Academic Advisory Committee on Social and Ethical Implications of Computing and Data Science

Year 1 Annual Report

September 19, 2024

Advisory committee members:

David Harding, *Professor of Sociology (Chair)*
Matilde Bombardini, *Associate Professor of Business*
Sandrine Dudoit, *Associate Dean, Faculty and Research, College of CDSS (ex officio)*
Marta Gonzalez, *Associate Professor of City and Regional Planning*
Jennifer Johnson-Hanks, *Executive Dean, College of Letters and Science (ex officio)*
Ayesha Mahmud, *Assistant Professor of Demography*
Tejas N. Narechania, *Professor of Law*
Rasmus Nielsen, *Professor of Integrative Biology*
Dara O'Rourke, *Professor of Environmental Science, Policy, and Management*
Alex Saum-Pascual, *Associate Professor of Contemporary Spanish Literature and New Media*

Table of Contents

1. Background and Orientation

The “and society” advisory committee was appointed by – and reports to – the EVCP, with the following responsibility: “to advise the EVCP on how CDSS can fulfill its commitment to the study, understanding and teaching of the social and ethical implications of computing and data science. The advisory council will make recommendations to the EVCP concerning an integrated approach to faculty hiring, research, graduate curriculum, and undergraduate teaching programs that it believes CDSS should implement individually or jointly with faculty in other schools and colleges, as well as advise the EVCP on collaborations and synergies that it believes CDSS and other colleges and schools should pursue…the committee is not limited to considering the social and ethical implications of computing and data science solely in the context of CDSS. If it envisions opportunities that serve Berkeley’s research and teaching in those areas outside of CDSS, it is free, as well, to provide advice about those opportunities to the EVCP.”

The above charge is very broad, and the work of this committee is expected to last five years. Therefore, this committee has focused on general recommendations of strategies, structures, and initiatives regarding research and education in the social and ethical implications of computing and data science within and outside of CDSS. We leave to those charged with implementation the details of exactly how any particular recommendation would be carried out. We emphasize that, ultimately, efforts to further incorporate “and society” into CDSS and to build out capacity across campus at the intersection of computing and data science and the social sciences and humanities will be the collaborative work of faculty both in CDSS and across campus.

In parallel with the work of this committee, a wide-ranging and diverse set of faculty from across campus initiated a process of proposing two “and society” departments within CDSS, one named “Human Technology Futures” and another that incorporates multiple divisions focused on the application of data science in different domains, including (most relevant to this report) Computation and Data Science for Economics and Business, Computation and Societal Policy, and Societal-Scale Data Infrastructure. Our committee has provided feedback on those proposals and endeavored to stay informed of the department formation process (indeed, multiple members of this committee have contributed to that process). However, given the breadth of this committee’s charge, it has (a) left the details of department planning and implementation to faculty directly involved in that process, and (b) explicitly avoided being limited in its own proposals or recommendations by the department formation process or proposal. While some of the recommendations and proposals below could (and should) be implemented by the new department(s) in formation, others may be best implemented within other administrative structures, both within and outside CDSS. However, we are optimistic that the formation of those departments within CDSS will serve as a strong foundation for fully incorporating “and society” in the medium to long term and will serve as a catalyst to jumpstart such efforts. The formation of these new departments provides an opportunity to think through the meaning and implementation of “and society” in research and curriculum.

Also in parallel to the work of this committee, CDSS has adopted College Bylaws that include an Undergraduate Study Committee with representation from each of the majors in CDSS and each of the three College Essential Skills, including "Human and Social Dynamics of Data and Technology." In addition, Berkeley Division Bylaw 73 provides for the inclusion of "And Society" faculty members in the college faculty via Articles 2 and 4–6, and these colleagues could serve on the Undergraduate Study Committee, among other roles. Although it is not explicitly within the charge of this committee to address faculty governance, we note that the specific inclusion of "and society" faculty in CDSS governance through these measures as well as the new departments is an important foundation for further progress.

During the first year of its work, this committee focused on (i) developing a conceptual framework specifying what it believes "and society" entails, (ii) identifying challenges and making recommendations for undergraduate teaching, and (iii) cataloging both existing structures on campus that facilitate research at the intersection of data science and the social sciences and humanities, as well as what peer universities are doing at this intersection. Further recommendations around research and around graduate curricula and public engagement are on the agenda for future years (see section 10).

A primary aspect of this year's work was taking stock of what other institutions are doing at the intersection of data science and the social sciences and humanities (see sections 6 and 8). While this work is incomplete and ongoing, one preliminary conclusion is that Berkeley's incorporation of "and society" into the core of CDSS and other data science initiatives on campus is relatively unique and has the potential to position UC Berkeley as a leader in the integration of the social and ethical implications of AI and big data into computing and data science research, teaching, and public engagement.

2. Framework

Before any progress can be made, the exact meaning of "and society" needs to be clearly defined and agreed upon. This is central to identifying the goals of a college with "and society" in its name and building consensus around the "and society" aspects of its mission. As the committee began its work, it became clear that no single definition or scope was provided at the time of the founding of CDSS, and that many different faculty on campus had very different ideas of what "and society" was supposed to mean. Therefore, the committee's first task was to more clearly identify what "and society" might mean for CDSS. The committee endeavored to create a framework that is broad and inclusive, encompassing multiple different perspectives while being specific enough to provide clear direction for future efforts.

