>Nominal Teaching Workload in Data Science</i>
Full Time Faculty in HDSI (11)	Mikhail Belkin, Professor	HDSI	3 courses
	Justin Eldridge, Assistant Teaching Professor	HDSI	6 courses
	Aaron Fraenkel, Assistant Teaching Professor	HDSI	6 courses
	Yian Ma, Assistant Professor	HDSI	3 courses
	Arya Mazumdar, Associate Professor	HDSI	3 courses
	Gal Mishne, Assistant Professor	HDSI	3 courses
	Yusu Wang, Professor	HDSI	3 courses
	Babak Salimi, Assistant Professor	HDSI	3 courses
	Zhiting Hu, Assistant Professor	HDSI	3 courses
	Berk Ustun, Assistant Professor	HDSI	3 courses
	Lily Weng, Assistant Professor	HDSI	3 courses
Joint Faculty in HDSI	R. Stuart Geiger, Assistant Professor	Communications & HDSI	1.5 courses
	Mikio Aoi, Assistant Professor	HDSI & Neurobiology	1.5 courses
	Jingbo Shang, Assistant Professor	CSE & HDSI	1.5 courses

(4 FTE)	Benjamin Smarr, Assistant Professor	Bioengineering & HDSI	1.5 courses
	Barna Saha, Associate Professor	CSE & HDSI	1 course
	Arun Kumar, Assistant Professor	CSE & HDSI	1 course
	Yoav Freund, Professor	CSE & HDSI	1 course
	Jelena Bradic, Associate Professor	Mathematics & HDSI	1 course
	Rayan Saab, Associate Professor	Mathematics & HDSI	1 course
	Alex Cloninger, Assistant Professor	Mathematics & HDSI	1 course
	Margaret Roberts, Associate Professor	Political Science & HDSI	1 course
	Armin Schwartzman, Professor	Biostatistics & HDSI	1 course
Fellows & Others	Bradley Voytek	HDSI Fellow, Cognitive Science	2 cross-listed courses
	Ilkay Altintas	HDSI Fellow, SDSC	
	Virginia De Sa	HDSI Associate Director, Cognitive Science	1 cross-listed course
	Dimitris Politis	HDSI Associate Director, Mathematics	1 cross-listed course
	Rajesh K. Gupta	HDSI Director, CSE	1 cross-listed course

In addition to the 28 faculty members listed above (15 FTE), the Institute currently has one offer pending for a joint faculty appointment with Philosophy and planning to fill one LPSOE slot. It anticipates additional 1-2 new faculty members to join the Institute for a total faculty strength of 16-18 FTE including 3 FTE LPSOE and 8 U18 lecturers.

Together, these provide a capacity of 47 courses by the current ladder-rank faculty and a combined total capacity of 57-73 courses/sections. The current Data Science undergraduate program accounts for 35 courses/sections per year. Conservatively, the Institute has the capacity to offer 6-12 graduate courses per quarter that enables it to adequately serve the proposed MS program consisting of 20 new courses. Further, the current TA support provides for 8-10 TA FTEs who will be drawn from the graduate student pool in Data Science, providing approximately \$600K per year in additional financial support for Data Science graduate (MS and PhD) students.

I. Plan for Evaluation of the Program

The MS program will be formally evaluated like all other UC San Diego graduate programs, consistent with senate regulations every 8 years. This formal evaluation will include an external review and UC San Diego graduate council oversight. In addition, as mentioned earlier, the HDSI faculty plans to perform a mid-flight review after three years for assessment of program success and needs, evaluation of program placement outcomes. As with the formal evaluation, the internal review process -- both annually and mid-flight -- will include student feedback and surveys, teaching evaluations, alumni and industry feedback.

Despite the newness of the institute, members of the Institute's faculty have significant prior and ongoing experience in building, launching and directing MS and MAS programs in the CSE, Bioengineering, Mathematics, Cognitive Science and Biostatistics units. The academic coordinator position was specifically designed with a view towards broadening participation, improving student learning experience and career placement outcomes. The Institute is planning to be physically co-located in the same building with a segment of the Teaching and Learning Commons (TLC) starting 2021. This colocation will present us with additional opportunities for engagement with UC San Diego's expertise in improving learning experience and outcomes for all our students. Before the formal launch of the program, we plan to define learning outcomes for individual courses as they complete their approval process, and how these learning outcomes compose into overall degree goals. In doing so, we will be working with TLC to define the assessment plans, metrics to drive these assessments. Among the important metrics, we will pay special attention to completion rates, minimization of attrition, average time to degree, diversity and placement of our students, including the fraction that continues to a doctoral degree at UC San Diego or elsewhere.

2. Program

A. Undergraduate Preparation for Admission into MS/DS

The HDSI faculty have spent significant time discussing and formulating a plan that enables maximum participation of interested students into the envisioned graduate program. We are also keenly aware of our primary obligation to ensure successful and timely completion of the graduate degree program. Balancing these requirements has required us to structure the incoming stream of students into essentially three broad categories:

1. students who come with preparation in computing and/or information sciences at a level to master algorithmic programming and cloud computing skills;
2. students with preparation in mathematical subjects at a level to master statistical analysis and probability necessary for meaningful data analysis;
3. students who enter the program from other areas of science that rely upon collecting and analyzing observational or experimental data in order to advance scientific understanding. These are students with a degree in natural sciences such as physics, chemistry, biology, environmental sciences, etc. or coming from a social science background such as economics, political science, psychology, etc. Application examples may be causal inference in economics, assessing statistical significance of a pharmaceutical experiment or psychological treatment, the study of social networks in political science, etc.

We note that these are broad and overlapping categories. Even when students come prepared in both advanced computing and mathematics/statistics, Data Science education challenges them to apply these skills meaningfully in diverse applications, as well as improve their visualization/presentation skills.

Graduate admissions process will use text analysis methods to automatically sort and bin *admitted students* into three pools and thus drive the subsequent advising process including prior communication to the students regarding their preparation options using online and other offers by UC San Diego and other organizations.

Within the first week of arrival, each student will be scheduled for a 1-1 meeting with a faculty advisor and/or MS program academic coordinator. After meeting with their faculty advisor, newly admitted students may be directed to take specific upper-level undergraduate courses from different areas, in order to solidify their backgrounds when/if there is some perceived weakness; up to two such courses may count towards their M.S. units.

B. Admission Requirements

Admissions requirements for the MS/DS program are:

- Bachelor's degree in a relevant field
- Undergraduate GPA of at least 3.0 on a 4.0 scale
- Three letters of recommendation.
- TOEFL or TSE scores (international applicants only)

C. Foreign Language

Foreign language is not required for this degree.

D. Overview of the Proposed Masters Program

To meet its goals, the Data Science graduate program consists of two components: formal courses, as well as a terminating thesis or a course-direct comprehensive examination. Formal courses are structured into three groups discussed below. For each group of courses, we describe the courses being created. In some cases, courses are cross-listed with existing similar courses from partner departments which can be substituted accordingly. However, no assumption is made about availability and ability of Data Science graduate students to register into the cross-listed courses offered by partner departments.

E. Plan

The program will follow Master Plan I and Plan III. Both are discussed in detail in Section 2.

F. Unit Requirements

For the conferral of the M.S. degree in Data Science, 48 units (**12 courses**) will be required. Out of the 12 courses, at least 10 must be graduate-level (including courses marked as mezzanine) courses and independent research; at most two can be upper-level undergraduate courses. In addition, two out of the 10 graduate courses can be courses in areas not directly related to data science, e.g., economics, biology, medicine, etc. upon approval of the student's faculty advisor.

G. Structure of the Proposed Masters Program: Required and Recommended Courses

The course numbering in Data Science uses the following legend: HDSI Course Naming Conventions. Courses in the series end with "0" such as DSC 110, DSC 210 have as pre-requisite courses in lower division or a mezzanine senior course respectively.

Areas Description	UD UG series	Grad series	Notes
Data Management & Data Systems, Data Security,	DSC 10X	DSC 20X	Introductory &

Data Visualization			Mezzanine Courses
Computational & Mathematical Foundations	DSC 11X	DSC 21X	
Data and Life Science	DSC 12X	DSC 22X	
Digital Infrastructure, Computing Systems, Cloud, Cyber-infrastructure, Traditional & non-traditional computing systems, Sensing systems, IOT, CPS	DSC 13X	DSC 23X	
Data Science Theoretical Foundations (builds upon lower division DSC 4X series)	DSC 14X	DSC 24X	
Applied Machine Learning: Data Mining (incl. Graph mining, time-series mining), recommender systems, ML-based vision, Deep learning applications. Natural Language Processing	DSC 15X	DSC 25X	Multiple domain-specific ML applications.
Arts, Humanities, Society, Policy and Social Sciences	DSC 16X	DSC 26X	
Data and Physical, Environmental Sciences	DSC 17X	DSC 27X	
Capstone Project Courses	DSC 18X	NA	
Special Topics	DSC 19X	DSC 29X	Topics: 291, Projects: 292, Seminars: 293.
Directed Research	DSC 199	DSC 298, DSC 299	298: Independent Research; 299: teaching credit
TA/Tutor Training		DSC 599	

Program Structure:

Courses in Data Science Masters Program are structured into three groups of courses: **Group A**, **Group B** and **Group C**. **Group A** courses are introductory level courses taught at the level of undergraduate senior or mezzanine courses. **Group B** are core graduate level courses with prerequisites from Group A courses. **Group C** are advanced, specialized and free-standing courses, often part of the required courses in the Data Science specialization of Graduate Program in other departments. In all three groups, required courses are indicated as such; they can not be substituted by other courses without exception approval from the graduate program committee. We have specified instructors with partial appointments in HDSI who have provided written support letters to be instructors for specific courses. Primary instructor for each course is

mentioned in bold. As with any graduate program, interest of a jointly-appointed faculty in teaching a course is qualified by their actual availability to teach the course in a given quarter with the written permission of their home department. We outline a process that will be published on how course offerings will be decided as a part of yearly course and degree plans.

Group A: Foundational Knowledge and Skill Areas

We have identified *five* critical foundational knowledge and skills that each student graduating from the MS/DS program is expected to receive, at a graduate level. It is, therefore, important that all our entering students either have background preparation or have courses available in the program to ensure a successful completion of the stipulated MS degree program:

1. **Programming skills:** ability to efficiently translate algorithmic knowledge and analysis methods into suitable programming platforms, especially using cloud computing resources.
2. **Data organization methods and skills:** ability to cast data from raw sources into formats (structured or semi-structured) that are amenable to scalable automated analysis, visualization on various platforms, data wrangling.
3. **Numerical Linear Algebra:** knowledge of underlying mathematics that supports the ability of students to conceptualize transformation operations and convert them into computational algorithms such as Principal Component Analysis (PCA).
4. **Multivariate Calculus:** the mathematical study of a function of multiple variables as required in understanding optimization methods such as gradient descent that underlie much of modern machine learning.
5. **Probability and Statistics:** understanding randomness in data that is important for methods such as PCA and stochastic gradient descent; quantifying the accuracy in estimation and model-fitting; performing multiple hypothesis tests; and optimality in estimation and prediction.

Among the incoming students interested in the Data Science graduate program, there may be some lacking the basic background at the undergraduate level in one or more of the above areas; this would prohibit them from taking the relevant graduate level courses.⁸ Accordingly, we have devised **five** foundational knowledge area courses described in the catalogue copy and listed below. A student can receive credit towards the M.S. degree for a maximum of **four** courses from the list of courses below.

1. **DSC 200: Data Science Programming [New]**, 4 units: Computing structures and programming concepts such as object orientation, data structures such as queues, heaps,

⁸ Incoming students graduating from Data Science or related majors may have taken courses in all of the five areas thus shortening the path to graduation by receiving credit for equivalent courses.

lists, search trees and hash tables. Laboratory skills include Jupyter notebooks, RESTful interfaces and various software development kits (SDKs). *Instructors: Aaron Fraenkel, Yoav Freund*

2. **DSC 202: Data Management and Database Visualization Methods [New]**, 4 units: Data management using relational data model, schema design, querying with SQL, and introduction to NoSQL systems. Popular database visualization methods are covered as well. *Instructors: Jingbo Shang, Amarnath Gupta*
3. **DSC 210: Numerical Linear Algebra [New]**, 4 units: Linear algebraic systems, least squares problems, orthogonalization methods, ill-conditioned problems, eigenvalue and singular value decomposition, principal component analysis. *Instructors: Rayan Saab, Alex Cloninger, Gal Mishne*
4. **DSC 211: Multivariable Calculus and Optimization [New]**, 4 units. Continuity and differentiability of a function of several variables, gradient vector, Hessian matrices, Taylor approximation, fundamentals of optimization, Lagrange multipliers, convexity, gradient descent. *Instructors: Yian Ma, Rayan Saab*
5. **DSC 212: Probability and Statistics for Data Science [New]**, 4 units: Probability, random variables, distributions, central limit theorem, maximum likelihood estimation, method of moments, confidence intervals, hypothesis testing, Bayesian estimation, introduction to simulation and the bootstrap. *Instructors: Jelena Bradic, Dimitris Politis, Armin Schwartzmann*

Student advising notes:

- a. Incoming students must meet with their assigned HDSI faculty advisor to customize their individual program of study. Students must exemplify they have the required preparation in the above five areas in order to be exempted from taking some (or all) of the above courses. For example, having taken MATH 170A (or equivalent) would indicate that the student does not need to take DSC210. Having taken MATH 173A (or equivalent) would show that the student does not need to take DSC211, and having taken MATH 181A (or equivalent) would show that the student does not need to take DSC212. Similarly, having taken BENG216 would show that the student does not need to take DSC200.
- b. In addition, incoming students can transfer up to two upper-level undergraduate or graduate courses taken under a different program and/or university, as long as (i) these courses are related to one of the above five foundational areas, (ii) these courses have not already been used for credit towards a different degree, and (iii) these courses are

approved by the student's HDSI faculty advisor to establish their relevance to data science and avoid course duplication.

