

Sociotechnical integration in data science education

Prof. Cathryn Carson, University of California, Berkeley

Cathryn Carson is an STS scholar and a historian of science and technology who has been active in interdisciplinary collaborations in undergraduate and graduate education, including nuclear engineering and data science.

Ari Edmundson is an STS scholar and intellectual historian who has collaboratively developed integrated course materials and dedicated courses to embed critical thinking about human contexts and ethics in data science curricula.

Ramesh Sridharan is a computer scientist, educator, and machine learning engineer who brings real-world experience with machine learning and AI solutions and a commitment to positive social change.

Alex Strang is an applied mathematician with a passion for teaching, mentoring, and leadership, with research foci on uncertainty quantification in hierarchical Bayesian interest and processes on large networks.

Lisa Yan is a computer science and data science educator and researcher with expertise in data analysis of student learning in large courses, computing ethics education, and teaching assistant education.

Lisa Yan, University of California, Berkeley

Ari Edmundson, University of California, Berkeley

Alexander Strang, University of California, Berkeley

Sociotechnical Integration in Data Science Education

Abstract

Colleges and universities are witnessing the emergence of the new interdisciplinary field of data science/data analytics, which typically includes computing, statistics, mathematics, data systems engineering, and real-world applications. As a new interdisciplinary engineering discipline, the field offers a transformational opportunity to develop a meaningful, collaborative, integrated sociotechnical curriculum from the ground up to bridge the division between “society” and “technology.” Data technologies and data-driven decision-making are a powerful entry point into engineering education for faculty in science and technology studies (STS), engineering studies, and the humanities and social sciences more broadly. All data is constructed through human and social processes, and data analytics and algorithmic systems remix this data to drive future decision-making. Engineering-minded students initially may find it intuitive to draw a box around “the technical system” (say, an algorithm for risk assessment) and separate it from “everything else,” defining their responsibility as optimizing “the technical system.” However, STS provides a toolkit for systematically blurring the boundaries of the box around the “technical system,” showing how considering historical, social, and political settings is essential for competent technical practice.

This paper analyzes UC Berkeley’s experience with ongoing curricular sociotechnical integration and its generative tensions for a team of instructors including STS, Electrical Engineering and Computer Sciences (EECS), and Statistics. The paper presents two case studies of sociotechnical integration in undergraduate technical courses that include practical exercises of applying principles and techniques to real-world situations. In a junior-level data science fundamentals course, a multi-week course unit teaches students linear modeling and feature engineering by way of a real-world case study about the politics of predicting housing prices for property tax assessment. In a senior-level inference and decision-making course, students are asked to complete an integrative course project in which they apply models to draw inferential conclusions about real-world data. An interdisciplinary team of instructors has enriched the course’s existing case studies with STS frameworks to provide students the necessary scaffolding to engage in substantive critical work on final projects.

This paper reflects on the broader goal of building a sociotechnically integrated undergraduate data science curriculum including a dedicated STS class on “human contexts and ethics” and a pedagogical training class. Through these case studies and reflections, the paper shares institutional and interdisciplinary lessons learned from co-designing multiple courses with instructors across disciplines.

Introduction

Sociotechnical integration is a compelling framework for addressing the demands of educating future engineers and similarly situated practitioners [1]. One of its most exciting opportunities comes when a new field is emerging and new educational programs can be designed from the ground up. This paper analyzes UC Berkeley's experience with sociotechnical integration in the case of a new interdisciplinary undergraduate engineering discipline: data science. It reports on experiments in collaborative curricular design and implementation by an interdisciplinary team of instructors that currently achieves impact at the scale of several thousand students each year.

Sociotechnical integration draws from the field of STS (science and technology studies) and related disciplines in the humanities and social sciences. Over the last decades, these fields have developed theoretical frameworks and case studies to illuminate how social and technical aspects of situations in the world cannot be cleanly separated from each other, but instead merge together in a sociotechnical whole [2,3].