We propose a framework consisting of the following five pillars:

1. Social science and humanistic perspective on research methods: Social scientists and humanists have distinct, but complementary, approaches to rigorous research design relative to statisticians and computer scientists (although there is also considerable overlap between statistics and the social sciences in this regard due to their long history

of cross pollination and collaboration). One way to think about these approaches is that they represent decades of work on how to apply the scientific method to research questions regarding social, economic, and cultural life. From social science, we would highlight the iterative relationship between theory and inductive and deductive modes of data analysis; the links between theory, concepts, and measurement; representativeness, external validity, and case definition and selection; integration of qualitative and quantitative analysis; and multiple perspectives on causal inference. Social scientific approaches also emphasize understanding mechanisms (how and why questions) and context (variation across populations and settings). For example, in studying the pandemic, whereas data scientists tended to focus on forecasting, social scientists tended to focus on understanding the social and institutional contexts through which the pandemic spread and health care was delivered. From the humanities, we would emphasize the need to contextualize these theoretical and applied approaches within scientific and human history, understanding that any technology is a manifestation of human experience, spanning from individuals to entire cultures and societies, from the past to the present. Understanding technological development within the world of languages, histories, and cultures, fosters social justice and equality by revealing how different peoples have tried to make moral, spiritual and intellectual sense of the world. Finally, a humanities-informed approach to data and technology augments statistics and data science courses that teach students to deal critically and logically with complex, subjective and imperfect information about the world, encouraging the questioning of the nature, origins, and intent of any given methodology, data set, or technology.

2. Science and Technology Studies (the "science of science"): Undergraduates who intend to work in scientific fields, especially those with implications for society and policy, should have some training in science and technology as social phenomena and social practice. Social scientists and humanists provide intellectual frameworks for understanding topics such as science as a profession, the creation of scientific knowledge, the impact of science and technology on society, politics, and material well being, the economics of science and the role of science in the economy, the role of science in public policy, among many other topics. Science and technology studies provide us the capacity to learn from past episodes of scientific and technological innovation and apply those lessons to best harness new developments while mitigating potential harms.
3. Social and ethical implications of artificial intelligence and big data: Training the next generation of responsible data scientists requires firm grounding in the Human Contexts and Ethics of data science, including social scientific, humanistic, and normative intellectual tools. From issues of algorithmic bias to data privacy to surveillance capitalism and beyond, the promise of technological progress in AI and big data is also fraught with fundamental risks relating to power, inequality along multiple dimensions, human agency, climate change, and institutional change. Navigating and mitigating these risks requires knowledge of the historical dynamics and patterns that have brought us to the current technological moment as well as intellectual frameworks for careful analysis of problems of social and ethical impacts. One central aspect of this is the regulation or governance of AI, which has two dimensions: (1) how we regulate AI / data science (e.g., auditing for bias, or regulating competition among providers, or copyright

for training data); (2) the implications of AI use in core government functions (e.g., law enforcement uses of AI, using AI to make policy or to assess public input).

4. Creative Applications of Technology: The interest in teaching creative applications of technology is twofold. On the one hand, digital technology has revolutionized the art and design world, enabling artists to create complex visuals, interactive narratives, and generative poetry and music, among a myriad of other artistic disciplines. Learning about the histories and artistic traditions behind digital arts allows for a reconceptualization of digital technology itself. On the other hand, teaching creative applications of digital technologies enhances our understanding of digital literacy (i.e., all the social practices and conceptions engaged in meaning making via digital codification), moving beyond standardized operational practices (e.g. content-based teaching, uncontextualized task performance, repetition of skills, etc.) towards a competency-based learning model that stresses the capacity to "learn how" rather than "learn about."
5. Domain Knowledge and Subject Matter Expertise: Data scientists apply their methods and tools in many social domains, from health to environment to criminal legal systems and beyond. Formulating research questions, interpreting data, and developing implications requires a deep understanding of the substantive domain in which the work is being done. In contrast to the methodological emphasis of pillar #1, this pillar emphasizes domain-specific knowledge and subject matter expertise (e.g. the causes and consequences of racial and ethnic inequality, the drivers of political polarization, or the psychology of climate-change denial). Social scientists and humanists have long standing research programs and deep knowledge of almost all domains of social life. Most methodological development is domain informed and motivated, in that methodologists develop new methods to solve specific scientific problems. Perhaps more importantly, the next generation of data scientists needs general and robust conceptual tools and practices for integrating existing knowledge into their work as data scientists, no matter the domain in which they are working at any moment.

3. Undergraduate Curriculum Challenges

Before we discuss challenges, we begin by noting that UC Berkeley is at the forefront of undergraduate data science education, not just in the United States but internationally. Indeed, a 2018 National Academy of Sciences report recognized Berkeley's model as a prototype for other institutions.[1] Since the founding of what was then called the "Data Science Education Program" in the (then) Division of Data Science in 2015, and indeed since the first version of Data 8 was piloted as early as 2012, Berkeley faculty from CS, Statistics, STS, and many other disciplines have built out a world-renowned curriculum in data science that reaches to almost every corner of campus and engages more undergraduates than almost any other program on campus. In addition to existing majors and minors in CS and Statistics and the DS major and minor, faculty and staff both within and outside CDSS provide "connector" courses linking other disciplines to Data 8, "push-in modules" to bring data science into courses in other disciplines

[1] National Academies of Sciences Engineering & Medicine, Data Science for Undergraduates: Opportunities and Options. Washington D.C.: The National Academies Press, 2018.

across campus, a CDSS Discovery program to engage undergraduates in research. The EECS and Statistics faculty also teach introductory courses like Data 8, Stat 2, and CS 10 that are available to students with minimal technical backgrounds, meet the requirements for many majors and minors on campus, and, in the case of Stat 2, teach statistical thinking without formula, coding, and technical details..

Of particular note for this report, "and society" is already represented in many ways in these programs. For example, the DS major and minor require upper-division training in Human Contexts and Ethics, every DS major must take one lower division and two upper division courses in a "domain emphasis" – many of which are in the social sciences and humanities, Statistics majors complete a similar "Applied Cluster" requirement of three upper-division courses – many of which are also in the social sciences and humanities, and all CDSS majors must complete Breadth requirements identical to those in L&S as well as an essential skills requirement in "Human and Social Dynamics of Data and Technology." The "and society" components of the DS education programs is one of the features that makes it distinctive in the wider higher education landscape.