- c. A student can receive credit towards the M.S. degree for a maximum of two courses (8 units) taken at the upper-division undergraduate level, subject to the approval of the student's faculty advisor. These two courses can be transferred (as discussed above), or taken during the course of the M.S. program. For example, a student can make use of the equivalencies discussed in part (a) above.

Group B: Core Knowledge and Skill Areas [M.S. students take at least five courses]

Building upon the foundation courses in Group A, the Master's program identifies several *core* graduate courses. Two of the core courses are required for all M.S. students, including those with a Bachelors in Data Science. The two **required** courses are:

1. **DSC 240: Machine Learning [New]**, 4 units: A graduate level course in machine learning algorithms: decision trees, principal component analysis, k-means, clustering, logistic regression, random forests, boosting, neural networks, deep learning. *Instructors: Misha Belkin, Yian Ma, Jelena Bradic, Gal Mishne*
2. **DSC 260: Data Ethics and Fairness [New]**, 4 units: Ethical considerations regarding privacy and control of information. Principles of fairness, accountability, and transparency. Use of metadata to information algorithms. Algorithmic fairness. Policy issues such as the Fair Information Practices Principles Act, and laws concerning the "right to be forgotten." *Instructor: R. Stuart Geiger*

In addition, a Master's student must select **at least three** out of the following seven core courses⁹

3. **DSC 203: Data Visualization and Scalable Visual Analytics [New]**, 4 units: Commonly used algorithms and techniques in data visualization. Interactive reasoning and exploratory analysis through visual interfaces. Application of data visualization in various domains including science, engineering, and medicine. Scalable interactive methods involving exploring with big data and visualization methods. Techniques to evaluate effectivity and interpretability of analytical products for diverse users to obtain insights in support of assessment, planning, and decision making. *Instructors: Ilkay Altintas, Juergen Schulze*
4. **DSC 204: Scalable Data Systems [New]**, 4 units: Storage/memory hierarchy, distributed scalable computing (i.e., cluster, cloud, edge) principles. Big Data storage, management

⁹ More course options will be added in areas such as Artificial Intelligence and Accountability/Trust and Data-driven Policy.

and processing at scale. Dataflow programming systems and programming models (MapReduce/Hadoop and Spark). *Instructors: Yoav Freund, Mai Nguyen, Ilkay Altintas*

5. **DSC 241: Statistical Models [New]**, 4 units: linear/nonlinear models, generalized linear models, model fitting and model selection (cross-validation, knockoffs, etc.), regularization and penalization (ridge regression, lasso, etc.), robust methods, nonparametric regression, conformal inference, causal inference. *Instructors: Ery Arias-Castro, Jelena Bradic*
6. **DSC 242: High-dimensional Probability and Statistics [New]**, 4 units: Concentration inequalities, Markov processes and ergodicity, martingale inequalities, empirical processes, sparse linear models in high dimensions, Principal component analysis in high dimensions, estimation of large covariance matrices. [This class may be cross-listed with the Mathematics Department.] *Instructor: Jelena Bradic, Rayan Saab*
7. **DSC 243: Continuous Optimization [New]**, 4 units: Linear/quadratic programming, optimization under constraints, gradient descent (deterministic and stochastic), convergence rate of gradient descent, acceleration phenomena in convex optimization, stochastic optimization with large data sets, complexity lower bounds for convex optimization. *Instructor: Yian Ma, Rayan Saab*
8. **DSC 244: Large-Scale Statistical Analysis [New]**, 4 units: Exploratory data analysis, diagnostics, bootstrap, large-scale (multiple) hypothesis testing, false discovery rate, empirical Bayes methods. [This class may be cross-listed with the Mathematics Department.] *Instructor: Armin Schwartzmann, Jelena Bradic*
9. **DSC 250: Advanced Data Mining [New]**, 4 units: Graph mining and basic text analysis (including keyphrase extraction and generation), set expansion and taxonomy construction, graph representation learning, graph convolutional neural networks, heterogeneous information networks, label propagation, and truth findings. [Prerequisite: DSC 190A or CSE158 or equivalent] *Instructor: Jingbo Shang*

In addition, MS students can take advantage of electives to complete their course of study. These courses can be advanced courses in core Data Science subjects listed under Group B as research topics (DSC 291) courses, or they can be graduate (or upper-division undergraduate) courses in other departments subject to approval by the student's HDSI faculty advisor.

Below is a (partial) list of such general elective courses:

General Elective Courses:

DSC 205: Geometry of Data, *Instructor: Gal Mishne*

DSC 231: Embedded Sensing and IOT Data Models and Methods, *Instructor: Rajesh Gupta*

DSC 251: Machine Learning in Control: Estimation of stability and uncertainty, optimal control, and sequential decision making. *Instructor: Yian Ma*

DSC 252: Statistical Natural Language Processing, 4 units. Diving deep to the classical NLP pipeline: tokenization, stemming, lemmatization, part-of-speech tagging, named entity recognition, parsing, and machine translation. Finite-state transducer, context-free grammar, Hidden Markov Models (HMM), and Conditional Random Fields (CRF) will be covered in detail. *Instructor: Jingbo Shang*

DSC 253: Advanced Data-driven Text Mining, 4 units: Unsupervised, weakly supervised, and distantly supervised methods for text mining problems, including information retrieval, open-domain information extraction, text summarization (both extractive and generative), and knowledge graph construction. Bootstrapping, comparative analysis, learning from seed words and existing knowledge bases will be the key methodologies. *Instructor: Jingbo Shang*

DSC 254: Statistical Signal and Image Analysis. *Instructor: Gal Mishne, Armin Schwartzman*

DSC 213: Statistics on Manifolds. *Instructor: Armin Schwartzman, Alex Cloninger*

CSE 234: Data Systems for Machine Learning. *Instructor: Arun Kumar*

MATH 281A-B-C. Mathematical Statistics (4-4-4 units), *Instructor: Jelena Bradic, Ery Arias-Castro*

MATH 284. Survival Analysis (4 units), *Instructor: Lily Xu*

MATH 285. Stochastic Processes (4 units), *Instructor: Ruth Williams*

MATH 287A. Time Series Analysis (4 units), *Instructor: Dimitris Politis*

MATH 287B. Multivariate Analysis (4 units), *Instructor: Ery Arias-Castro*

COGS 243: Statistical Inference and data analysis (4 units), *Instructor: Angela Yu, Virginia de Sa*

The program also provides for domain specializations using domain courses as electives; see Group C courses in the next section.

Group C: Specialization Areas [3 courses, 12 units; can be all research units]

Group C courses aim to provide either practical experiences in chosen specialization areas, or advanced training for students preparing for doctoral programs. All courses in this group are electives, including use of Project or Thesis course units as DSC 292 or DSC 299. Courses in this group also serve as a means to directly engage faculty in departments across the campus who are directly interested in Data Science related topics and instruction. Consequently, we make important courses taught by HDSI affiliated faculty visible to the Data Science graduate students. However, their availability is subject to schedule and enrollment constraints of the individual departments.

Based on written approval from participating departments, courses available in a given domain in a given year will be announced beginning of the academic year with a pre-registration deadline for capacity planning purposes.

We describe five distinct specialization areas and anticipate additional specialization options to be added to the program in future as we engage with participating departments. Specializations availability will be planned over a 2-year rolling window to ensure specialization areas selected by a given cohort of MS students will be available.

We have solicited support letters from departments that offer the courses listed in the specializations below (Appendix B). A complete set of support letters will be provided prior to review by the academic senate.

Specialization Area: Bioengineering

This specialization area is targeted at students with background and training in life sciences areas (Bioengineering, Bioinformatics, Systems Biology) who seek to apply their knowledge of Data Sciences in biomedical applications. This specialization is offered in coordination with the “Biomedical Data Science and Engineering” specialization of the MS degree in Bioengineering offered by the Department of Bioengineering. The specialization area is open to all students in MS/DS program who meet course prerequisites for the courses in the track.

BENG 218: Biomedical Technologies and Data Measurements, 4 units

BENG 203: Genomics, Proteomics and Network Biology, 4 units

BENG 211: Systems Biology & Bioeng. I: Biological Components, 4 units

BENG 212: Systems Biology & Bioeng. II: Network Reconstruction, 4 units

BENG 213: Systems Biology & Bioeng. III: Building and Simulating Large Scale In Silico Models, 4 units

BENG 221: Mathematical Models for Bioengineering, 4 units

BENG 230A: Biochemistry, 4 units

BENG 230B: Cell and Molecular Biology, 4 units

BENG 276: Numerical Analysis in Multiscale Biology, 4 units

COGS 278: Genetics and Individuality, 4 units

PHYS 278: Biophysical Basis of Neuronal Computation, 4 units

Specialization Area: Business

The business specialization studies managerial decision making and explores opportunities where the use of data and models can positively impact business performance. Tracks in this specialization center on each of the primary functional areas within a business.

Track: Marketing

MGT 475: Research for Marketing Decisions (4 units)

MGT 477: Consumer Behavior (4 units)

MGT 489: E-commerce (4 units)

MGTA 455: Customer Analytics (4 units)

MGTA 479: Pricing Analytics (4 units)

Track: Supply Chain and Technology

MGT 450: Project Management (4 units)

MGT 451: Technology Innovation and Strategy (4 units)

MGTA 456: Supply Chain Analytics (4 units)

MGTA 463: Fraud Analytics (4 units)

Track: Finance

MGT 407 Valuation in Corporate Finance (4 Units)

MGTF 402 Investment Analysis (4 units)

MGTF 404 Financial Econometrics and Empirical Methods (4 Units)

MGTF 405 Business Forecasting (4 Units)

MGTF 406 Behavioral Finance (4 Units)

MGTF 415 Collecting and Analyzing Financial Data (4 Units)

Specialization Area: Machine Vision & Interaction Design (MVID)

Machine Vision and Interaction Design concerns recognition and analysis of visual objects and effective interfaces for human machine interactions.

COGS 202. Computational Modeling of Cognition, 4 units. *Instructor: Eran Mukamel*

COGS 220. Information Visualization, 4 units. *Instructor: Jim Hollan*

COGS 225. Image Recognition: Computer Vision I. *Instructor: Zhuowen Tu*

COGS 283. Big Visual Data Processing, 4 units. *Instructor: Zhuowen Tu*

Specialization Area: Computational Neuroscience

Brain and Mind explores deeper into understanding of the brain structures and behavior and its connection to cognition. The specialization consists of tracks that dive deeper into understanding of brain and neural structures, their modelling or imaging through multiple imaging modalities, or spatio-temporal signal processing to understand neural processes.

COGS 260: Brain Computer Interfaces. *Instructor: Virginia De Sa*

BGGN 246: Computational neurobiology. *Instructor: Terry Sejnowski*

BGGN 260 / BENG 260/ PHYS 279: Neurodynamics. *Instructor: Gert Cauwenberghs*

COGS 260 / NEU 282: Algorithms for the Analysis of Neural Data. *Instructor: Eran Mukamel*
COGS 280: Neural Oscillations. *Instructor: Bradley Voytek*

Specialization Area: Networks

Complex networks such as biological networks, (bio)chemical networks, cognitive networks, computer networks, computation networks, collaboration networks, epidemiological networks, social networks, transportation networks and telecommunication networks permeate our lives. The structure and behavior of such networks are increasingly informed by large amounts of data. Understanding and controlling these networks draws on theory and methods from multiple areas of mathematics, statistics and computer science, as well as specific knowledge stemming from application domains. This specialization allows students to explore theoretical foundations as well as applications.

Math 261A: Probabilistic Combinatorics and Algorithms (especially graph theory). *Instructor: Todd Kemp, Andrew Suk*

Math 277A: Topics in Computational and Applied Math: Diffusion Geometry and Metric Graph Learning. *Instructor: Alex Cloninger*

Math 289A: Topics in Probability and Statistics: Stochastic Networks. *Instructor: Ruth Williams*

Math 289B: Topics in Probability and Statistics: Random Graphs and Random Matrices.
Instructor: Ioana Dumitriu

DSC 205: Geometry of Data, *Instructor: Gal Mishne*

BNFO 286. Network Biology and Biomedicine. *Instructor: Trey Ideker*

POLI 287: Multidisciplinary Methods in Political Science: Social Networks. *Instructor: James Fowler*

SIOB 276: Quantitative Theory of Populations and Communities. *Instructor: George Sugihara*

ECE 227: Big Network Data. *Instructor: Massimo Franceschetti*

MAE247: Cooperative Control of Multi-Agent Systems. *Instructor: Jorge Cortes*

Additional Specialization Areas may include: Design, Civil Society & Public Policy, Environmental Data Science, Cyber-Physical Systems, Automation and Control, and Biomedical Informatics.