As an educational strategy, sociotechnical integration seeks to broaden students' understanding of the human and societal aspects of phenomena that may be typically classified as technical, and to develop students' capacity to recognize, disrupt, and ultimately work across conventional divides that separate technical from non-technical disciplines and practices. Within engineering education, and STEM education more broadly, this approach supports students in becoming competent practitioners without "boxing in" their work as purely technical practice. Through sustained sociotechnical integration, students can gain critical skills, concepts, vocabulary, and experience with social, political, and historical aspects of real-world practice. As compelling as sociotechnical integration can be to students, instructors, and STS scholars, it can be challenging to implement, requiring continuing collaboration of instructors from very different disciplinary backgrounds and a readiness to tackle institutional barriers. Different forms of implementation are possible in response to particular institutional contexts and constraints [4,5,6].

Data science/data analytics has been emerging in a wide range of educational institutions, including colleges and schools of engineering and computing as well as liberal arts and sciences, for approximately a decade. It commonly bridges across computing, statistics, mathematics, data systems engineering, and real-world applications. While building on precursors especially in computing education (including "responsible computing") [7,8] and statistics education (including real-world considerations and case studies) [9,10], the participation of instructors from liberal arts disciplines, specifically STS, creates a transformational setting to develop a meaningful, collaborative, integrated sociotechnical curriculum from the ground up.

At our institution (UC Berkeley), a Data Science undergraduate degree program has been under development since 2014 and available to students as a Bachelor of Arts (BA) degree since 2018 and an undergraduate minor since 2019. The institution enrolls over 30,000 undergraduate students, and many courses in data science and related areas exceed 1,000 students, with large teams of teaching assistants (TAs) supporting instruction. Currently about 800 students each year graduate with a Data Science BA, and several thousand students each year take a junior-level or above data science class.

The data science discipline has been defined, at our institution, to include the human contexts and ethics (HCE) of data technologies and data-driven decision systems. HCE draws upon STS as well as related disciplines such as history of science and technology and critical theory. Data-centric processes, approaches, and technologies are a powerful entry point for sociotechnical integration: all data is constructed through human and social processes, and data analytics and algorithmic systems remix this data to drive future decision-making.

Within the Data Science undergraduate program at our institution, HCE is both a degree requirement (a dedicated required course) and an element to be embedded within other courses through the lens of sociotechnical integration. This emphasis on sociotechnical integration is not driven by accreditation criteria. Rather, it responds to the distinctive character of the field of data science; to our institution's stated intention to prepare broadly educated, publicly minded future leaders; and to the positioning of our instructional team, in which STS has been involved from the start. Our approach to sociotechnical integration has drawn heavily from classical traditions in STS, including the sociology of science and technology, controversies and case studies, the history of science and technology, and feminist STS [11,12], as brought to bear on core questions in data science.

Sociotechnical integration in undergraduate data science education at our institution is done by an interdisciplinary team of instructors with backgrounds in STS, Electrical Engineering and Computer Science (EECS), and Statistics. Ideally, the integrated sociotechnical framing can be reinforced across the curriculum, in technical courses as well as (and in connection with) in the dedicated Human Contexts and Ethics course, in TA pedagogy/training, and where possible in co-curricular settings as well. This work has been growing in phases since the initial courses in the program were created, and it takes different forms in different courses, depending on the different opportunities presented at different stages of a student's educational pathway. In this paper we focus on two courses, a junior-level data science fundamentals course (Data 100) and a senior-level inference and decision-making course (Data 102).

Data 100

Data 100 is a junior-level core data science course that is offered every semester. Students use Python, linear algebra, and statistics to learn fundamental data science techniques at the intersection of computer science and statistics: exploratory data analysis, visualization, data cleaning, models for regression, classification, clustering, and bootstrapped inference techniques. In the Data Science BA degree, students often take this course in conjunction with advanced mathematics requirements like linear algebra and probability. This course serves as a prerequisite for more advanced Data Science courses, like Data 102. Because the course also is a fundamental requirement in several undergraduate and graduate programs, from engineering and sciences to public policy, it is the largest junior-level course at UC Berkeley and regularly enrolls 1000 students; Fall 2024 had 963 students (900 undergraduate, 63 graduate).

Each semester, Data 100 is taught and managed by a large course staff consisting of two faculty co-instructors (chosen from a set of 6-7 professors), a dozen part-time undergraduate and graduate Teaching Assistants (TAs), and around twenty undergraduate readers and tutors. Each week, co-instructors teach two 80-minute lectures, and TAs run a one-hour discussion section of 35-40 students. Formative assignments include programming labs and homework assignments

that incorporate programming, written exploratory analysis, and mathematical proofs. Given the class size, curricular development is limited; there is an emphasis on designing learning experiences that can be efficiently autograded. Homework assignments are therefore broken up into many subparts.