Nevertheless, Berkeley has always maintained its excellence through constant innovation and improvement, so this committee discussed four potential areas of curricular improvement with respect to data science and the "society" part:

1. Many social science and humanities courses are currently disjoint from the many technical courses that CS, DS and Stat majors take. We need courses to "bridge" these otherwise hard to reconcile sets of concepts. It is too hard a problem for students to tackle on their own. For example, oftentimes DS majors with a domain emphasis related to the humanities or social sciences do not have courses that make explicit links between domain knowledge and the technical skills they learn in their other courses for the DS major. Addressing this issue will likely require faculty in the social sciences and humanities with the relevant technical expertise to design or redesign courses such that disciplinary and technical knowledge are interwoven, ideally through collaboration or consultation from CDSS faculty. For example, the new Data 6 course, developed by a team of faculty from CS, Statistics, and the social sciences in collaboration with colleagues at other universities, is an example of integrating data science, social science, and ethics into a single course.

2. Relatedly, the undergraduate programs in CDSS would benefit from tighter linkages between "and society" concepts and technical concepts. Where "and society" courses exist and/or are required, they stand alone without strong linkages to other major requirements, especially in the upper division. As we discuss further below, faculty on campus are beginning to develop models of how to integrate "and society" into core technical courses in the CDSS majors /minors. Additionally, "and society" courses in CDSS would benefit from an integration of data science and statistics concepts into them.

3. CDSS students should engage in creative ways of thinking that augment the creativity learned in their technical courses (e.g., exploratory data analysis, visualization, open-ended modeling). The study and creation of digital art and literature has the potential to teach creative applications of digital technologies that can support and generate new ways of thinking about societal problems and new ways of asking questions (pillar #4).

4. For humanities and social sciences majors there is a structural lack of access to more technical courses. The courses that are currently offered are too advanced for them to successfully absorb and require a sequence of prerequisites that is too long. While CDSS introductory courses (e.g. Data 6, Data 8, STAT 2, STAT 20, CS 10) are relatively accessible and provide an important foundation, by themselves they are insufficient to train non-majors in the use of data science tools and methods in their home disciplines. For example, a social science major who would like to understand how to apply machine learning to social science problems must currently take a long sequence of technical courses not otherwise required for their major (and sometimes already at capacity), close to the equivalent of minoring in DS. Filling this gap in the application of DS tools and methods will require faculty in the social sciences and humanities to offer applied courses in their disciplines, again, ideally through collaboration or consultation with CDSS faculty.

4. Undergraduate Curriculum Recommendations

We propose four strategies for improving the representation of the five pillars in the framework above into the undergraduate curriculum, both within and outside CDSS.

1.To better integrate the social and ethical dimensions of computing and data science into the undergraduate curriculum, we recommend creating specialized “Society Modules” on the essential skills of “Societal Reasoning for Technology and Data.” Such modules, sometimes termed “embedded X” in the CS and DS teaching literature, can be integrated into existing core required courses for the CS, Statistics and DS majors. These modules will introduce ideas, topics, and case studies from the five pillars defined above and will be designed to fit with the existing course syllabus. They will be an integral part of the course, and the topics covered will be considered as one of the core learning objectives, rather than optional or additional material.

2. To better integrate statistical and computational reasoning into social sciences and humanities major programs, we recommend creating specialized statistical and computational reasoning modules to embed in core courses in disciplines outside of CDSS. These modules would both help CS/DS/Stat students bridge what they learn in their technical courses with the application of those skills and methodologies to social sciences and humanities research questions, and they would help social science students bridge what they learn about the social dimensions of their field to real-world applications and their limitations and consequences. These modules will introduce ideas and topics from data science and statistics (related to pillars

#1 and #3) in the context of case studies (pillar #5) and will be designed to fit with the existing course syllabus. They will be an integral part of the course, and the topics covered will be considered as one of the core learning objectives, rather than optional or additional material.

3. Create non-compulsory, but well-staffed courses that make statistics and coding accessible to humanities and social sciences majors that are interested, but currently discouraged by the gap in preparation and time commitment they need to overcome to take courses currently offered. These courses would be more applied than major requirements in existing CDSS majors and likely taught in disciplinary departments outside of CDSS, perhaps in collaboration with CDSS faculty.

4. Create non-compulsory, but well-staffed courses that "bridge" computing, statistics and data science, their applications to social science or humanities problems, and the ethical and social issues surrounding those applications. Such courses could serve majors both within and outside CDSS.

5. Create/identify courses that explore digital poetics, aesthetics and rhetorics where students engage in creative applications of their technological skills. These courses can focus on the creative application of tools or skills, but can also engage in historicizing and contextualizing digital technologies within art and literary histories.

We have identified several key components, considerations, possible next steps, and ideas for implementation of these strategies:

a. We see the module approach as being less disruptive, compared to developing entirely new courses, for faculty teaching loads, potential additional hires to teach new courses, student course loads, and course scheduling. It is also less resource intensive than creating new courses. This approach works within the current structure, and is also a way to foster dialogue and collaboration amongst faculty in different disciplines within CDSS and UC Berkeley more broadly, while also exposing students to a wide range of faculty from diverse areas they might not encounter otherwise. In the longer term, the experience of jointly developing society modules will lead to co-teaching of courses by faculty with complementary expertise.

b. One way to jumpstart many of these efforts would be to convene a series of teaching workshops for interested faculty from across campus to collaborate on curriculum development and to learn the various pedagogical techniques and course management infrastructures in use across campus. As an example, efforts by Berkeley faculty in HCE, Statistics, and CS provide models of how to collaborate, successfully develop, and integrate Society Modules into existing DS courses, including Stat 20, Data 8 and Data 100 as well as the DS pedagogy course for UGSIs (Data 375).[2] These faculty emphasize the importance of long-term partnerships, institutional resources, and training course staff, who play a major role in developing and refining course materials in large CS, Stat, and DS courses.