H. Thesis and Comprehensive Examination Requirements

A successful completion of the MS program will require completion of one of the following two choices that must be selected and approved by the MS program committee before the end of first quarter into the MS program:

- a. **Thesis Option:** The student must sign up for a minimum of 8 and maximum of 12 units of DSC 299 (Independent Research) which can also be used to meet Group C requirements. All courses must be completed for a letter grade, except the DSC 298 units which are taken only as Satisfactory/Unsatisfactory basis. The student will perform a thesis research under the guidance of a thesis advisor and a thesis committee consisting of at least three members. It is required that at least two members of the committee are members of the HDSI faculty council, one of the three committee members can be an industry fellow with an adjunct appointment or a faculty member drawn from another department or division. Chair of the committee shall be appointed by the MS program committee. Alternatively, the industry fellow may be requested to serve as the fourth member of the committee. The committee must be approved by the Graduate Division by the end of third quarter in the MS program;
- b. **Course-Directed Comprehensive Examination Option:** Comprehensive examination option is designed to evaluate a student's ability to apply their Data Science knowledge into practical problem solving in a specific domain. In order to do so, students using this option are tested in three course-hosted comprehensive examinations, each in a different subject area. This plan conforms to the guidance issued by the Graduate Council dated February 11, 2016 confirming compliance of this plan to Senate Regulation 700 and has been adopted on UC San Diego Campus¹⁰. In this format of the Comprehensive Examination, the students must answer comprehensive questions in their chosen domain in each of the three selected courses. The comprehensive examination is integrated into the host courses, and in most cases, the associated work serves dual purposes, contributing independently to the student's course grade and comprehensive examination score. The comprehensive examination typically consists of a specific class assignment or examination, or a portion thereof, that has been explicitly approved by the MS program committee. Determination of the outcome on the comprehensive exam is separate from the grade in the host course. The students are required to successfully pass the comprehensive examination in three courses drawn from each of the three areas: computing, math/statistics, systems. Students are permitted up to five attempts, that is, five different courses. No more than three course-hosted comprehensive examinations can be taken in a single quarter, and no comprehensive examination can be repeated in a single quarter. The courses marked for comprehensive examination can be taken only for a letter grade. Course-hosted examinations are registered in the beginning of each quarter and students must register in advance by the specified deadline for the examination. The examination is supervised by a faculty committee responsible for the content, evaluation

¹⁰ For instance, CSE has adopted it as a Comprehensive Plan (Interdisciplinary)
<https://cse.ucsd.edu/graduate/degree-programs/ms-program/ms-2015-comp-interdisc>

and administration of the examination which is separate from the course instructor who is responsible for the course grade but not success in the comprehensive examination.

I. Field Examinations

No field examinations are required.

J. Qualifying Examinations

No Qualifying examinations are required.

K. Thesis and/or Dissertation

Students opting for Plan I are required to file an approved thesis to satisfy requirements for completion of the program.

L. Final Examination

No final examination is required. Students opting for Plan II are required to follow instructions regarding course-directed comprehensive examination.

M. Explanation of Special Requirements Over and Above Graduate Division Minimum Requirements

There are no special requirements over and above Graduate Division requirements.

N. Relationship to Master's and Doctor's Programs

The program is designed to support a future Doctoral Program in Data Science currently under planning.

O. Special Preparation for Careers in Teaching

Teaching is not a component of this degree program.

P. Sample Program

Course pathways through the MS/DS program differ based on the background preparation and training of the student. For instance, students with background in computer science, information science are likely to take DSC 210, DSC 240 and DSC 241 whereas students with general engineering background may take DSC 202, DSC 240, DSC 241 and students with background preparation in mathematical sciences or mathematics-oriented majors such as economics may choose DSC 200, DSC 202, DSC 241. In all cases, the intent of the course options in Group A courses is not to be seen as electives but rather as requirements that they need to fulfill, either by taking the course or by showing prior knowledge to be implemented by student advising. The following table describe some possible pathways through the proposed Master's Program:

	Fall	Winter	Spring	Notes
Student A with background in Computing/Information				
Year 1	DSC 210, 211, 240	DSC 212, 241, 242	DSC 243, 244, COGS 278	Comprehensive exam option
Year 2	DSC 260, COGS 280, 283			
Student B with background in Mathematics/Statistics				
Year 1	DSC 200, 240, 260	DSC 202, 241, 242, MATH 287A	DSC 204, 244, MATH 287C	Student transfers MATH 185 and 189 having taken them beforehand (and not used towards B.S.).
Year 2				Comprehensive exam option
Student C with Domain Specialization Background				
Year 1	DSC 200, 210, 211	DSC 202,212, 250	DSC 298, 204, 244	Student takes all 5 courses of Group A but only 4 count towards M.S. degree
Year 2	DSC 298, 240, 260	DSC 298		Thesis Option.

Q. Normative Time from Matriculation to Degree

We expect common pathways through the program to take two years through the program. However, students with training in computing and mathematics would be able to reduce time spent in the foundation courses, thus completing typically in three to five quarters. A minimum in-residence time of three quarters will be needed even for students with a Data Science bachelor's degree to ensure they are able to complete the core and specialization requirements.

3. Projected Need

A. Student Demand for the Program

Section 1(C) outlines the programmatic reasons for launching the graduate program as a key vehicle for advancing knowledge and practice in Data Science. The driver for such a program, indeed, of HDSI as an institution, is to satisfy the growing demand for the graduate program in data science both internally as well as externally. Industry surveys have repeatedly shown a soaring need for data scientists. Chief and credible among these are reports by McKinsey¹¹, IBM¹²Bloomberg¹³. As mentioned, in a survey of our students graduating from Data Science major, roughly a third of them have indicated their interest in a graduate degree in Data Science program. In 2019, nearly a half (1571) of over 3000 applicants to various graduate degree programs at UC San Diego who indicated interest in Machine Learning and Artificial Intelligence related topics in Electrical Engineering, Computer Science and Cognitive Science, directly indicated their interest in Data Science programs at HDSI.

Thus, the demand for a graduate training in Data Science is high and continues to grow. Further, the program not only serves the HDSI mission of educating talent in the area of Data Science, but also serves as a vehicle for continued engagement and proliferation of Data Science training across various graduate programs through new foundational, core and elective course offerings that engage domain experts into the field of Data Science. The program provides an excellent means to create new educational opportunities for students, especially for underserved and economically-disadvantaged student populations who can benefit from graduate scholarships offered by HDSI as a part of its core endowment-supported activities that we have mentioned earlier.

Based on the demographic data of students interested in constituent areas of machine learning, information theory and statistics, one would expect a skewed demographic balance at about one-fifth of domestic students. However, with growth of areas such as computational biology and computational social science, we expect international fraction to be closer to campus average of 60% (7%, Social Sciences, 15% in Arts and Humanities, 23% Biological Sciences, 32% Health, 45% GPS, 50% Physical Sciences, 70% JSOE and RSM) and better than the founding areas of Mathematics (90%), Computer Science (80%), Electrical Engineering (83%). Indeed, the demographics on our Undergraduate Program support an expectation of one-third to one-half of domestic students. Following a similar reasoning, we expect a better ratio of California

¹¹ <https://mck.co/2W4LriY>

¹² <https://bit.ly/3dkiDIS>

¹³ <https://bloom.bg/3chfute>

resident applicants over computer science where we saw four times as many resident applicants for the MS program over the doctoral program. We shall be monitoring and reporting on these numbers in our annual program review as a part of HDSI annual report.

B. Opportunities for Placement of Graduates

Data Science as an academic subject area is rapidly emerging with Schools or Colleges of Data Sciences (such as Berkeley, Wisconsin, MIT, Columbia) or Departments such as NYU, Michigan, Yale, Cornell, UC Irvine, Virginia just to name a few. Regardless of institutional home as a department, division, school or college, the faculty demand for Data Science is expected to rise along with the need for a pipeline of graduate students in the coming years. While no survey data is available for the doctoral demand and a subject is addressed in greater depth in our Doctoral program proposal currently in preparation, we have plenty of data and evidence for placement of master's students. A detailed market analysis in support of the opportunity in 'Big Data' and long-term trends in this domain comes from a recent McKinsey Global Institute report that identified 'Game Changer' opportunities for US growth¹⁴. The May 2011 McKinsey Global Institute report, "Big Data: The next frontier for innovation, competition and productivity"¹⁵, predicted the need for over 500,000 data scientists by 2018. McKinsey projected a shortfall of 1.5 million additional managers and analysts in the U.S. who can "ask the right questions and consume the results of the analysis of big data effectively." These numbers were later analyzed by commercial outfits such as kdnuggets.com. Analysis of LinkedIn Workforce report dated August 2018 states "Nationally, we have a shortage of 151,717 people with data science skills". Kaggle, the largest community of data scientists and now a part of Google has over 2 million subscribers. The LinkedIn profile of Data Scientists lists 132,083 people with Data Scientist titles that are spread across IT services, Computer Software, Financial Services, Banking, Insurance and Higher Education.

In a separate 'bottom-up' study, the EMC Corporation, a publicly traded company with 60,000 employees, interviewed 497 data science and business intelligence professionals from around the world. The results of their study on the need for Data Scientists pointed to some interesting trends in the computing industry. About two-thirds of the individuals polled believe the demand for data scientists will outpace supply in the next five years with nearly 30% coming from professionals in disciplines other than computer science. The study also cited the lack of training and resources as the biggest obstacles to data science in organizations. These observations directly support the case for the need for rigorous scientific training for the professionals moving into the data science field.

¹⁴ http://www.mckinsey.com/insights/americas/us_game_changers

¹⁵

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Data Scientists constitute a separate category of jobs that are currently posted along with IT and business analytics positions. As of this writing, Indeed.com posts openings for 5953 Data Science jobs, Glassdoor lists 21,166 Data Science jobs with a salary range from \$76,000 to \$148,000 for an average of \$118,700. This compares with an average salary of \$76,500 for software engineers, \$70,700 for computer engineers and \$110,200 for computer scientists. Indeed, driven by the opportunities available, a whole industry has sprung up on Data Science placements.¹⁶ Data Science graduates will be well qualified for job titles such as data analysts, business intelligence and predictive analysis professionals. The students are likely to find employment across many areas including internet companies, banking, insurance, investments, engineering and healthcare. We will work with Career and Placement Services as well as Alumni Board to ensure mentoring and placement of graduates from the Data Science program.

C. Importance to the Discipline

Section 1(B) addresses the intellectual underpinning driving the emergence of the discipline of Data Science where UC San Diego already has an oversubscribed undergraduate program as well as active postdoctoral program offered by HDSI. A graduate program is crucial to the establishment of Data Science as an academic area. The proposed Masters program is a necessary step toward a complete graduate program that establishes the complete pipeline of talent into academic (teaching and research) careers in the emerging discipline.

D. Ways in which the program will meet the needs of the society

Data Scientists are highly sought after, showing a societal need for individuals with this professional competency. Furthermore, Data Sciences are already having an impact on many aspects of society, including e-commerce, financial industries, technology companies, health care, and academia. There are few aspects of society that will not be affected by Data Science. The proposed program directly serves the current and growing need of professionals in the area and its applications.

E. Relationship of the program to research and/or professional interests of the faculty

As of this writing, the Institute has appointed 2 teaching assistant professors, eight ladder-rank full-time professors (5 assistant, 1 associate, 2 full) as well as 10 ladder-rank joint faculty (4 at 50% and 8 at 25%) into the Institute, each of which directly lists Data Science as their main professional research interest.

F. Program Differentiation

Sections 1(E) and 1(F) cover in detail related programs at UC San Diego and in the UC system. Nationally, there is a distinction between online or near-online masters programs being pursued

¹⁶ <https://www.dataquest.io/blog/career-guide-find-data-science-jobs/>

by schools such as Georgia Institute of Technology (Georgia Tech), University of Illinois, Johns Hopkins, University of Michigan and UC Berkeley versus residential graduate degree programs offered by New York University (which also offers a PhD program in Data Science), Columbia, Cornell and many other schools. In contrast to the former programs that usually seek to offer a minimal set of core (and preparatory) courses, the graduate program offerings leading to a PhD degree seek broader and deeper training through a large and diversified set of electives, often under programs such as “CS+X” or “Stats+X” categories. The proposed MS program is designed to serve the academic needs of the growing discipline through a large number of elective courses: of the 20 courses that we are offering or planning to offer, only seven courses are in the foundations and core, while the rest are elective courses. We fully anticipate increasing participation from faculty and departments across the campus in creating additional electives/specialization courses thus extending the reach and impact of Data Science as a discipline.

4. Faculty

The HDSI faculty community consists of over 200 faculty affiliates organized into 44 different research clusters. The core faculty of HDSI consists of 47 faculty who are appointed at various levels of FTE scale reflecting the extent of teaching responsibilities within Data Science¹⁷. The table below lists the Institute faculty. A two-page CV for each faculty member is provided as a separate attachment to the proposal.