Beyond summative midterm and final exams, students also complete two course projects—longer, multi-week individual explorations. These course projects are similar to in-depth homework assignments because of their rigid structure and build student understanding of the data science modeling tasks in the course: regression and classification.

Curriculum Process and Timeline

Given the larger investment that students have in Data 100's multi-week projects, we first focused our curriculum development on these assignments and branched out to other course components like lecture and discussion section. While the initial course development process resembled the traditional "import" of social contexts in technical coursework (e.g., as exemplified by case studies in engineering ethics courses [7]), our final product reflected the iterative, collaborative, continuing conversation around sociotechnical integration, which was conducted across multiple semesters and Data 100 instructors.

We have worked together with multiple Data 100 instructional teams to implement a two-part project centered on the politics of property appraisal in Cook County, Illinois. The project consists of two Jupyter notebooks, an 80-minute lecture, and a lesson plan for a 20-minute small-group discussion. Across each of its components, the project teaches students to approach every data science problem as sociotechnical, incorporating perspectives drawn from the social sciences and humanities into their decision-making process at each stage in the data science lifecycle.

Collaboration between a Human Contexts and Ethics (HCE) team of instructors and students and Data 100 course staff has run for approximately five years. It began when students familiar with the course reviewed every homework and project to search for opportunities to develop students' sociotechnical thinking skills. An existing homework assignment used the Ames, Iowa residential property datasets, a classic dataset in statistics education, to teach students how to use linear regression to predict housing prices based on sales data. Housing is a rich topic for discussing human contexts and politics, so this homework appeared to be ripe for collaboration. The dataset lacked an interesting real world context to explore, so we searched for alternatives.

In this context, an undergraduate student drew attention to a newly published dataset on GitHub that would become the cornerstone of a two-week long project. The Cook County Assessor's Office (CCAO) had recently begun a multiyear effort under the newly-elected county assessor to develop an Open Data initiative as part of an effort to curb a corrupt and regressive system. It was being publicly presented as a means of leveraging the power of machine learning and the transparency of open data to create a more just and equitable system and restore public trust in the Assessor's Office [13,14,15,16].

The HCE team developed the project by researching property tax assessment practices generally, metrics of “fairness” in housing appraisal, and the specific case study of the CCAO. The first version of the homework was introduced together with a four part pre-recorded lecture introducing students to the case study, offering some historical context on the racial politics of housing and property assessment; breaking down how the CCAO approached the problem in terms of the data science lifecycle; and modeling how to understand the stakes of the case study from a sociotechnical perspective. The initial version project did not yet fully integrate the “social” and the “technical” components, leaving students with the feeling that they were two distinct “steps” in the process. To remedy this, the HCE team, the Data 100 instructor, and a faculty member in Sociology subsequently applied for and were rewarded with a small grant to further develop the project, later joined by additional faculty members. These included extensive modifications to the homework, expanding it into a two part project (now Project A1 and A2); and major revisions to the lecture and discussion lesson plan.

Learning Objectives

The CCAO project teaches students how to construct a model to address a real world problem in such a way that success depends on the ability to understand the human and social context of the problem and the data. First, the lecture, notebook, and discussion questions all challenge students to consider the meaning of different kinds of errors and their distribution. While students first evaluate their model using root mean square error (RMSE), they are then confronted with RMSE’s inability to capture regressivity, which their linear models tend to produce. The purpose of this exercise is not, however, to show students the technically correct way to create a more accurate model, but rather to suggest that the choice of model and the loss function used to evaluate it have politics. There is no single “correct” solution. Rather, the model redistributes the tax burden in ways that map onto different conceptions of what counts as “fair.” It asks students to consider how someone’s positionality impacts the meaning and experience of different forms of error (e.g. a 1% overassessment negatively impacts a low-income homeowner more than a high-income homeowner) and their assessment of what counts as “fair.”

The project thus also teaches students not to take for granted that a mathematical measure of accuracy is the same thing as “fairness,” and that rather, making the sociotechnical translation of accuracy into fairness is part of the work being done by the CCAO in attempting to bolster its own legitimacy. Far from self-evident, accuracy is made to stand in for fairness—and the CCAO’s project depends on whether or not it can make this translation hold. This also helps students recognize the limits of data science in solving social problems: whether the public accepts outputs of a model and thereby trusts the CCAO is a matter of communication, rhetoric and politics, not just technical sophistication. In short, students learn to treat algorithmic fairness as a sociotechnical construct.