[2] Cathryn Carson and Ari Edmondson (HCE/History), Lisa Yan (EECS), Andrew Bray (Statistics)

c. Identify existing technical courses in the CS, Statistics, and DS majors that would be well-suited for integrating a module on society. Examples of potential existing courses where a module could be added are Data 100, 101, 102, and EECS 16A (for an example of this integration in Data 100, see this 2021 Presidential Chair Fellows Project). The “Society Modules” need to be clearly identified and advertised in the course catalog, course descriptions, departmental website, etc. If certain courses already include components that satisfy this requirement, we can promote this feature of the course and clearly identify it as a course that offers an “and society” perspective.

d. Identify existing courses in the social sciences that would be well-suited for integrating a module on statistical and/or computational reasoning. A starting point would be those courses that incorporated DS-modules created as part of the DS Modules program (no longer in operation) that integrated DS concepts into non-DS courses. The modules need to be clearly identified and advertised in the course catalog, course descriptions, departmental website, etc. If certain courses already include components that satisfy this requirement, we can promote this feature of the course and clearly identify it as a course that offers a data science perspective.

e. For module development, create a list of social science and humanities consulting faculty members to closely collaborate with technical course instructors to develop “Society Modules,” and, similarly create a list of CS/DS/Stat consulting faculty members available to collaborate with “and society” faculty interested in developing data science modules for their courses. Note that the roster of “society module consultants” would include a wide range of social scientists and humanists.

We envision the consulting faculty member being involved in all aspects of the module development – designing curricula and material, integrating the module with existing course material, training teaching assistants, fostering diverse perspectives, equity and inclusion in the way the materials are presented and taught, etc. Consultants would work closely with instructors during this development stage, and would also ideally join the instructors to co-teach the module on its first iteration. Consultants and instructors can also collaborate with teaching experts at the university to ensure smooth integration and pedagogical coherence.

f. We suggest that the module creation be compensated by offering course release for faculty during the development stage (or the stipend equivalent to account for the extra labor). Likewise, faculty creating new courses would be compensated, for example, in terms of salary or course release. Rules for allocating teaching credit or other forms of compensation would need to be developed to make this proposal practical.

g. Instructors developing new bridge courses will be offered support (e.g., support from faculty members in the newly formed departments in CDSS or other paid instructors) and resources for course development. Such courses would help students to bridge what they learn in their technical courses with the application of those skills and methodologies to social sciences and humanities research questions. This might require revamping current social science/humanities courses or hiring new faculty who are able to teach these “bridge” courses. Currently circulating

ideas to develop upper-division data science “connector” courses could be one way to realize this proposal. Such connectors might link other courses to DATA 100, for example.

h. We recommend a staged implementation with a pilot course in the next academic year.

5. Existing Curricular Resources at UC Berkeley

These resources provide a foundation on which to build new curricula to better integrate “and society” into CDSS and data science into other social science and humanities disciplines.

- Data Science Modules Team, github repo
- D-Lab workshops and github repo of workshop materials
- CDSS Discovery Program
- Human Contexts and Ethics Program, and HCE toolkit
- Course materials from new MA in Computational Social Science (Social Science Division)
- Digital Humanities (including summer only minor and certificate program)
- UC Berkeley DataHub
- Additional ideas from our Fall Brainstorming Document, including graduate curricula that could perhaps be adapted for undergraduates.

6. Comparative Perspectives: Data Science “and Society” in the Curriculum at Peer Universities

Summary:

- We considered how data science teaching (and to some extent research) is organized at peer institutions
- We looked at the presence of data science majors and minors and whether a “society” class was taught
- Most universities we considered have either zero or one courses in their Data Science major/minor/option that pertains ethical and societal implications of data science
- We also considered whether DS centers/institutes have a society component and whether they have faculty appointed specifically to the center (as opposed to a collection of faculty from other departments)
- The following table was compiled as a summary and to the best of our knowledge
- If a cell is missing, it means we were not able to determine Y/N with confidence

	MIT	Michigan	Harvard	UW	Pitt	NYU	CMU

Major DS	N	Y	N	N	Y	Y	N
Minor/Option DS	Y		N	Y		Y	N
Dedicated DS classes (vs stat/math?)	N	N	N		Y		N
Required “And society” class in major	N	N	N	Y	Y	Y	N
DS faculty in one unit?	N	N	N	N		Y	N
Humanists included in DS center?	N	N	N	N	N	Y	N
Social Scientists included in DS center?				N		Y	
Institutes/Center DS	Y	Y	Y	Y	Y	Y	N
STS dept/program	Y	Y	Y	Y	N	Y	Y

MIT

1. MIT Statistics + Data Science Center https://stat.mit.edu/
    a. Interdisciplinary PhD programs in Statistics and …(economics, poli sci)
    b. Minor in statistics and data science - no “society” course requirement
2. IDSS (MIT Institute for Data Science and Society)
    a. PhD Program in Social and Engineering Systems
3. AI + Society as a research area within EECS
4. Program in Science, Technology and Society
    a. Second major UG in STS
5. MIT Subjects:
    a. Course IDS - Data, Systems and Society
    b. Course STS - Science, Technology and Society
    c. Many of these courses coincide with courses already offered in other departments