HDSI Faculty Council		
<i>Faculty Group</i>	<i>Names and Title</i>	<i>Specialization Areas</i>
Full Time Faculty in HDSI (11)	Mikhail Belkin, Professor	Machine Learning, Learning Theory, AI: understanding structure in data, analysis and algorithms for non-linear high-dimensional data
	Justin Eldridge, Assistant Teaching Professor	Machine Learning Theory: improving correctness of learning algorithms, clustering, process of learning
	Aaron Fraenkel, Assistant Teaching Professor	Machine Learning, Design of anti-fraud, anti-abuse systems
	Yian Ma, Assistant Professor	Machine Learning Theory: scalable inference methods, time-series data and sequential decision making, Bayesian inference algorithms
	Arya Mazumdar, Associate Professor	Machine Learning, Information Theory: error correcting codes for use in storage systems
	Gal Mishne, Assistant Professor	Signal processing and machine learning for graph-based modeling, processing and analysis of large-scale high-dimensional real-world data; unsupervised data analysis in neuroscience
	Yusu Wang, Professor	Computational Geometry,

¹⁷ <https://datascience.ucsd.edu/about/faculty/hdsi-faculty-council/>

		Topological/geometric methods for data analysis
	Babak Salimi, Assistant Professor	Data Management, Causal Inference, Fairness in Decision Support Systems
	Zhiting Hu, Assistant Professor	Machine Learning and Natural Language Processing with applications in controllable content generation, enterprise AI platforms and healthcare.
	Berk Ustun, Assistant Professor	Interpretability, Fairness, Accountability in Machine Learning, Applications of ML to Medicine, Finance, Justice and Business.
	Tsui-Wei (Lily) Weng, Assistant Professor	Robust, Reliable and Trustworthy AI Systems, Deep Reinforcement Learning, Fairness in machine learning and Robustness against adversarial attacks.
Joint Faculty in HDSI (4 FTE, 12 headcount)	R. Stuart Geiger, Assistant Professor	Computational Social Science: computational ethnography, socio-technical systems
	Mikio Aoi, Assistant Professor	Computational Biology: large-scale Bayesian nonparametric inference, Bayesian optimization, neuronal analysis
	Jingbo Shang, Assistant Professor	Data mining, natural language processing, and machine learning: mining and constructing structured knowledge from massive text corpora with minimum human effort
	Benjamin Smarr, Assistant Professor	Computational biology, dynamical systems, stochastic processes and Biological Circuits
	Barna Saha, Associate Professor	Theoretical Computer Science: algorithm design and analysis, probabilistic method and large scale data analytics

	Arun Kumar, Assistant Professor	Databases, Data management and software systems, data preparation, model selection, and model deployment, ML/AI-based data analytics
	Yoav Freund, Professor	Machine Learning and its applications in bioinformatics, computer vision, finance, network routing, and high-performance computing.
	Jelena Bradic, Associate Professor	Statistics: causal inference, ensemble learning, robust statistics and survival analysis with applications to gene-knockout experiments, understanding cell cycles, developing new policies or detecting effects of treatments onto survival
	Rayan Saab, Associate Professor	Mathematics: signal processing and analysis, sparse and low-dimensional representations of high dimensional data, compressed sensing
	Alex Cloninger, Assistant Professor	Applied harmonic analysis, machine learning, neural networks, analysis of high-dimensional data
	Margaret Roberts, Associate Professor	Political Science: automated content analysis, political methodology and politics of information
	Armin Schwartzman, Professor	Biostatistics: Signal and image analysis; functional and manifold-valued data; high-dimensional data; modern multivariate statistics; large scale multiple testing; applications in biomedicine and the environment.
Fellows & Administrative	Bradley Voytek, Associate Professor HDSI Fellow	Cognitive neuroscience: neural modeling and simulation, along with large-scale data mining and machine learning techniques, to understand the physiological basis of human cognition and age-related cognitive decline

	Ilkay Altintas, Research Scientist HDSI Fellow	Scientific workflows and solution architectures for data and computational science, eScience applications.
	Virginia De Sa, Professor HDSI Associate Director	Cognitive Science, Neuroscience, Computer Science: learning processes.
	Dimitris Politis, Distinguished Professor, HDSI Associate Director	Statistics: time series, bootstrap methods, and nonparametric estimation methods
	Rajesh K. Gupta, Distinguished Professor, HDSI Director	Embedded and Cyber-physical Systems: sensor data organization, metadata models and methods.
0% Appointme nts	Angela Yu	AI (artificial agents): learning and decision making under uncertainty, social cognition.
	Eran Mukamel	Computational neuroscience: modeling and analysis of large-scale data sets to understand complex biological networks of the brain
	Frank Wuerthwein	Physics: experimental particle physics, distributed high-throughput computing with large data volumes
	George Sugihara	Complex system dynamics, methods for forecasting chaotic systems, neurobiology, gene expression in cancer
	Henrik Christensen	Robotics, computer vision, AI: systems-oriented approach to machine perception, robotics and design of intelligent machines
	Julian McAuley, Assistant Professor	Machine Learning: Social Networks, using artificial intelligence in fashion choice, and data science in various applications.
	Larry Smarr, Distinguished Professor	High-performance computing and networking: advanced

		cyberinfrastructure, experimental systems
	Lucila Ohno-Machado, Professor	Biomedical Informatics: accessible and usable health data and its use in evidence-based health decisions
	Michael Pazzani, Professor Emeritus HDSI Distinguished Scientist	Machine learning, explainable artificial intelligence, personalization, internet search, and recommendation systems
	Michael Holst, Professor	Mathematical and Computational Physics: biochemistry and biophysics, computational fluid dynamics, computer graphics, materials science, and numerical algorithms relativity
	Robin Knight, Professor	Bioengineering, Cellular and Molecular Biology, Computer Science: Microbial analysis
	Ronghui (Lily) Xu, Professor	Mathematics and Biostatistics: machine learning, statistical inference for complex data-types in the presence of high-dimensional covariates
	Ruth Williams, Distinguished Professor	Mathematics: Probability theory, stochastic models of complex networks (e.g., in internet, systems biology)
	Shankar Subramaniam, Distinguished Professor	Bioengineering and Systems Biology
	Shannon Ellis, Assistant Teaching Professor	Human Genetics, Data Science Education
	Tara Javidi, Professor	Information Theory, Machine Learning: wireless mesh networks.
	Terry Sejnowski, Distinguished Professor	Computational Neurobiology, Neurosciences, Neural Networks
	Vineet Bafna, Professor	Bioinformatics, Computational Biology
	Young-Han Kim, Professor	Information Theory: network information theory.

In addition to the 28 faculty members listed above (15 FTE), the Institute currently has 1 offer pending and planning to fill two 100% (including one LPSOE). It anticipates additional 1-2 new faculty members to join the Institute for a total faculty strength of 16-18 FTE including 3 FTE LPSOE and 8 U18 lecturers.

5. Courses

Section 2(G) describes the structure of the program and courses. eCourse description of the courses and their instructors is attached in Appendix C. We note that the courses have been devised to ensure broadest possible access to Data Science masters program by motivated students from diverse educational backgrounds. Accordingly, the program makes provision for course credit for a maximum of 2 undergraduate courses as well as selection of 4 out of 5 courses (Group A: Foundational Areas) that ensure adequate preparation of students to enable successful completion of degrees. Group B: Core Areas specify a minimum of 5 courses that constitute the body of knowledge and skills in methods/tools areas of data science. This list includes two required courses on Machine Learning and Data Ethics & Fairness. Finally, elective courses (including thesis research) seek to specialize data science skills in specific areas or application domains.

6. Resource Requirements

As mentioned earlier that no new or additional resource requirements are expected from the campus in support of the proposed MS program. Instead, the Institute's continuing and planned expenses in graduate scholarships, faculty recruiting and cyber-infrastructure resources (including personnel) will be key enablers for the successful operation of the proposed MS program.

7. Graduate Student Support

There are four primary sources of funding for graduate students: (a) Graduate division support of graduate students based on campus policy on distribution of scholarships support to graduate students. It was earlier known as "block grants" derived on the basis of campus policies for academic units based on their need and extramurally funded research activities; (b) Teaching assistant support. Currently, TA support is at 8 TA FTE per year and expected to rise with increase in undergraduate enrollment in our majors and minors (from currently at 700 students in the major, 5000 students in classes annually to 1000 majors and 10000 students in classes annually in three years). Graduate students will be trained and once determined to be qualified per university regulations, they will be offered TAships; (c) Extramurally funded research projects including training grant(s). Extramural funding is likely to be the largest source of funding for our graduate students, given the extensive growth and consistent availability of research funded by organizations such as NSF, DARPA, DOE, ARL and others. Data Science areas are among the most intensely invested areas of research both by public and private organizations (foundations). Based on budget analysis provided as a part of 3-year FTE planning, we anticipate annual research support of \$200K/year per faculty appointed in the institute; (d)

endowment-supported graduate student scholarships. We have currently budgeted \$600K per year for this program. We expect to grow this program with the growing industry contributions and philanthropic support to the Institute. Financial aid will be available to approximately one quarter of our best students in the early years. As we scale the program, the ratio of financial support may drop to no less than 15% of the total student population. In addition, as outlined in our EDI strategy (Section 1(F)), the Institute will directly offer scholarship for URM students.

8. Governance

The program is offered by the Halicioglu Data Science Institute, established as an academic unit by the Academic Senate on June 6, 2018. HDSI faculty council is the governing body of all academic programs by the Institute. A copy of Bylaws is attached in Appendix B.

9. Changes in Senate Regulations

No changes to Senate regulations are proposed.

Appendix A: Listing of Research Areas

The following table lists topical areas covered in doctoral research efforts engaging core HDSI faculty organized by seven core themes of HDSI.

AI: Automated Reasoning, Knowledge Representations, Cognition
Knowledge Representations, Distributed representations, learning multiple levels of representation or based on composing learned functions
Multi-agent Systems combined with graph signal processing, network analysis
Automated decisions, Computer augmented decision making (with applications in geospatial analysis, health)
Intelligence amplification and application to decision making, Augmented Cognition
Machine Learning: Theory, Algorithms, Systems
Adversarial ML, ML for security and privacy, algorithmic fairness
Reinforcement learning, Learning as optimization, multi-task learning, transfer learning, learning to learn
Algorithms, game-theoretic setups such as GANs, realistic study of the limits of machine learning and applied statistics
NLP, language technologies, unstructured text analysis
Accelerated ML Systems: architectures, algorithms, tools and libraries for accelerated ML systems
Data Infrastructure: Data Viz, Programming and DB Systems
Data Visualization, Visual Analytics, HCI for data science
Databases/data systems for data systems
Data mining, data integration from multiple modalities (text, time-series, imaging etc)
Methods and System design to ensure data security and privacy
Distributed/cloud computing
IOT and Cyber-Physical Systems, AI sensors
Software engineering and PL for Machine Learning, ML Systems
Mathematical Foundations of Data Science: Causal Inference, Hypothesis Testing, Optimization Theory
Causal inference in machine learning, Sequential decision making methods and their statistical analysis
Non-parametric data analysis
Multiple hypothesis testing and high dimensional data analysis, false discovery rate
Applied probability problems for the analysis of data science methods

Submodular optimization, transport theory (optimal transport/Wasserstein distance, parallel transport, especially on non-Euclidean spaces), optimization theory and algorithms that use large data to reduce computation without compromising statistical validity
Digital Humanities: DS in Society, Ethics/Policy
Ethics, data science in public interest
Philosophy of information: how data science allows us to learn about the world, information transfer from data to models, prediction/interpretation tradeoffs
Privacy and public policy, Accountability measures and methods from Data Science
Understanding humans, data science in language, literature and arts
Computational social science, data-driven sociology
Computational Linguistics, Speech versus intentional language, Conversational design and human behavior
Systems and Applications: CPS/IOT, Architectures, Health, Economics, Robotics
Brain-inspired Computing Machines, Neuromorphic Architectures, Hyperdimensional Computing
Medical signal processing, computational medicine, medical data integration challenges (patient records, device records, insurance records etc); Causal inference in Medical Informatics
Data-driven developmental economics, new economic theories based on automated data-driven measurements and methods
Statistics and economics: statistical game theory (focus on statistical and computational properties of the Nash equilibrium and its implications to fairness), market efficiency (antitrust), and other market design problems
Geospatial data collection and analysis
Robust and commonsense learning in robotic systems
ML augmented organizational workflows, data science applications in organizational behavior, business
ML methods for dynamic, time, causal inference with implications for Political Science, Economics and/or Healthcare
IOT for health, autonomous vehicles
eSciences: Real-Time Instrumentation Data, Sustainability
Environmental Data Sciences
Data science in ecology and conservation, Sustainability at scale
Data Science in Precision Imaging Systems: Applied Optical and Electron Microscopy
Data Science for High-throughput Biology, Sequencing, Mass Spectrometry in support of Bioinformatics, Quantitative Biology

Appendix B: Letters of Support Received

In addition to the three founding departments of the Data Science undergraduate program (Computer Science and Engineering in the Jacobs School of Engineering, Mathematics in the Division of Physical Sciences and Cognitive Sciences in the Division of Social Sciences) we sought and received letters from all departments and faculty interested in teaching courses in Data Science graduate program, or hosting Data Science students in their courses.