Because the project requires students to consider the wider social context that created and framed the problem for them, and allows students to make their own decision about how to proceed, this project also navigates some of the well-known shortcomings of using case studies to teach engineering ethics, namely, that case studies excessively emphasize exceptional breakdowns rather than the texture of everyday practice [21], or that they so narrowly define the stakes and takeaways of a situation in advance that the exercise feels like a didactic game [22].

Sociotechnical Integration Curriculum Changes

Project A1 and A2

Project A1 and A2 offer students the opportunity to build their own linear model to predict the values of residential properties using data drawn from the Cook County Assessor's Office GitLab. This dataset contains property characteristics data on all residential properties in Cook County (square footage, number of bathrooms, etc.) and sales data on homes that were recently sold. Project A1 emphasizes Exploratory Data Analysis, and familiarizes students with the dataset, its structure, and what it represents. This lays the groundwork for Project A2, in which students specify and fit a linear model to several features of the property characteristics data to predict property values at scale. Students then analyze the error of their model, first using RMSE, and then by exploring the limitations of RMSE. In this way, students reflect on questions of fairness and explain the meaning and impact of different kinds of errors, in part by drawing on the analysis presented in the guest lecture.

Discussion Lesson Plan

The discussion lesson plan emphasized helping students think through how different kinds of distributions of errors in the prediction of housing prices relate to their intuitions about fairness. The first discussion question presents students with a series of alternatives (e.g. “A homeowner whose home is assessed at a lower price than it would sell for,” “An assessment process that systematically overvalues inexpensive properties and under-values expensive properties.”), and asks them to explain which strike them as most unfair and why. This question helps students think through their own intuitions about fairness and check their understanding of how property assessment relates to taxation. The second discussion question shows students two different plots representing two different ways to visualize how errors might be distributed in a linear model, and asks students how they would align with the different scenarios discussed in the first part of the discussion.

Course Staff Meetings

Implementing the CCAO project required extensive work with the undergraduate TAs who teach the discussion sections and help students with the project during office hours. Many of the undergraduates TAs for Data 100 have not yet fulfilled the HCE course requirement for the Data Science major, are unfamiliar with the stakes of property taxation, and—most importantly—have little experience leading discussions on sociotechnical topics. Therefore, course staff meetings proved an invaluable opportunity to help prepare TAs to teach the material, explain its significance, and field questions. It also provided a chance to get feedback from TAs who had taught the material in prior semesters, who told what was most compelling and confusing to both TAs and students. This feedback helped us iterate and improve upon all parts of the project across several semesters.

Discussion/Reflection/Evaluation of Data 100

Across eight semesters of implementation of the CCAO project in Data 100 we have witnessed significant improvement in student learning and performance as a result of continuous iteration and updating of the project components. In particular, the current version of the project seamlessly integrates the social and technical components of the homework so that “technical” success on the project requires students to invest considerable effort in understanding the social context of their model. Naturally, the quality of student answers to the more qualitative questions varies considerably, but connecting these questions more closely to students’ modeling decisions, instead of asking them at the end of project sections, has tended to promote more widespread and sincere engagement.

Our emphasis on real-world and sociotechnical contexts also makes the project especially engaging to students in practice. In evaluations, students frequently list the project as one of their favorite parts of the course because the concrete context made students feel that they were actually building something meaningful and related to their daily lives. Students appreciate that the project emphasizes ethical themes and that it is relatively open-ended. Although many students comment on the project’s difficulty, they also emphasize that it was especially rewarding.

The experience of working on this project also helped us devise better strategies for fostering quality interdisciplinary collaboration in sociotechnical curriculum development. The first stages of project development were largely modular, with new curricular materials developed internally by the human contexts and ethics and then handed off to the instructors of Data 100 for implementation or feedback. By contrast, later stages of the project involved more direct high level discussion about learning outcomes and hands-on collaboration when writing questions and homeworks. Our more recent collaborative work developing sociotechnical curricula for Data 102 build upon these experiences, so that lengthy discussions about learning objectives and collective brainstorming about desirable outcomes involving faculty from HCE, Statistics, and Computer Science preceded curriculum development.