U Michigan

- MIDAS Michigan Institute for Data Science https://midas.umich.edu/
    a. 5 research pillars https://midas.umich.edu/research-pillars/ of which only one is related to society
        - *Measuring and improving society*
            - Unstructured Data for Social Science - “MIDAS collaborates with the AI Lab and the Institute for Social Research to offer research

connection meetings to connect experts for unstructured data (text, image, video, etc) and domain researchers who plan to address significant research questions using such data”
        - Supporting the development of new data and their access - “MIDAS supports and collaborates with our faculty and campus units who develop new data sources and data infrastructure for social science research, and enable the wide adoption of such new resources.”
- Teaching:
    a. Graduate Data Science Certificate Program (open to all graduate students)
        - Areas of certificate (none related to “society”):
            - Analysis Methods (AM):
            - Data Management (DM)
            - Algorithms and Applications (AA)
    b. UG Major in Data Science
        - EECS + Statistics as in Berkeley
        - Could not find any course related to “Society” in core

Harvard

1. Harvard Data Science Initiative - Specific programs within umbrella that relate to “society”
    a. Bias^2: Bias2 Program supports research, features speakers, and engages the data science community towards using data science to uncover bias, and understanding and combating the use of badly-conceived data science that can reinforce bias and inequity
    b. SPUDS (Summer Program for Undergraduates in Data Science): ten-week summer program for undergraduate students doing research
    c. Trust in Science: in collaboration with the Harvard Kennedy School’s Program on Science, Technology & Society (STS). At a time of seemingly widespread loss of confidence in science and expertise, the Project seeks to illuminate the varied factors that currently impede trusting relations between the producers and users of scientific information.
    d. Faculty affiliates from other departments, no dedicated faculty
2. Harvard Kennedy School program on Science, Technology & Society: enhancing the quality of research, education, and public debate on the role of science and technology in contemporary societies. Among the fields that significantly contribute to the STS Program’s core mission are science and technology studies, anthropology, comparative politics, history, government, law, and sociology.
    a. STS offered as a special field to PhD students in Public policy
    b. STS offered a secondary field to PhD students in other disciplines
3. Teaching:
    a. Master of Science in Data Science: includes AC221 Critical Thinking in Data Science

b. Could not find UG program in data science
c. Can add Data Science as secondary field to any PhD program in the Graduate School of Arts and Sciences (one of possible 3 elective is AC221 above)

U Washington

1. Teaching
    a. Master of Science in Data Science
        - Requires Data 512: Human-Centered Data Science
        - Collaboration between 6 units: Applied Mathematics, Biostatistics, EECS, Information School, Statistics and Human Centered Design and Engineering
    b. Advanced Data Science specialization within PhD in Computer Science
    c. UG:
        - Option in Data Science for CS majors must take SOC 225 Data & Society seminar required
        - Option in Data Science for Bioengineering majors must take Societal Implications of Data Science: SOC 225, BH 201, BH 311, or BH 444
        - Other Options in Data Science require taking: INFO 350 “Information Ethics and Policy”, STAT 303 “Introduction to the Ethics of Algorithmic Decision Making”
    d. Master of Science in Data Science
2. eScience Institute
    a. Data Science support (similar to D-Lab?) “Our data scientists and researchers act as matchmakers who help researchers across UW and beyond find the most appropriate tools and technology for their work.”

U Pittsburgh

1. Teaching:
    a. Data Science Major
        - School of Computing and Information + Statistics and Mathematics Department
        - Required course CS 0590 Social Implications of Computing Technology
    b. Professional Master’s Data Science Online
2. Responsible Data Science RDS@Pitt
    a. Organization of research and teaching mission here
    b. Examples of initiatives under umbrella:
        - Sustainable Social Computing Lab
        - Pitt Disinformation Lab

NYU

1. Center for Data Science

    - Interdisciplinary faculty appointed specifically to center
    - AI, Misinformation, and Policy Seminar Series
2. UG teaching
    - Major and minor in data science
    - DS-UA 202 Responsible Data Science (more focused on legal, compliance, transparency, etc)
3. Dedicated PhD and Master's in Data Science

Carnegie Mellon

1. Department of Statistics and Data Science
    - Primarily statistics department
2. UG teaching:
    - No specific degree in data science (only Statistics or Statistics+Machine Learning, etc.)
    - No "society" mandatory courses
3. Master of Science in Applied Data Science
    - No "society" classes in standard curriculum
4. PhD programs in Statistics+Public Policy (not explicitly data science)

7. Existing Campus Research Initiatives and Structures

We first reviewed CDSS plans for research related to the social and ethical implications of computing, data science, and artificial intelligence. From the documents we reviewed, CDSS had developed limited plans to date for further integrating "And Society" research into the college, at least with regard to a comprehensive plan. CDSS proposal documents and "listening tour" notes did comment on extensive work already occurring at Berkeley, but did not lay out clear plans for connecting, integrating, or supporting this work going forward. Since then, CDSS has supported the development of the department proposals discussed above (Section 1), both of which represent significant research initiatives.

We then surveyed research going on across the Berkeley campus. We were interested in research themes, methods, and individuals leading this work. We organized over 20 research initiatives within the Committee's overall framework for defining the "And Society" component of CDSS's mission. We discuss on-going research initiatives across campus within the five categories that comprise our Committee's overall framework for defining the "and society" component of CDSS's mission. Our attempt to categorize these various research efforts is best described as a first effort, as there may be additional initiatives that we have overlooked, and we recognize that some of these initiatives may span several components of our framework.

Nevertheless, we think this survey is useful in both understanding the nature and breadth of existing research initiatives across campus

Social science and humanistic perspectives on research methods (Pillar #1)

Initiatives advancing this research include:

- Public Interest Technology | CDSS at UC Berkeley:
    - “design, build, and govern new technologies in ways that advance the public interest”
    - “promote greater understanding of core ethical, political and societal dimensions of technological and social change requires collaborations among researchers and educators in a wide range of fields, including computing and data science, humanities, social sciences, public policy, and law”
- Human Technology Futures (berkeley.edu)
    - “colleagues at UC Berkeley across the arts, humanities, and interpretive social sciences who are committed to shaping how computing and data science engage with human concerns and how our own fields take up and transform the opportunities that new technologies present”
- Berkeley Initiative for Transparency in the Social Sciences (bitss.org)
    - “ improve the credibility of science by advancing transparency, reproducibility, rigor, and ethics in research.”