Divisional Deans & Directors

1. Steve Boggs, Dean of Physical Sciences
2. Lisa Ordóñez, Dean of Rady School of Management
3. Al Pisano, Dean of Jacobs School of Engineering
4. Peter Cowhey, Dean of Global Policy Institute
5. Carolyn L. Sandoval, Interim Director, Teaching+Learning Commons
6. Carol Padden, Dean of Social Sciences
7. Robert E. Continetti, Dean Designee, HDSI
8. Kit Pogliano, Dean, Division of Biological Sciences
9. Cristina Della Coletta, Dean Arts & Humanities

Department Chairs

1. Lei Ni, Department of Mathematics
2. Douglas Nitz, Department of Cognitive Science
3. Kun Zhang, Department of Bioengineering
4. Dean Tullsen, Department of Computer Science and Engineering
5. Cheryl Anderson, Family Medicine & Public Health (email support, letter pending)
6. Karen Messer, Biostatistics, School of Medicine (email support, letter pending)
7. Bill Lin, Department of Electrical and Computer Engineering

Faculty, Instructors

1. Jorge Cortes, MAE247
2. James Fowler, POLI 287
3. Trey Idekar, BNFO 286 / MED 283
4. Massimo Franceschetti, ECE 227
5. Vineet Bafna, CSE 283 / BENG 203, CSE 280A
6. Alex Cloninger, DSC 210, Math 170A, Math 277A
7. Arun Kumar, CSE 234, DSC 202, DSC 204
8. Rayan Saab, DSC 210, DSC 211, DSC 242, DSC 243
9. Armin Schwartzman, DSC 244, DSC 212, DSC 241, DSC 242
10. Angela Yu, COGS 243
11. Brad Voytek, COGS 280
12. Ronghui (Lily) Xu, MATH 284

13. Siavash Mirarabbaygi, ECE

External Review Letters

1. Liza Levina, Vijay Nair Collegiate Professor and Chair, Department of Statistics, University of Michigan
2. Babak Shahbaba, Professor and Chancellor's Fellow, Director of the UCI Data Science Initiative, Department of Statistics and Computer Science, UC Irvine
3. Daniela Witten, Dorothy Gilford Endowed Chair, Professor of Statistics and Biostatistics, University of Washington, Seattle
4. Zachary Chase Lipton, BP Junior Chair Assistant Professor, Machine Learning Department, Tepper School of Business, Carnegie Mellon University

Appendix C: Catalogue Copy Description [Draft]

Data Science (DSC)

All courses, faculty listings, and curricular and degree requirements described herein are subject to change or deletion without notice. Updates may be found on the Academic Senate website: <http://senate.ucsd.edu/catalog-copy/approved-updates/>.

The Graduate Program

The graduate program offers a master of science degree and a (planned) doctor of philosophy degree in data science. To be accepted into either course of study, a student should have a BS or BA degree in relevant fields or work experience in Data Science, or be able to demonstrate an equivalent competency.

Admission to the graduate program is done through the Graduate Division, UC San Diego. The application deadline is in December. Admissions are always effective the following fall quarter. For admission deadline and requirements, please refer to the departmental web page: <http://datascience.ucsd.edu>.

Admission decisions for the MS and PhD programs are made separately. A current MS student who wishes to enter the PhD program must submit a petition, including a new statement of purpose and three new letters of recommendation, to the HDSI graduate admissions committee.

Data Science Program

The field of Data Science spans mathematical models, computational methods and analysis tools for navigating and understanding data and applying these skills to a broad and emerging range of application domains. A whole range of industries – from drug discovery to healthcare management, from manufacturing to enterprise business processes as well as government organizations – are creating demand for data scientists with a skill set that enables them to create mathematical models of data, identify trends and patterns using suitable algorithms and present the results in effective manners. The target systems can be, for example, biological (e.g., clinical data from cancer patients), physical (e.g., transportation networks), social (e.g., social networks) or cyber-physical (e.g., smart grids). In all these cases, there is a combination of core knowledge in information processing coupled with the skills to abstract, build and test predictive and descriptive models that must be taught and learnt in the context of an application

domain. These application areas are in many domains served by Engineering, Physical Sciences, Social Sciences, Health & Life Sciences, and Arts & Humanities.

Master of Science Program

The goal of the masters program is to teach students knowledge and skills required to be successful at performing data driven tasks, and lay the foundation for future researchers who can expand the boundaries of knowledge in Data Science itself.

Course Requirements

There are Foundation, Core, and Elective and Research requirements for each masters program. These course requirements are intended to ensure that students are exposed to (1) fundamental concepts and tools (Foundation), (2) advanced, up-to-date views in topics central to Data Science for all students (the Core requirement), and (3) a deep, current view of their research or application area (the Elective requirement). Courses may not fulfill more than one requirement.

The master of data science program is structured as a total of 12 4-unit courses grouped into *foundational*, *core* and *specialization* areas as described below. Successful completion of the program requires completion of a Thesis or a course-based comprehensive examination that tests integrative knowledge across multiple courses. Out of the 48 units, at least 40 units must be using graduate-level courses. In addition, 2 out of 10 graduate courses can be in areas not directly related to data science but a domain specialization such as economics, biology, medicine etc upon approval of the student's faculty advisor.

Foundational Courses

These courses seek to provide five critical foundational knowledge and skills that each student graduating from the master's program is expected to receive at a graduate level: programming skills, data organization and methods skills, numerical linear algebra, multivariate calculus, probability and statistics. The program is designed so that students lacking in any (and all) of these foundational knowledge and skills can take credit for a maximum of **four courses** from the following five courses: DSC 200, DSC 202, DSC 210, DSC 211 and DSC 212.

Core Courses

These courses build upon foundational courses. All students must take **two** required core courses: DSC 240 and DSC 260. In addition, students can select at least **three** out of seven core courses: DSC 203, DSC 204, DSC 241, DSC 242, DSC 243, DSC 244, DSC 250.

Elective and Specialization Courses

Students can choose from following elective or specialization tracks a total of **three** 4-unit courses to complete course requirements.

General Elective Courses:

DSC 205, DSC 231, DSC 251, DSC 252, DSC 253, DSC 254, DSC 213
CSE 234, MATH 181 A-B-C, MATH 284, MATH 285, MATH 287A-B, COGS 243.

Specialization Areas:

Upon prior approval from graduate advisor, students can sign up for an available specialization area for an “Master of Science in Data Science with specialization in *specialization-area*” degree:

Specialization: Bioengineering

BENG 218, BENG 203, BENG 211, BENG 213, BENG 221, BENG 230A-B, BENG 276, COGS 278, PHYS 278

Specialization: Business (Marketing)

MGT 475, MGT 477, MGT 489, MGTA 455, MGTA 479

Specialization: Business (Supply Chain and Technology)

MGT 450, MGT 451, MGTA 456, MGTA 463

Specialization: Business (Finance)

MGT 407, MGTF 402, MGTF 404, MGTF 405, MGTF 406, MGTF 415

Specialization: Machine Vision and Interaction Design

COGS 202, COGS 220, COGS 225, COGS 283

Specialization: Computational Neuroscience

COGS 260, BGGN 246, BGGN 260, COGS 260 (or NEU 282), COGS 280

Specialization: Networks

MATH 261A, MATH 277A, MATH 289A, MATH 289B, DSC 205, BNFO 286, POLI 287, SIOB 276, ECE 227, MAE 247

Availability of all specializations is not guaranteed. Additional specialization areas may be added by student petition.

Thesis or Comprehensive Exam

The degree can be pursued under either the Thesis Plan I or the Comprehensive Examination Plan II. The comprehensive examination option follows a course-based comprehensive examination plan under the supervision of a comprehensive examination committee. For full-time students, all the requirements can be completed within one to two years. Students must register for a minimum of three quarters for residency requirements. To maintain good

academic standing, students must be making timely and satisfactory progress toward completion of degree requirements and must maintain a minimum overall GPA of 3.0 at UC San Diego.

Electives and Research

The number of Elective and Research units required varies by degree (see below). Electives are chosen from graduate courses in DSC, CSE, Cognitive Science, ECE and Mathematics or from other departments as approved. Please refer to the department's website for a list of approved Electives. Courses must be completed for a letter grade, except Research units that are taken on a Satisfactory/Unsatisfactory basis. Seminar and teaching units may not count toward the Electives and Research requirement, although both are encouraged.

Plan I: Thesis Option

Plan II: Comprehensive Examination

Under this plan, the student must complete a practical course-based comprehensive examination designed to evaluate the student's ability to integrate knowledge and understanding as well as utilize associated skills to produce a solution to a real-world problem. The examination requires the student to produce one or more artifacts representing a solution to the prescribed problem(s). The artifact(s) may, for example, take the form of one or more of the following: Source code, design documents, formal reasoning expressed via one or more proofs, mathematical model(s), and/or expository prose or oral presentation. More information regarding the comprehensive examination can be found in a separate document provided by the Institute.

Research

Students electing Plan I may choose to pursue a thesis research topic with an adviser while enrolled in four units of DSC 298. A maximum of four units of DSC 298 may be applied to the Electives and Research requirement.

Student with Disabilities

In order for the program to respond, a student requiring accommodation for disability must make a request for accommodation upon submission of the student's intent to apply to the Graduate Program.

Information concerning accommodation requests is available at: <http://disabilities.ucsd.edu/students/obtainaccommodations.html>. Distance learning sites must confirm their ability to support students with disabilities.



OFFICE OF THE DIVISIONAL DEAN—PHYSICAL SCIENCES
9500 GILMAN DRIVE, #0352
LA JOLLA, CALIFORNIA 92093-0352

TELEPHONE: (858) 534-6882
FAX: (858) 534-5224

April 30, 2020

To: Graduate Council

From: Dean Steven Boggs

Re: Master of Science in Data Science

Dear Colleagues,

I am writing to express my strong support for the new proposed Master of Science in Data Science degree to be offered by the Halicioğlu Data Science Institute (HDSI). The proposed program is a strong collaboration between HDSI and other units on campus, including the Mathematics Department.

I am particularly excited by Domain Area structure of the program, which enables the MS degree to serve students with many different fields of interest and will keep the program flexible to new and evolving fields. Besides the broad range of areas proposed, I can imagine multiple new domain areas relevant to Physical Sciences. The program is also well balanced in serving students interested in pursuing careers in industry or academia.

A number of proposed courses have strong overlap with existing Mathematics courses. I am encouraged that HDSI is working closely with Mathematics with the intention of cross listing these courses, coordinating instead of duplicating efforts.

I strongly support this proposal. I anticipate there will be high demand from students for this program. It would be a welcome and timely addition to the academic offerings at UC San Diego.

Sincerely,

Steven E. Boggs
Dean, Division of Physical Sciences
Chancellor's Associates Endowed Chair in Physics



Lisa D. Ordóñez, PhD
Dean
Stanley and Pauline Foster Endowed Chair
Rady School of Management

9500 Gilman Drive # 0553
La Jolla, California 92093-0553
Tel: (858) 822-0830
lordonez@ucsd.edu
rady.ucsd.edu

April 27, 2020

To: Graduate Council
From: Dean Lisa Ordóñez
Re: Master of Science in Data Science

Dear Colleagues,

Several faculty members at Rady and I have had an opportunity to review HDSI's proposed degree program for a Master of Science in Data Science. We are very supportive of the proposal and believe it will be well-received by other units on campus. This proposal is a timely one and will help address an unmet and growing need for graduate education in data science. Further underscoring this need, at Rady we currently have more than 1,300 students applying for our Master of Science in Business Analytics. Demand for our program has grown by an average annual rate of 40.5% since seating the first cohort in 2016.

There is a lot to like in the details of HDSI's proposal. First, a fundamental goal of the program is to "lay the foundation for future researchers who can expand the boundaries of knowledge in Data Science itself". This is an important aspect that will help produce capable students who can then become inputs to a wide array of doctoral programs on campus. Students with graduate training in data science would be aptly prepared for Rady's PhD program particularly in the fields of quantitative marketing, information systems, and finance. This is similarly true for many departments on campus including CSE, CogSci, Bioinformatics, Public Health, etc.

We are very interested in this possibility which brings us to a second desirable aspect of the proposal concerning domain specialization. The proposed program requires students to select a domain and take 12 units from a corresponding track. The courses that compose a track help increase a student's knowledge of the domain and importantly the application of data science *in* the domain. At Rady we intend to be a participating department and offer tracks in our most relevant sub-fields. We see the potential for a graduate student in Data Science to become excited about topics in the business domain and then desire pursuit of a doctoral degree. Successful outcomes like this would serve the research mission of the university.

Lastly, we view the Master of Science in Data Science program as complementary to our degree programs and believe the experience of graduate students focusing in the areas of data science and its applications will be positively impacted by its existence. While domain specialization ensures students in the proposed program attain breadth, I can also see how students in our MS in Business Analytics and Master of Finance degree programs would mutually benefit from some of the new graduate courses that are created as part of this proposal. While our students typically take their breadth electives within Rady, some students seek courses in departments across

campus. These electives require approval by program directors who assess fit. Our MS Business Analytics students would find course offerings such as DSC 243: Optimization, DSC 250: Introduction to Text Analysis, and DSC 260: Data Ethics and Fairness as natural options to help them achieve greater breadth.

In closing, I am supportive of the Master of Science in Data Science program presented in this proposal. It is very well thought out and designed with aspects that make it unique within UC San Diego. I anticipate that the program will be successful in achieving its goals.

Best regards,

A handwritten signature in black ink, appearing to read "Lisa D. Ordóñez". The signature is fluid and cursive, with the first name "Lisa" being the most prominent.

Lisa D. Ordóñez, PhD
Dean
Stanley and Pauline Foster Endowed Chair
Rady School of Management



ALBERT ("AL") P. PISANO, DEAN
IRWIN AND JOAN JACOBS SCHOOL OF ENGINEERING
WALTER J. ZABLE PROFESSOR OF ENGINEERING
7313 JACOBS HALL

9500 GILMAN DRIVE
LA JOLLA CALIFORNIA 92093-0403
TEL: (858) 534-6237 FAX: (858) 822-3904
EMAIL: DeanPisano@eng.ucsd.edu

05 May 2020

TO: Graduate Council

FROM: Albert P. Pisano, Dean of Engineering

RE: Master of Science in Data Science

Colleagues:

I am writing to express my strong support for the new proposed Master of Science in Data Science degree, to be offered by the Halicioğlu Data Science Institute (HDSI).