Data 102

Data 102, Data, Inference, and Decisions, is a senior-level undergraduate, topics course in Data Science. that is also offered every semester. It fulfills the modeling, decision making, and learning requirement of the Data Science BA degree, and it is often one of the final technical requirement courses students take before graduation. It regularly enrolls approximately 250 students, with 245 students enrolled in the Fall 2024 semester. Students often take Data 102 concurrently with other required advanced courses, such as the HCE requirement. This overlap in student population was a boon to cross-course collaboration between Data 102 and HCE instructors, which we expand on below.

Data 102 aims to help students master the modeling, inferential, and decision-theoretic aspects of data science by returning to, and expanding on, topics students learned in their prerequisites. Data 102 expands and links ideas introduced in prior classes so that students leave with robust, networked knowledge in their domain. For example, Data 102 expands upon students’ previous exposure to logistic regression in Data 100 by introducing the generalized linear model. Major topics include binary decision making via hypothesis testing and classification, Bayesian inference with conjugate and hierarchical models, prediction using generalized linear models and

nonparametric models, causal inference, bandit problems, and reinforcement learning. The course concludes with a series of case-study lectures.

Data 102 follows a similar structure to Data 100 and includes two 80-minute lectures, discussion and programming lab sections, and homework. In contrast to Data 100, homework assignments in Data 102 are the most challenging component of the course and are assigned every two weeks. Each homework involves a combination of mathematical and theoretical exercises, computational and programming exercises, and interpretation exercises that ask students to critically evaluate results and situate them in a broader context. There are two summative midterm exams.

Given the advanced technical material in Data 102, the curriculum naturally lends itself to problem-based learning strategies with a deeper exploration of case studies and real-world datasets used on assignments. We discuss how this technical context provides an opportunity for more integrated, sociotechnical education through the course's primary student deliverable: their open-ended, group final project.

Curriculum Process and Timeline

Over the course of a summer and fall, a team of four instructors and one undergraduate assistant worked to incorporate and reinforce HCE learning in Data 102. The interdisciplinary team included faculty in computer science, statistics, and STS. Two of the four members were active Data 102 instructors. The remaining two instructors provided expertise in STS/HCE, in computer science and data science pedagogy, and specifically in teaching HCE-aligned courses within a data science degree program.

The team spent a summer exploring, developing, and aligning goals. Over the summer the two Data 102 instructors, who were the fall co-instructors, provided course context, while reading materials suggested by the other team members [17,18]. Both Data 102 instructors reported that the reading and regular discussion process provided essential education needed to target and prepare course revisions. At the same time, the undergraduate assistant reviewed the existing Data 102 syllabus. After reading through all of the lectures, homeworks, and labs, he prepared a list of potential case studies for expansion. By the end of the summer the team had agreed on three key areas of course revisions to trial during the fall: developing a new HCE-focused lecture, revising assignments to incorporate additional HCE focus, and challenging students to think about the ethical implications of their work in their final projects.

The team spent the fall implementing and testing the revisions proposed during the summer. During this time the Data 102 instructors prepared course revisions, and brought design questions regarding those revisions back to the team for discussion. These included assignment revisions, changes to the final project specification, and the construction of two new HCE focused capstone lectures for the final case-study unit. The team refined learning goals, suggested alternate sources and reading, reviewed and provided constructive criticism on proposed changes, helped prepare rubrics, and gave practical guidance on teaching HCE topics in a data science course.

The team met biweekly for six months, beginning in June 2024 and ending in December 2024. During that time they maintained regular contact. Both the frequency and duration of the collaboration was essential to its success. Sustaining regular meetings created a collaborative environment that fostered trust, aligned goals, and maintained forward momentum. The meetings also provided many opportunities for free-flowing discussion, brainstorming, and HCE education for the Data 102 instructors. The HCE education was critical for sketching out ideas and defining connections between technical course content and the HCE ideas. The team transitioned from ideating during the summer, led and directed by the HCE experts, to teaching the material during the fall, led by the Data 102 instructors as they developed more HCE expertise and were able to request specific areas of support leading up to each unit's content changes.