Science and Technology Studies (the “science of science”) (Pillar #2)

Initiatives advancing this research include:

- Berkeley Program in Science and Technology Studies :: Center for Science, Technology, Medicine, & Society
    - “Berkeley program in STS brings together a diverse community of scholars studying the origins, growth, and consequences of scientific and technological knowledge and practice. Focusing on the changing conditions of knowledge production today, cutting-edge theoretical and conceptual inquiry, and engagement with public policy”
- Algorithms in Culture - Center for Science, Technology, Medicine, & Society
    - “We explore the implications of [algorithms] and deployment in politics, media, science, organizations, culture, and the construction of the self”

Social and ethical implications of artificial intelligence and big data (Pillar #3)

Initiatives advancing this research include:

- Human Contexts and Ethics | CDSS at UC Berkeley
    - “training the next generation of responsible data scientists–while rethinking the meaning of responsible data practice”
- Berkeley Artificial Intelligence Research Lab
    - “supporting an inclusive community for researchers across AI and social science disciplines to advance understandings around theories and practices for responsible and equitable AI. The initiative explores innovations for creating more responsible data, models and management approaches that enable us to better support a more inclusive and equitable society”
- Center for Human-Compatible Artificial Intelligence (humancompatible.ai)
    - “develop the conceptual and technical wherewithal to reorient the general thrust of AI research towards provably beneficial systems.”
    - “refocusing AI away from the capability to achieve arbitrary objectives and towards the ability to generate provably beneficial behavior”
- CITRIS and the Banatao Institute (citris-uc.org)
    - the CITRIS Policy Lab supports interdisciplinary research, education, and thought leadership to address core questions regarding the role of formal and informal regulation in promoting innovation and amplifying its positive effects on society
- AI, Platforms, and Society Center - Berkeley Law
    - “Artificial Intelligence, Platforms, and Society Center offers a unique forum for students, academics, practitioners, and technology companies to explore technical and governance strategies to support responsible technology development and use, including the role of the private sector and state, federal, and international laws, regulations, and policies in shaping these strategies”
- Center for Technology, Justice, and the Courts - Berkeley Law
    - “explore the evolving justice issues caused by new technologies.”
- Technology and Work - UC Berkeley Labor Center
    - “ensure that AI and other digital technologies benefit rather than harm workers”
- Algorithmic Fairness and Opacity Group - Berkeley
- Kavli Center for Ethics, Science, and the Public
    - “providing an inclusive, multi-disciplinary framework for understanding the ethical implications of science and technology”

Creative Applications of Technology (Pillar #4)

Initiatives advancing this research include:

- Berkeley Center for New Media

- Matrix Exhibitions | BAMPFA
- Digital Humanities | (berkeley.edu)
    - “thoughtful application of digital tools and methodologies to humanistic inquiry.”
- Language and AI | Language Center (berkeley.edu)
    - Berkeley Language Center is proud to support scholarship and exploration around how language and AI deeply impact the future of our world

Domain Knowledge and Subject Matter Expertise (Pillar #5)

Initiatives advancing this research fall into two broad categories. Some begin with some specific domain expertise, such as in law or environmental science, and consider ways to bring data science techniques to bear on those domains. Others begin with data science, and aim to make data science methods more easily applied to a range of domains. We include both below: the first five examples are of the former; the last two of the latter.

- Center for Cultural Analytics | Berkeley Institute for Data Science (BIDS)
    - “ develops and refines computational methods as its members interrogate cultural production across a wide range of disciplines. The research will focus on the data-driven analysis of cultural phenomena.”
- Computational Research for Equity in the Legal System Training Program (CRELS) | Berkeley Institute for Data Science (BIDS)
    - “intersection of the studies of inequality, criminal justice, data science, and the social implications of artificial intelligence (AI) and big data”
- DSE | The Eric and Wendy Schmidt Center for Data Science & Environment at Berkeley
    - “combines the power of computing and environmental science with open science principles and a commitment to inclusivity—all towards the purpose of building tangible, replicable, and accessible solutions to problems compromising the health of our environment. We aim to co-create these solutions in direct partnership with those who hold the knowledge and expertise of their local needs and environmental context.”
- Fisher Center for Business Analytics (berkeley.edu)
    - “advances the creation of business value from data science”
- CEGA - CEGA (berkeley.edu)
    - “improve the lives of people living in poverty by generating insights and tools for decision-makers backed by rigorous, inclusive, and transparent research
- D-Lab (berkeley.edu)
- EPIC Data Lab | UC Berkeley
    - “The mission of the EPIC Data lab is to democratize data work via no-code and low-code interfaces, leveraging scalable AI-powered program synthesis techniques, targeting collaboration across a range of personas and teams.”
    - “ the EPIC Data lab has a special emphasis on applications in social justice.”
- Center for Healthcare Marketplace Innovation
    - “We are developing AI and machine learning tools that reflect our deep understanding of healthcare system economics – the incentives, regulatory

environment, behavioral biases, and information structure – that makes problem-solving in healthcare inherently complex."