There is a "Data Science and Machine Learning" concentration, one of twelve concentrations, in the Electrical Engineering MS program offered by the Electrical and Computer Engineering (ECE) Department. Further, among related programs that partially cover some of the subject areas of the MS/DS program, the Computer Science and Engineering (CSE) Department has a MS program in Computer Science which offers elective courses in Machine Learning. Additionally, there are discussions between HDSI and the Department of Bioengineering regarding collaboration regarding a new program. I note that there are currently five jointly appointed faculty between HDSI and Engineering (or six, if counting the HDSI Director himself), and with this level of strong collaboration, I am confident that HDSI and Engineering will move forward together in a mutually-beneficial way.

I anticipate there will be high demand from students for this program, and look forward to an exciting new crop of students.

Sincerely,

A handwritten signature in black ink, appearing to read "Albert P. Pisano", with a large, stylized flourish above the name.

Albert ("Al") P. Pisano
Member, US National Academy of Engineering
Walter J. Zable Distinguished Professor & Dean
Irwin and Joan Jacobs School of Engineering
University of California, San Diego



PETER F. COWHEY

Dean, School of Global Policy and Strategy
Qualcomm Chair In Communications and Technology Policy

9500 Gilman Drive
La Jolla, CA 92093-0519
T: (858) 534-1946
pcowhey@ucsd.edu
gps.ucsd.edu

May 4, 2020

To: Graduate Council

From: Peter Cowhey, Dean, GPS

Re: HDSI Proposal for a MS in Data Science

I have reviewed the HDSI proposal in consultation with several GPS faculty. We support the proposal. I offer these brief observations about some issues that may be of specific concern to the Graduate Council.

1. *Is there any conflict with existing or planned degrees at GPS?*

We see no conflict with GPS degrees.

2. *Is the proposed degree complementary to teaching in GPS or other graduate programs on campus? Could GPS students benefit from the classes?*

It is easy for us to imagine that the HDSI courses could be of value to students in a variety of PhD programs. Whether or not they would be valuable to GPS students depends greatly on which types of preparation for admission to the HDSI degree will be required. If students are required to possess an undergraduate major/minor in computer science or similar degree program (e.g., math), then most GPS students would not benefit from the classes. If the beginning classes at HDSI assume a far more varied form of preparation, then some HDSI classes might benefit GPS students. Many of our students have a strong capacity for quantitative analysis, including big data applications.

3. *Can we identify any conflicts between the HDSI degree and planned degree programs outside of GPS?*

Our faculty flagged the possibility of a conflict with the planned Social Science MA in computational social science. However, Social Science and HDSI programs have very different starting points and therefore are probably distinctive even though they each focus on tools for data analysis. We know that Dean Ordonez feels that there is no conflict with Rady's programs.

4. *Does GPS plan to offer its own version of a data science degree or partner with HDSI?*

Neither the proposed Social Science MA nor this HDSI program gets to the more topically focused mixes of policy analysis and big data analysis. GPS students would benefit from a joint degree or certificate program with HDSI that was explicitly designed to blend applied policy analysis in fields like environmental economics/development planning with more intense courses in data science. The GPS "Big Pixel" research project already includes HDSI researchers and exemplifies how satellite observation data can be blended with existing policy analytic work to yield new policy findings and refined research tools. We hope to explore this idea with HDSI in the coming year.

Therefore, in summary, GPS is supportive of the HDSI proposal for a MS in Data Science.

Sincerely,



Peter F. Cowhey



CAROLYN L. SANDOVAL, PH.D.
INTERIM DIRECTOR, TEACHING + LEARNING COMMONS
9500 GILMAN DRIVE #0175-W
LA JOLLA, CALIFORNIA 92093-0175

May 20, 2020

To: Graduate Council

From: Carolyn L. Sandoval, Interim Director, Teaching + Learning Commons

Re: Master of Science in Data Science

Dear Colleagues:

On behalf of the Teaching + Learning Commons, I am writing to express my enthusiastic support for the Halicioğlu Data Science Institute (HDSI) proposal for a program of graduate study in Data Science. The proposal is quite impressive and the transdisciplinary nature of the program is a unique strength that will likely be attractive to many potential students. Additionally, the plan for ensuring that students have the requisite background knowledge and experience is well thought out and addresses potential barriers to access and equity.

The Teaching + Learning Commons is well positioned to assist the HDSI faculty and administration on the implementation of the data science graduate program in the following areas:

1. Support for the development of a comprehensive assessment plan focused on desired program learning outcomes;
2. Support for faculty and graduate Instructional Assistants in course design and learner-centered and equity-based teaching practices;
3. Academic support for graduate students in the program.

I look forward to working with HDSI Director Rajesh Gupta and others to support the success of the program.

Please let me know if you need additional information.

Sincerely,

A handwritten signature in cursive script that reads "Carolyn L. Sandoval".

Carolyn L. Sandoval
Interim Director
Teaching + Learning Commons

UC San Diego

May 22, 2020

TO: James Antony, Dean
Office of Graduate Studies

RE: Proposed M.S. program in Data Science


We have had a series of conversations with Director Rajesh Gupta and Associate Director Virginia De Sa about the proposed masters' of science program in Data Science. I am pleased to say that these discussions have been fruitful, and we are in full support of the proposed program.

In the Division of Social Sciences we have several majors and a masters' program that provide coursework and training in computational areas: the majors in Cognitive Science, Psychology, Linguistics, and the B.S. major in Political Science and Data Analytics and the newly approved masters' program in Computational Social Science (CSS). In addition, we have coursework across almost all departments in the division on analytics and tools for social science research involving large data sets. A few departments offer courses in the ethics of data science. A number of our faculty in the division will be providing courses for the proposed program.

Though it might seem there are similarities between the proposed program and the recently approved CSS masters', there are important differences between them: 1) the prerequisites for Data Science include more advanced computationally-related courses at the undergraduate levels than for CSS; 2) there are several specializations that will be offered in the proposed masters' program and more is planned in the future, allowing masters' students to be drawn from a number of different disciplines, and not just the social sciences; and 3) the CSS focuses on questions of social science: human behavior, human society and human interaction. We expect the 2 programs to target and attract different populations of prospective masters' students. Based on market research, we know there are students who will prefer one or the other masters' programs, depending on their training and background prior to entering graduate study and their career goals after completing the program.

It is clear that a great deal of thought has gone into this proposed program, and we like the idea of the campus having options for graduate study in computation and data science. Having several options allows each of our programs to appeal to different populations, and to attract students who are best suited for each. We welcome innovative and high-quality education on this campus, and this proposed program is a solid addition to graduate study here at UC San Diego.

Sincerely,



Carol Padden
Dean, Division of Social Sciences

Dean, Division of Social Sciences

University of California San Diego • 9500 Gilman Drive # 05020 • La Jolla, California 92093-0502

Tel: (858) 534-6073 • Fax: (858) 534-7394 • socialsciences.ucsd.edu



OFFICE OF THE EXECUTIVE VICE CHANCELLOR
ACADEMIC AFFAIRS

9500 GILMAN DRIVE
LA JOLLA, CALIFORNIA 92093-0001
Phone: (858) 534-3130
URL: <http://evc.ucsd.edu>

From: Robert E. Continetti, Sr. Associate Vice Chancellor, Academic Affairs

To: Graduate Council

RE: Support Letter for HDSI Master of Science Degree in Data Science

May 22, 2020

Dear Graduate Council,

As Dean Designee of the Halıcıoğlu Data Science Institute (HDSI), I have had the pleasure of working with the HDSI faculty throughout the discussions and consensus-building process in constructing the enclosed proposal for Master of Science Degree in Data Science.

While the end result speaks for itself, perhaps I can provide an insight into the thought process driving the conceptualization of the graduate program. The entire graduate program proposal as led by Director Gupta was driven by a singular objective to provide the most compelling learning opportunities for graduate students interested in Data Science. To do so, the proposal draws upon multiple prior and ongoing experiences in building graduate programs, some which were thankfully developed by Professor Gupta himself in his capacity as chair of Computer Science and Engineering. It also helped that Professor Gupta had led the construction of the Data Science undergraduate major and minor that is now offered by HDSI. Such a focus on student experience and faculty engagement is commendable and in doing it, the proposed program manages to engage a large number of faculty from other departments, all of which have provided a letter of support for the program.

The end result is a program that is well-structured to draw and grow talent from different areas in to Data Science, and create a talent pool to advance the field of Data Science. HDSI has successfully recruited 13 new faculty members in addition to transfer of two full-time teaching professors. In addition, 8 faculty members from other departments have transferred 25% appointment into HDSI. These together provide significant teaching capacity for being able to offer the planned graduate courses for the benefit of graduate students at UC San Diego.

I enthusiastically endorse the proposal for Master of Science Degree in Data Science.

Sincerely,

Robert E. Continetti
Sr. Associate Vice Chancellor, Academic Affairs

Cc: Director Gupta



KIT POGLIANO
Dean of Biological Sciences
Professor of Molecular Biology
Richard C. Atkinson Endowed Chair
Division of Biological Sciences

9500 Gilman Dr.
La Jolla, CA 92093-0376
Phone: (858) 822-5738
E-mail: kpogliano@ucsd.edu

September 22, 2020

Graduate Council
UC San Diego
MC: 0002

Re: Letter of Support for the MS degree in Data Science from the Halicioglu Data Science Institute.

Dear Colleagues,

I am writing to express my enthusiastic support for the proposed MS degree in Data Science from the Halicioglu Data Science Institute, which will provide a pathway to Data Sciences for students graduating with degrees in Biological Sciences and other disciplines. This will fill a major need in both academia and industry for individuals who work at the interface of Computer Sciences and Biological Sciences, who will drive discovery by develop methods to mine the larger and larger data sets collected in many fields, including multi-omics, imaging, neurobiology, and medicine, and to connect and analyze diverse data sets. Students in this program will have a unique opportunity to work in an interdisciplinary setting with students and faculty different disciplinary backgrounds, providing them with invaluable experience that will prepare them for a wide array of careers. I am confident that this program will be a resounding success.

As Dean of the Division of Biological Sciences, I look forward to working more closely with the Halicioglu Data Science Institute in the coming years. Last year, I worked with my faculty and chairs to develop a six year strategic plan that prioritized growth in quantitative and computational biology for faculty hiring and curricular development. Indeed, we have already launched this initiative, and recently hired five ladder rank and teaching professors in this area (Yonatan Aljadeff, Mikio Aoi, Marcus Benna, Barry Grant, Ashley Juavinett), including a cluster of four Computational Neurobiology faculty, one whom is jointly appointed with HDSI (Aoi). I look forward to resuming our plan to jointly recruit faculty in this area, possibly this year, if we can identify appropriate joint recruits, or as soon as the hiring freeze is lifted. We also look forward to working with HDSI in our ongoing efforts to integrate computational analysis, bioinformatics, and programing into our curricula by developing new courses and modules for students at an array of levels (from Freshmen to PhD students), and new degree specializations. This effort has already started, and we are grateful to HDSI, especially to Eric Mjoen, who is working with my faculty and staff to develop collaborative projects that will bring together students from HDSI and Biological Sciences to work in interdisciplinary teams on academic and industrial projects. I am confident that this is the start of a productive partnership that will benefit our research, our students and the biotechnology and biomedical communities.

The outstanding faculty that HDSI has already recruited alone and in partnership with other units on campus, the success of their existing programs, and the need for additional talent in this area, make me confident that this degree program will be a success. I am therefore pleased to endorse this proposal and look forward to working with HDSI in the coming years on this and in other areas of mutual interest.

Sincerely,

A handwritten signature in blue ink that reads "Kit Pogliano". The signature is fluid and cursive, with the first name "Kit" written in a slightly larger, more prominent font than the last name "Pogliano".

Kit Pogliano, PhD

October 7, 2020

To: Rajesh K. Gupta, Director, Halicioğlu Data Science Institute (HDSI)
From: Cristina Della Coletta, Dean Arts & Humanities
RE: Proposed Master of Science (M.S.) Degree in Data Ethics

Dear Professor Gupta:

I am very pleased to offer my support for the creation of a new Master of Science (M.S.) degree in Data Science at UC San Diego.

The M.S. in Data Science will enhance UC San Diego's ability to train graduate students in this evolving and dynamic transdisciplinary field. The program is strategically designed around multiple specialization tracks, allowing students from diverse academic background to develop shared core competencies while also supporting individual selections of domain-elective courses.

As noted in the Executive Summary: "the goal of the program is to teach students knowledge and skills required to be successful at performing data-driven tasks." This is the type of training that is increasingly important in many other graduate programs. From the perspective of the Division of Arts & Humanities, the M.S. in Data Sciences has a strong potential to benefit students interested in pursuing doctoral programs with emphasis in the areas of digital humanities, practical or applied ethics, history of science, and environmental and medical humanities, among other options. Of course, the M.S. in Data Sciences will also serve as a selection tool for the envisioned doctoral degree program in Data Science.

The proposal for the M.S. in Data Science is well-argued and meticulously presented. It makes a compelling case for attracting and supporting diverse student populations. I believe the M.S. degree program in Data Sciences will provide a welcome addition to graduate studies at UC San Diego.

I look forward to seeing this program take off, and to further opportunities of collaboration between the Division of Arts and Humanities and HDSI.

Sincerely,



Cristina Della Coletta
Dean, Arts & Humanities



LEI NI, CHAIR
DEPARTMENT OF MATHEMATICS, 0112
9500 GILMAN DRIVE
LA JOLLA, CALIFORNIA 92093-0112

Phone: (858) 534-1177
Fax: (858) 534-4932
Email: leni@ucsd.ed

April 25, 2020

To: Professor Rajesh Gupta, Director, HDSI, UC San Diego

Re: The Master Program at HDSI

I am writing to express the Mathematics Department's support for the new proposed Master of Science in Data Science degree to be offered by the Halicioğlu Data Science Institute (HDSI) in collaboration with various academic units on the UC San Diego campus.

The proposed program will provide students with the knowledge and skills for a career in Data Science for industry, civil service or academia. As mathematics and statistics are integral components of the interdisciplinary field of Data Science, and indeed many UCSD Mathematics Department faculty are affiliated with HDSI, the department is pleased to see the further connections that this proposal will make with the Mathematics Department. In particular,

- (a) several Mathematics Department undergraduate courses can be used to satisfy background requirements (Math 170A, Math 173A and Math 181A, with Math 180A as a prerequisite, can be used by students to satisfy foundational requirements),
- (b) several core DSC courses (DSC 240, DSC 242 and DSC 244) are likely to be cross-listed with equivalent Math Department courses, especially when taught by Mathematics faculty appointed or affiliated with HDSI, and
- (c) several Mathematics Department graduate courses are listed as possible domain-elective courses for students to choose from in pursuing their specializations.

Many of the courses in (a) already have significant contingents of Data Science undergraduate majors enrolled in them. We expect to accommodate the MS in Data Science students in these as well and to also accommodate these students in the graduate courses falling under (b) and (c).

The Mathematics Department looks forward to cooperating with HDSI on this program to further catalyze connections and collaborations related to data science.

Lei Ni, Chair

Department of Mathematics
LN:mr



Douglas A Nitz, Professor and Chair
Department of Cognitive Science, 0515
9500 Gilman Drive
La Jolla, CA 92093-0515

April 29, 2020

To: Professor Rajesh Gupta, Director, HDSI at UCSD

Re: Data Science Masters Program

I am writing on behalf of the Cognitive Science Department in full support of the Masters Program in Data Science associated with HDSI at UCSD. The merits of the program have been under evaluation by Cognitive Science since early stages of the program's development, in particular through Virginia de Sa and Bradley Voytek.

As presented, the program is described in great detail and is very thorough in its approach, finding an excellent balance between the needs of student populations of importance to the University and the capabilities of HDSI and collaborating departments. The program is unique in bringing together the very best that UCSD has to offer educationally in this rapidly growing field of research and engineering. The design clearly reflects the power of combining data science practice to the highly varied scientific problems addressed in cognitive science.

Data science is highly relevant to multiple sub-disciplines in cognitive science and so the program is synergistic with our own efforts to prepare the next generation of professionals. There is no doubt that the program will represent an important opportunity to many our students upon their graduation and that the presence of masters students in this field. Cognitive Science is prepared to take on HDSI Data Science Masters students in several of our graduate courses including: Cogs 278 (computation on genetics), Cogs 220 (information visualization), Cogs 225 and 283 (visual processing and machine learning). Cogs 202 (computational models for cognition) is in development in collaboration with HDSI. The faculty responsible for teaching these courses are excited to accept a talented group of HDSI Masters students and to continue the valuable collaboration between Cognitive Science and HDSI.

We look forward to helping make the program a success.

A handwritten signature in cursive script that reads "Doug Nitz".

Douglas A. Nitz, Professor and Chair
Department of Cognitive Science
University of California, San Diego



KUNZHANG, PH.D.
PROFESSOR
CHAIRMAN
DEPARTMENT OF BIOENGINEERING
9500 GILMAN DRIVE 0412
LA JOLLA, CALIFORNIA 92093-0412

TELEPHONE (858) 822-7876
FAX (858) 534-5722
E-MAIL: k4zhang@ucsd.edu

May 4, 2020

To: Professor Rajesh Gupta,
Director, HDSI at UCSD

Re: Data Science Masters Program

I am writing on behalf of the Department of Bioengineering to enthusiastically endorse and support the MS program in Data Science associated with HDSI.

As you are aware, we are in the process of launching a MS Program in Bioengineering with specialization in Biomedical Data Science and Engineering and your planned program dovetails extremely well with our proposed program.

Given the vast importance of Data Sciences in modern society, it is imperative that we train qualified professionals who can join the workforce solving problems where big data is the paradigm. I have reviewed your proposed program and the design of the curriculum is excellent and will be ideal for training students.

The proposed MS Program specialization will benefit from your courses and we anticipate that students in your MS Program will take courses in our Department. I am also excited that we have an outstanding joint hire in Ben Smarr (newly appointed Asst Prof, jointly between HDSI and Bioengineering) who will serve as a bridge between our programs. I should also add that my colleague Dr. Shankar Subramaniam is launching a new course on Biomedical Data Sciences in the academic year 2020-21, which would be valuable to the two new MS Programs in BE and HDSI.

Several other courses offered by Bioengineering including graduate courses in technologies that generate vast data in biomedicine and complex modeling courses that transform data into knowledge would be valuable for our Programs.

I look forward to working with you and helping make our MS Programs harbingers of the future.

Best regards,

Kun Zhang
Professor and Chair
Department of Bioengineering

A handwritten signature in black ink, appearing to read 'Kun Zhang'.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
OFFICE OF THE CHAIR

9500 GILMAN DRIVE
LA JOLLA, CALIFORNIA

May 4, 2020

To: Professor Rajesh Gupta, Director, HDSI

Re: Data Science MS

Dear Dr. Gupta,

On behalf of the department of Computer Science and Engineering, I would like to express my strong support of this proposal to create a graduate program in Data Science. While Computer Science and Data Science are closely linked fields, and in fact have some significant overlap, I note that HDSI has worked with CSE from the beginning to ensure that we provide a distinct educational opportunity for our students, first at the undergraduate level and now at the graduate level. In fact, I should note that several of our growing list of co-appointed faculty (four now) have been instrumental in helping create this proposal.

Artificial Intelligence and Machine Learning are, and likely always will be, fundamental Computer Science disciplines. However, the quickly-emerging field of Data Science, which also incorporates foundations from statistics and math, computer science theory, Cognitive Science, etc., and addresses a new set of timely problems across any number of disciplines, is well served by a concentrated focus on this domain. This is true at the undergraduate level, but likely even more true at the graduate level, both in the course curriculum and the research.

Again, the overlap between the two areas has only been positive to this point. CSE has hired faculty that make us stronger in Data Science, and HDSI has already hired several people that improve our presence in computer science.

This proposal has sought to minimize impact on our over-subscribed AI courses. It specifies one undergraduate course as a prerequisite, and one graduate CSE course as a general elective. To this point, HDSI has worked with us to help cover those courses and we expect this level of cooperation and shared vision to continue. We may in fact see some relaxing of the heavy demand on some of our CSE graduate AI courses (coming from non-majors) as HDSI gets their offerings in place.

As a department, we feel that this (adding graduate programs) is the natural progression for HDSI and Data Science as an important discipline at UCSD. This was our expectation when we supported the formation of HDSI and when we helped launch the DSC undergrad program. HDSI has been allowed to hire faculty, and have already attracted several with strong research credentials. Preventing them from attracting graduate students directly interested in Data Science would be a huge blow to the program and the faculty they have attracted.

Like some other units, we feel that we have been partners in the growth and direction of Data Science at UCSD, and HDSI, and couldn't be more pleased to see them take this next natural step.

Sincerely,

A handwritten signature in blue ink, appearing to read "Dean Tullsen". The signature is fluid and cursive, with a long horizontal stroke at the end.

Dean Tullsen

Professor and Chair

Dept of Computer Science and Engineering

Subject: Re: HDSI Graduate Programs

From: "Anderson, Cheryl" <c1anderson@health.ucsd.edu>

Date: 5/1/2020, 9:34 AM

To: Rajesh Gupta <gupta@eng.ucsd.edu>, "Messer, Karen" <kmesser@health.ucsd.edu>, "Continetti, Robert" <rcontinetti@ucsd.edu>

Dear Rajesh,

You can count on our support. Karen and I will send a joint letter of support.

Best,

Cheryl

From: Rajesh Gupta <gupta@eng.ucsd.edu>

Sent: Friday, May 1, 2020 2:10 AM

To: Messer, Karen <kmesser@health.ucsd.edu>; Continetti, Robert <rcontinetti@ucsd.edu>; Anderson, Cheryl <c1anderson@ucsd.edu>

Subject: HDSI Graduate Programs

Dear Karen, Cheryl

I attache the latest Graduate Program proposal for MS in Data Science by HDSI is currently going through the review process. We have received a number of support letters. I attach a sample for review.

It will help to get a letter of support pointing out joint appointment of Armin and future engagements that benefits students from additional course offerings.

Regards,

Rajesh



PROF. BILL LIN
TEL: (858) 822-1383
E-MAIL: billin@eng.ucsd.edu

ELECTRICAL & COMPUTER ENGINEERING
9500 GILMAN DRIVE, MAIL CODE 0407
LA JOLLA, CALIFORNIA 92093-0407

DATE: May 7, 2020

TO: Graduate Council

FROM: Bill Lin, Chair, Department of Electrical and Computer Engineering

RE: Master of Science in Data Science

Dear Colleagues,

It is my pleasure to write this strong letter of support for the new proposed Master of Science in Data Science degree, to be offered by the Halicioğlu Data Science Institute (HDSI).

Demand for data scientists is clearly exploding in industry, as data science is being applied in all aspects of society. The proposed MS program in HDSI is very timely to serve these market needs. The proposed HDSI MS program complements our existing "Machine Learning and Data Science" specialization in the MS program offered by my department of Electrical and Computer Engineering (ECE) and will provide our ECE students additional course offerings in important areas. Further, the proposed HDSI MS program will facilitate closer engagements and collaborations between the faculty in ECE and HDSI. A number of ECE faculty are already affiliated with HDSI, and we expect to pursue joint recruitments in the future in areas that fall at the intersection of ECE and HDSI.

Overall, the proposed HDSI MS program will undoubtedly bring much greater visibility to UC San Diego as the preeminent university for artificial intelligence, machine learning, and data science. I look forward to cooperating with HDSI on this program as well as other initiatives.

Best regards,

A handwritten signature in black ink, appearing to read "Bill Lin".

Bill Lin, Chair
Electrical and Computer Engineering Department
University of California, San Diego



MECHANICAL AND AEROSPACE ENGINEERING
LA JOLLA, CALIFORNIA 92093

+1 (858) 534-0708

Mechanical and Aerospace Engineering
Jacobs School of Engineering
University of California
9500 Gilman Dr
La Jolla, California 92093
PHONE: +1 (858) 822-7930
EMAIL: cortes@ucsd.edu
April 29, 2020

To: Professor Rajesh Gupta, Director HDSI
Re: MAE247: Cooperative Control of Multi-Agent Systems
From: Jorge Cortes, Professor, Mechanical and Aerospace Engineering

Dear Rajesh,

This is to confirm that I support the listing of the graduate course “MAE247: Cooperative Control of Multi-Agent Systems” as an elective for the HDSI Masters program under the Networks specialization and warmly welcome qualified graduate students in the course.

Sincerely,

A handwritten signature in blue ink, appearing to read "Jorge Cortes".

Jorge Cortés
Professor

DEPARTMENT OF POLITICAL SCIENCE, 0521
TELEPHONE: (858) 534-6807
FAX: (858) 534-7130



9500 GILMAN DRIVE
LA JOLLA, CALIFORNIA
92093-0521

April 29, 2020

Dear Colleagues:

To: Professor Rajesh Gupta, Director HDSI
Re: POLI 287: Multidisciplinary Methods in Political Science: Social Networks
From: James Fowler, Professor

This is to confirm that I support the listing of the above course as an elective for the HDSI Masters program under the Networks specialization and welcome qualified graduate students in this course.

Sincerely,

A handwritten signature in cursive script that reads "James H. Fowler".

James H. Fowler
Professor
University of California, San Diego
fowler@ucsd.edu

UC San Diego Health

UC San Diego Health
Department of Medicine

9500 Gilman Drive MC-0688
La Jolla, CA 92093-0688
T: +1 858.822.4558
F: +1 858.534.4246
tideker@ucsd.edu
idekerlab.ucsd.edu

Trey Ideker, Ph.D.

Professor of Medicine

Adjunct Professor of
Bioengineering and Computer
Science

Director, NCI Cancer Cell Map
Initiative (CCMI)

Director, NIGMS National Resource
for Network Biology (NRNB)

Director, NIMH Psychiatric Cell
Map Initiative (PCMI)

April 29, 2020

Professor Rajesh Gupta, Director HDSI
UC San Diego
rgupta@ucsd.edu

Re: BNFO 286 / MED 283: Network Biology and Biomedicine

Dear Dr. Gupta,

This is to confirm that I support the listing of the above course as an elective for the Halicioğlu Data Science Institute (HDSI) Masters program under the *Networks* specialization and welcome qualified graduate students in this course.

Sincerely,



Trey Ideker, Ph.D.



MASSIMO FRANCESCHETTI
PROFESSOR OF ELECTRICAL ENGINEERING
9500 GILMAN DRIVE, MC 0407
LA JOLLA, CALIFORNIA 92093-0407

PHONE: (858) 822-2284
FAX: (858) 534-2486
E-MAIL: massimo@ece.ucsd.edu

April, 30, 2020

To: Prof. Rajesh Gupta, Director HDSI
Re: ECE227: Big Network Data
From: Massimo Franceschetti, Professor, ECE

This is to confirm that I support the listing of the above course that I teach for the machine learning and data science graduate program in ECE at the Masters level as an elective for the HDSI Masters program under the Networks specialization and welcome qualified graduate students in this course.