- Center for the Theoretical Foundations of Learning, Inference, Information, Intelligence, Mathematics and Microeconomics at Berkeley (CLIMB)
    - "established to address new conceptual and mathematical challenges arising at the interface between technology, science, and society"
- Center for Computational Biology
    - Development and application of computational and statistical methods to investigate biological systems, at the molecular, cellular, organism, and population level, from model organisms to humans.
    - Also an Augmented Graduate Group that supports a PhD and Designated Emphasis
- Computational Precision Health
    - Development and application of computational and statistical methods to improve the quality, efficiency, and equity of medicine and public health.
    - Also an Augmented Graduate Group that supports a PhD and Designated Emphasis
    - Joint with UCSF

8. Research Initiatives and Integrations at other Universities

Harvard Data Science Initiative. The Harvard Data Science Initiative (HDSI) represents Harvard's commitment to shaping the new science of data. It illuminates the new interdisciplinary pathways that faculty, students, and partners will use to solve real problems, in a world with critical ethical challenges regarding facts, data, and truth.
Observations:

- Includes focal topics such as "Bias^2: Using data science to uncover bias and addressing bias in data science".
- A cross-disciplinary center rather than a department.
- While dominated by STEM faculty, it includes faculty from the humanities and social sciences.

Stanford Data Science. The goal of Stanford Data Science is to weave data science research and methods into the University's fabric, giving our faculty and students the tools, skills, and understanding they need to do the cutting-edge research that will drive 21st-century scholarship and education. Learning from data is hard, but the future of knowledge creation and our university is at stake.
Observations:

- Includes Data Science for Social Good summer program.
- A cross-disciplinary center rather than a department.

- Heavily STEM dominated but includes a few faculty from humanities and a few social scientists.

University of Washington eScience Institute. The eScience Institute empowers researchers and students in all fields to answer fundamental questions through the use of large, complex, and noisy data. As the hub of data-intensive discovery on campus, they lead a community of innovators in the techniques, technologies, and best practices of data science and the fields that depend on them.
Observations:

- Seems focused on undergraduate education.
- A cross-disciplinary center rather than a department.
- Heavily STEM dominated but includes a few faculty from humanities and a few social scientists.
- Includes Data Science for Social Good summer program.

University of Michigan Michigan Institute for Data Science (MIDAS). MIDAS strengthens University of Michigan's preeminence in Data Science and Artificial Intelligence, and enables their transformative use in a wide range of research disciplines to achieve lasting societal impact.
Observations:

- A cross-disciplinary center rather than a department.
- While heavily STEM dominated, it also includes humanities and social science including researchers AI Ethics and AI Justice
- Education in data science is housed in Computer Science and Engineering and in the Department of Statistics not in MIDAS.

MIT Institute for Data, Systems, and Society. The mission of IDSS is to advance education and research in state-of-the-art analytical methods in information and decision systems, statistics and data science, and the social sciences, and to apply these methods to address complex societal challenges in a diverse set of areas such as finance, energy systems, urbanization, social networks, and health. As part of the Schwarzman College of Computing, IDSS spans all five schools at MIT, embracing the collision and synthesis of ideas and methods from analytical disciplines including statistics, data science, information theory and inference, systems and control theory, optimization, economics, human and social behavior, and network science. Also houses MIT Statistics and Data Science Center. The Statistics and Data Science Center is an MIT-wide focal point for advancing research and education programs related to statistics and data science.
Observations:

- A cross-disciplinary center rather than a department.
- STEM dominated but also includes faculty from humanities and social sciences.
- The "And society…" approach includes elements such as the Initiative on Combating Systemic Racism aimed at using big data to develop and harness computational tools that can help effect structural and normative change towards racial equity.

The University of Chicago Data Science Institute (DSI). The DSI executes the University of Chicago's bold, innovative vision of Data Science as a new discipline by advancing interdisciplinary research, partnerships with industry, government, and social impact organizations, and holistic data science education.
Observations:

- A cross-disciplinary center rather than a department.
- STEM focused and includes quantitative social scientists.
- Hosts a research initiative for Internet Equity.

Caltech does not appear to have a campus-wide data science initiative as such. However, they support an Information Science and Technology (IST) initiative seeking to seed and incubate new fields at the interface of computing and information sciences with other fields — to invent new "CS+X" fields.
Observations:

- The IST was founded on 2004 and funded by the Gordon and Betty Moore Foundation which also founded centers on (1) Biological Circuit and Design, (2) Social and Information Sciences Laboratory, (3) the Physics of Information, and (4) the Mathematics of Information.
- Hosts a number of other centers including Center for Social Information Sciences which focuses on studying the consequences of interactions between economics and computer science and trains the next generation to reason about this interface.
- Humanities does not seem to be represented.

Princeton Center for Statistics and Machine Learning is Princeton University's focal point for data science education and research on campus.
Observations:

-Running since 2018
-It has an Undergraduate Certificate Program in Statistics and Machine Learning
-7 core faculty mostly from EECS and 45 Affiliate faculty
-Research areas span biology, engineering and the humanities, share faculty with the Center for Digital Humanities

The Center for Digital Humanities at Princeton is an interdisciplinary research center working at the intersection of the humanities and technology. They create and apply digital tools to humanistic questions and critically engage with the promises and risks that technology poses to society.
Observations:

-Launched in 2014, 14 Affiliates, 11 students

The Yale Institute for Foundations of Data Science (FDS) is a hub for interdisciplinary collaboration. Launched in spring 2022, FDS connects data scientists with researchers across Yale. Together, their members inspire new possibilities, develop new approaches to solving problems, and advance foundational data science research. "What our institute is doing is really

more about developing core data science, as opposed to thinking about applying it within specific disciplines," its said faculty director
Observations:

-Launched in 2022
-Hosts events, supports a postdoctoral program
-One faculty director, 72 affiliated faculty spanning statistics, computer science, policy, science, management and sociology
-Research areas span biology, engineering and the humanities, share faculty with the Center for Digital Humanities
-No degree programs, no clear "and society" focus

The Cornell Center for Data Science for Enterprise and Society aims to unify programs and curricula in data science with an initial emphasis on questions grounded in data that are generated by human activity, including computational social science (e.g., sociology and government), the economics/computer science interface, aspects of digital agriculture in the production and management of agriculture, digital platforms supporting urban infrastructure (e.g., the sharing economy), and as a theme that is cross-cutting in many of these areas, the corresponding issues of privacy, security, and fairness
Observations:

-7 Research Professors for three years appointments
-1 faculty director and and a Executive Committee

The University of Pennsylvania Warren Center for Network & Data Sciences fosters research and innovation in interconnected social, economic and technological systems.
Observations:

-Two directors
-100 affiliated faculty
-No degree programs

Columbia University - Data Science Institute (DSI) at Columbia University strives to be a "force for change". They advance the state-of-the-art in data science; transform all fields, professions, and sectors through the application of data science; and ensure the responsible use of data to benefit society.
Observations:

-400 affiliated faculty
-43 core faculty
-Ph.D. with a specialization in Data Science, it spans Applied Mathematics, Computer Science, Electrical Engineering, Industrial Engineering and Operations Research, and Statistics departments, lacking Social Sciences
-M.S. in Data Science is jointly offered in collaboration with the Graduate School of Arts and Sciences' Department of Statistics, and The Fu Foundation School of Engineering and Applied Science's Department of Computer Science and Department of Industrial Engineering and Operations Research
-Executive Education

-Certificate in Data Science

Duke - Social Science Research Institute at Duke University brings together researchers with interests in problems that cross the various social and behavioral sciences, including problems that connect with the humanities and natural sciences.We promote multidisciplinary collaboration among such scholars as they work on important social issues that are challenging to address fully from within any given discipline.
Observations:

-No Data Science focus
-Launched in 2003
-3 faculty, one director
-Master in Interdisciplinary Data Science

9. Preliminary Research-Related Conclusions and Recommendations

Berkeley has an impressive scope, breadth, and depth of relevant research already underway across campus. We surveyed 25 initiatives, and there are likely more! This is often cutting edge research with real-world import and impact. Many of these initiatives, unsurprisingly, span several pillars of our analysis framework. It also appears that there is some overlap among these programs and initiatives.

One recommendation we have is to further this assessment by identifying a lead person or program to document and connect these initiatives, and to explore whether there are opportunities to combine forces, where relevant, to maximize economies of scale and impact. For example, it might make sense to have one interdisciplinary center on the “Governance of AI,” rather than having three or four centers spread across campus. However, advancing this type of interdisciplinary research will require either (1) responding to interdisciplinary incentives; (2) getting better at valuing work from outside a core discipline; or (3) some combination of both. Of course, there are trade-offs between coordination and efficiency on the one hand, and encouraging “bottom-up” innovation that follows faculty interest.

There is an important potential role for CDSS to better connect and promote research at the intersection of computing, data science, and social sciences, humanities, and ethics. The key here is not to just add a “coordination role” or “more meetings,” but to show that CDSS can add value by bringing teams together, showing results, increasing funding, and having real impact on Berkeley and the world. This could evolve into facilitating collaborative research. The Berkeley Institute for Data Science seems like a natural home for such efforts given its past and current work as well as its new leadership, though an infusion of resources would likely be necessary.

As we canvassed our colleagues regarding the intersection of data science, social sciences, and the humanities, we heard some concern across campus that CDSS will "scoop up" existing research, or fundraise around it, without really contributing back to the Berkeley community. We surface this issue as a potential barrier to the increased collaboration that will be necessary for realizing a shared vision around computing, data science, and society. We take no position on the likelihood or validity of this concern, but note that since it is clearly present, CDSS will need to communicate both its many past efforts and its future commitment to social and ethical research, to supporting interdisciplinary and inter-departmental research, and to fostering the development of new areas of research. At the same time, we strongly urge our colleagues outside of CDSS to recognize the many contributions that CDSS faculty and leadership have already made toward interdisciplinary and collaborative endeavors, to give our colleagues the "benefit of the doubt" when it comes to intentions, to be cognizant of the challenges inherent in standing up new programs of multidisciplinary research, and to be ready and willing to jump into these new endeavors and communicate the intellectual value of social science and humanities perspectives, methods, and domain knowledge.

Interestingly, other universities have mainly developed organizational structures that involve cross-disciplinary centers rather than a department to support this research. This shows that establishing departments may be challenging, yet it could put Berkeley at the forefront of advancing research initiatives and educational frameworks in data science and society, for example, by facilitating the hiring of scholars who would not be appealing to any existing department. This would particularly focus on data generated by human activities, such as computational social science (such as policy and government), the intersection of economics and computer science, sustainable and inclusive urban infrastructure, and a pervasive theme across these domains such as issues of privacy, security, and equity. In the case of CDSS, it may make sense to start with the proposed Human Technology Futures (HTF) and the (still to be named) data science applications departments as the central coordinating organizations. If they can establish areas of expertise, show impact, and build partnerships based on trust, they would be well positioned to expand and deepen their role over time.

10. Next Steps: Looking Ahead to 2024-25

The committee looks forward to discussing this report with campus stakeholders, including CDSS leadership, DIVCO and other Academic Senate committees, the Council of Deans, and senior campus leadership. The committee is open to refining or revising recommendations based on those conversations. Looking ahead to next year, the committee plans to further refine and elaborate its recommendations related to research, including continuing to explore what peer and near-peer institutions are doing at the boundary between data science and the social sciences and humanities.[3] The committee also plans to address two additional areas of importance: (a) graduate education, and (b) public engagement and community engagement.

[3] Two institutions that are more focused on data science and society are UVA and UNC. We also plan to add information about ongoing efforts at UCSD as they develop.

For example, the committee will consider the ways that CDSS and other campus resources can be leveraged to ensure that more graduate students leave Berkeley with data science skills that assist them in making an impact in the public and private sectors. And the committee will consider how campus initiatives related to data science can leverage the substantial campus expertise in community engaged scholarship.