With kindest regards

Massimo Franceschetti

A handwritten signature in cursive script that reads "Massimo Franceschetti".



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY OF CALIFORNIA, SAN DIEGO
THE IRWIN AND JOAN JACOBS SCHOOL OF ENGINEERING

9500 GILMAN DRIVE
LA JOLLA, CALIFORNIA 92093-0404
FAX: (858) 534-7029

May 4, 2020

To: Prof. Rajesh Gupta, Director HDSI

This is to confirm that I support the listing of the following courses that I teach as electives for the HDSI MS program under the Bio specialization and welcome qualified graduate students in this course.

1. CSE280A: Algorithms for population genetics
2. CSE283/Beng203

Sincerely,

A handwritten signature in black ink that reads "Vineet Bafna".

Vineet Bafna, PhD

Professor, CSE, #4218

UC San Diego

9500 Gilman Drive

La Jolla, CA 92093-0404

vbafna@cs.ucsd.edu

858-822-4978 (O)

858-534-7029 (F)

<http://www.cs.ucsd.edu/~vbafna>



DEPARTMENT OF MATHEMATICS
9500 GILMAN DRIVE, #0112
LA JOLLA, CALIFORNIA 92093-0112

+1 (858) 534-4889
FAX +1 (858) 534-5273
EMAIL acloninger@ucsd.edu

May 3, 2020

Dear Colleagues,

To: Professor Rajesh Gupta, Director HDSI

This is to confirm that I support the listing of the course *Math 277A: Topics in Computational and Applied Math: Diffusion Geometry and Metric Graph Learning* as an elective for the HDSI Masters program under the Networks specialization, and welcome qualified graduate students in the course.

Sincerely yours,

A handwritten signature in black ink, appearing to read "Alex Cloninger".

Alexander Cloninger, Ph.D.
Assistant Professor of Mathematics and
Halıcıoğlu Data Science Institute
University of California, San Diego



BRADLEY VOYTEK
ASSOCIATE PROFESSOR
UC SAN DIEGO

9500 GILMAN DR.
LA JOLLA, CA
92093-0515

2020 May 01

To: Professor Rajesh Gupta, Director, Halicioğlu Data Science Institute
Re: COGS 280: Neural Oscillations
From: Bradley Voytek, Associate Professor

This letter confirms my support for including the above class, COGS 280: Neural Oscillations, as an elective for the Halicioğlu Data Science Institute Master program, **Computational Neuroscience** Specialization Area.

I look forward to working with students from HDSI!

Sincerely,

Bradley Voytek, Ph.D.

UC San Diego
Department of Cognitive Science
Halicioğlu Data Science Institute
Neurosciences Graduate Program

Cc:
Professor Doug Nitz, Chair, Cognitive Science
Jennifer Morgan, MSO, HDSI

May 1, 2020

Yian Ma
Assistant Professor
Halicioğlu Data Science Institute
University of California, San Diego
La Jolla, CA 92023

Subject: HDSI Master Program Proposal

Dear Rajesh Gupta,

This is to confirm that I support the listing of courses including DSC 211, DSC 240, DSC 243, and DSC 251 for the HDSI Masters program and welcome qualified graduate students in this course.

Sincerely,

Yian Ma



SIAVASH MIRARABBAYGI
ASSISTANT PROFESSOR OF ELECTRICAL AND COMPUTER ENGINEERING
9500 GILMAN DRIVE, MC 0407
LA JOLLA, CA 92093

PHONE: (858) 822-6245
E-MAIL: smirarabbaygi@ucsd.edu

May 4, 2020

To: Prof. Rajesh Gupta, Director HDSI
Re: ECE208: Computational Evolutionary Biology
From: Siavash Mirarab, Assistant Professor, ECE

This is to confirm that I support the listing of the above course that I teach for the graduate program in ECE at the Masters level as an elective for the HDSI Masters program under any appropriate specialization and welcome qualified graduate students in this course.

With kindest regards

Sincerely,

A handwritten signature in black ink, appearing to read "Siavash".

Siavash Mirarab

Liza Levina
Vijay Nair Collegiate Professor and Chair
Department of Statistics, University of Michigan
<http://dept.stat.lsa.umich.edu/~elevina/>

323 West Hall
1085 South University
Ann Arbor, MI 48109-1107
734-764-3235

September 23, 2020

Dear colleagues,

It is my pleasure to write a letter of support for the new Master's in Data Science program proposed by HSDI at UC San Diego. The demand for degrees in data science at all levels, and especially Master's, is enormous, and job placement for graduates is spectacular, even in the midst of the pandemic-induced recession. Our Master's program in Data Science at the University of Michigan is now in its third year, hosted by the Department of Statistics and run jointly with the Department of Electrical Engineering and Computer Science in the College of Engineering, the Department of Biostatistics in the School of Public Health, and the School of Information. In the last admission cycle, we received 960 applications for the class of about 50. The students that apply to our Data Science program come from a wide range of undergraduate majors, work experiences, and backgrounds; notably, they are more diverse and include more domestic students than applicants to our traditional Master's in Statistics program.

The proposal developed by HSDI is well thought out, and does a good job balancing the needs for rigorous foundations with the exceptional breadth of potential data science applications. I expect that the program will both challenging and accessible, and will attract a broad range of students. I enthusiastically support the proposal. Please do not hesitate to contact me should you need more information.

Sincerely,



Liza Levina
Vijay Nair Collegiate Professor and Chair of Statistics



DONALD BREN SCHOOL OF INFORMATION AND COMPUTER SCIENCES
DEPARTMENT OF STATISTICS
DONALD BREN HALL 2224
IRVINE, CALIFORNIA 92697-1250
TELEPHONE (949) 824-0623

October 4, 2020

Dear Dr. Rajesh Gupta,

I am writing this letter to express my enthusiastic support for your new Master of Science program in Data Science at UC San Diego. As the Director of Data Science Initiative at UCI and the Vice Chair of the Graduate Affairs at the Department of Statistics, I truly believe that such programs are crucial for tackling the emerging data-intensive problems in science and engineering.

As you indicated in your proposal, currently there is a significant shortage of highly skilled data scientists in the U.S. workforce. To address this issue, we need to train a new generation of data scientists with interdisciplinary skills and a strong background in mathematics, statistics, and computer science, as well as the ability to collaborate with other scientists and practitioners. These students will be in huge demand in future. According to LinkedIn's 2017 U.S. Emerging Jobs Report, machine learning engineers, data scientists, and Big Data engineers rank among the top emerging jobs¹. More specifically, the State of California is expected to be one of the largest hubs for data science careers. This is primarily due to the booming economy and market dynamics of the Silicon Valley area, and the fast-growing data science units in the south, mainly in the Los Angeles, Orange County, and San Diego areas. HDSI is at the forefront of this data revolution and one of the few places in the country capable of leading this movement. The rapid growth of your undergraduate program from 50 majors to 206 since 2017 shows the success of your effort so far to harness the growing interest in data science. I have no doubt that your future MS program will be as successful. It would attract a group of students who desire a deeper knowledge of data science techniques and concepts. Such students tend to be more mature and well-prepared for taking leadership roles in industry. Your program can also serve as a gateway for students who might be interested in PhD level education in data science related fields.

I carefully studied the specific aspects of your proposed program. This is a very well-thought out and skillfully crafted proposal. It targets all major application domains of data science. You have put together a very impressive group of core and affiliated faculty members. Your goal of recruiting 15-20 students in the first year of the program is quite realistic; this is roughly the number of MS students we admit to our program every year. I also think that your plan for including 4 foundational courses in your program is essential and absolutely necessary. As you mentioned, students come to these interdisciplinary programs from diverse educational backgrounds. It is crucial to make sure they have the right foundations to be successful academically and professionally. At UCI, we have been dealing with the same issue. After seeing that some of our graduate students struggle with our first-year courses, we decided to offer a week-long bootcamp for our incoming students each year. We try to cover similar topics (but obviously not at the same depth) as your foundational courses. Also,

¹ <https://economicgraph.linkedin.com/research/LinkedIns-2017-US-Emerging-Jobs-Report>

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

similar to your proposed program, we encourage our students to take elective courses related to their domain of interest.

Overall, establishing this new program is a great idea, and based on your proposal, I have no doubt it would be one of the most successful programs in this field. Because of the geographical proximity of our two institutes, I would welcome any opportunity to collaborate with your faculty and contribute to the success of HDSI. Please do not hesitate to contact me if I could be of any help.

Regards,

A handwritten signature in cursive script, appearing to read "Babak Shahbaba".

Babak Shahbaba, Ph.D.
Professor and Chancellor's Fellow
Director of the UCI Data Science Initiative
Vice Chair of Graduate Affairs
Departments of Statistics and Computer Science
University of California, Irvine
babaks@uci.edu

Daniela Witten
Dorothy Gilford Endowed Chair
Professor of Statistics & Biostatistics



Prof. Rajesh Gupta
Professor and Qualcomm Endowed Chair
Department of Computer Science & Engineering
University of California, San Diego

September 30, 2020

Dear Prof. Gupta,

I have reviewed with interest the proposal for a Master's of Science in Data Science degree program put forth by the Halicioglu Data Science Institute (HDSI) at University of California, San Diego. I am writing to express my enthusiastic support for this program.

In 2014-2015, I led the development of the University of Washington's Master of Science in Data Science program. This program officially launched in 2016, and is now beginning its 5th year. This highly interdisciplinary professional Master's program is a truly collaborative endeavor, as it is hosted by the Departments of Statistics, Biostatistics, Computer Science and Engineering, Applied Math, Human Centered Design and Engineering, and the Information School. The program is designed to expose the students to all of the key aspects of data science: from statistics to software engineering to data ethics.

Data science is consistently ranked as one of the most in-demand jobs nationally and internationally, and the demand for data scientists is expected to continue to increase in the coming years. In light of this demand, it is not surprise that our program has experienced extremely high demand and rapid growth, with almost one thousand applicants per year in recent years. These applicants have varied backgrounds, undergraduate majors, and past work experience, and include both international and domestic applicants. I expect that the proposed MS program put forth by HDSI will experience comparably high demand and growth.

I congratulate HDSI on a very strong proposal for a new MS program. It balances rigorous foundations with key applications. I expect the program to provide a broad range of students with the fundamentals required for a successful career in data science. HDSI is well-positioned to lead this program, in light of its expertise in the disciplines involved.

I wish HDSI all the best with what expect will be a highly-successful program.

Sincerely,

A handwritten signature in black ink that reads 'Daniela Witten'.

Daniela Witten, Ph.D.
Dorothy Gilford Endowed Chair
Professor of Statistics and Biostatistics
University of Washington
dwitten@uw.edu



Zachary Chase Lipton
BP Junior Chair Assistant Professor
Machine Learning Department
Tepper School of Business
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
USA

September 23, 2020

Support Letter for HDSI MS in Data Science Degree Program

To whom it may concern,

It is my **great pleasure** to support the initiative to launch a new Masters of Science in Data Science (MS/DS) Degree Program Halıcıoğlu Data Science Institute (HDSI). By way of introduction, I am a graduate of the UCSD Computer Science and Engineering Department, where I completed my MS and PhD at CSE in 2015 and 2017, respectively. Currently, I am the BP Junior Chair Assistant Professor of Operations Research and Machine Learning at Carnegie Mellon University, where I direct the [Approximately Correct Machine Intelligence \(ACMI\) Lab](#). The lab consists of 1 postdoctoral researcher, 10 PhD students, and several more collaborators and visiting researchers, working on the key problems that tend to be overlooked when Machine Learning is applied carelessly in the wild. These include questions about generating actionable insights (capable of guiding decisions), robustness under distribution shift, and alignment between the ML systems and various social desiderata, including fairness and transparency.

I believe that the proposed MS program will be an extremely valuable addition to UCSD's current offerings. While the program has clear connections with several existing programs (e.g. the MS in CS and the MS in bioinformatics), it is clearly separated and will provide an ideal degree for a number of students whose needs may be overlooked by the current offerings. Specifically, this will be the first degree focused entirely on data science. Unlike a traditional CS MS degree, this program will enable student to focus the majority of their education (and crucially, course load) on data science and its applications, providing an opportunity to dive deep into these areas that may not be possible while simultaneously focusing on programming languages, systems, engineering, and theory. Moreover, with its diverse interdisciplinary focus, the program will provide a unique offering for many students to have direct contact with diverse application areas. Perhaps most importantly, with its core focus on the ethical and societal dimensions of deployed ML systems, this degree will prepare students for careers at the

burgeoning intersection of data science and policy. I also expect this degree to prepare many students for data science careers embedded in various other industries, and perhaps even public facing careers as data journalists come to play key roles in shaping our understanding of key events, including the current coronavirus pandemic.

While I was a student, HDSI Director Rajesh Gupta (then Chair of the CSE Department) led CSE through a time of unprecedented growth, leaving the department academically stronger, housed in a revamped facility, and on sound footing financially. I am confident both that this program will make a terrific offering to the university and that Rajesh has the vision and expertise to execute the program as planned.

If there is anything else that I can do to support this program, please do not hesitate to call me at 914-260-5590 or email me at zlipton@cmu.edu.

Best Regards,

A handwritten signature in black ink, appearing to read 'Zachary Lipton', with a horizontal line extending to the right.

Zachary Chase Lipton