Learning Objectives

As part of our team curriculum development, we tackled final project redesign with a backwards design approach by identifying new key sociotechnical learning goals to broaden the project's original objectives. In particular, we aimed to teach students conceptual frameworks and basic skills for performing responsible data science research and analysis. The capstone lectures on algorithmic fairness served as models for how we expected students to approach their own projects. These lectures demonstrated that a sociotechnical approach to research and design would require students to know *how* to look for and identify the most relevant human contexts of their research project, for example, paying attention to the identity and positionalities of the persons represented or otherwise implicated in the data in question.

Responsible sociotechnical practice involves more than just the application of interdisciplinary conceptual frameworks, but also specific knowledge about contexts that only comes through careful empirical research. A second major learning objective of the new interventions was thus to foster students' basic research skills by requiring students to conduct independent research on the most salient contexts and ethical stakes of their project, report their findings, and integrate it into their analysis of the data. So in addition to getting practice doing research, students also learn how to *use* qualitative research to interpret and draw conclusions about their data, cultivating synthetic and interpretative skills.

Sociotechnical Integration Curriculum Changes

Final Project Assignment

Traditionally, students complete the final project in teams of 4 over the last six weeks of the course. In their projects, they choose a dataset from a list of curated options or select one of their own, complete a data report, perform exploratory data analysis, then conduct two analyses, each involving either causal inference, Bayesian modeling, hypothesis testing, or predictive modeling. Finally, students are asked to reflect on the limitations of their analysis, and to provide a recommendation based on their conclusions. Team progress is evaluated through a series of checkpoints, and member participation is evaluated through a series of member evaluation polls. The project is meant to give students hands-on experience with a real dataset, and for many students, it is the most involved end-to-end data science experience that they complete during

their degree. This includes the process of encountering roadblocks, dead ends, and data or methodological limitations and finding a way around them in a limited timeframe.

Literature Review

In the past, the final project specifications did not ask students to read, provide, or cite any formal sources. As a result, very few projects cited prior academic work on their problems, even when relevant work was evidently available. Since referencing and responding to prior work is an essential component of research, and is an underemphasized skill in many data science courses, we introduced a Prior Work Section to the project specifications (Figure 1). Students were also asked to submit a longer list of prospective sources during their first project check in. The TAs reviewed the submitted sources to give project groups feedback if they had not selected sufficiently reliable sources. We asked students to submit this list first so that they would start reading before completing their project.

Critical Self-evaluation

We also expanded and reorganized the conclusion section of the project specification. We modified the conclusion to give students a clear structure for constructive discussion beyond result summary. In particular, we focused on two important categories of reflection: critical evaluation, and recommendations, both shown in Figure 1. The first reading in HW 6 was chosen to motivate students to complete their critical evaluation; the second motivated the recommendations portion. The critical evaluation section asked students to critically discuss their methods and results. We designed this section to promote healthy skepticism and honesty in reporting. The recommendations section asked students to make a policy recommendation based on their findings, then to reflect on the ethical considerations implicit in that recommendation. Since each question posed is prospective, student submissions were graded based on sincerity of demonstrated effort.

Prior Work Section: • Provide citations for 1-2 relevant sources. Ideally, these should address similar questions to yours. You must use high-quality sources: academic papers, accompanying blog posts written by the paper authors, or detailed reports by reputable institutions are ideal. If you aren't sure about the quality of a source, feel free to check with course staff at office hours. • For each source, explain how a question they address is relevant to your research question(s), and in what ways their approach is similar or different.
Conclusions Section: • <i>Outcomes summary</i>: Summarize your key findings. • <i>Critical Evaluation</i>: 1. What limitations are there in the data that you could not account for in your analysis? 2. What domain knowledge were you missing that would have informed your data treatment, question choice, or method choice? Identify one question you would ask a domain expert to better inform your effort. Explain how the answer could have impacted your choices. 3. How robust do you think your conclusions are to your modeling choices? Identify at least one modeling choice that you could plausibly change, and explain how your choice may or may not have biased your result. Explain how a user should reason with this bias. 4. How generalizable are your results? How broad or narrow are your findings? • <i>Recommendations</i>: 1. What future studies could build on your work? Propose a follow-up study that could confirm or refute your claimed results. 2. Based on your results, suggest a call to action. What interventions, policies, real-world decisions, or action should be taken in light of your findings? 3. Discuss potential impacts of your call to action. Your discussion should holistically address the most crucial human contexts and ethical stakes of your work. To help you discuss these impacts thoughtfully, we've provided several prompts below. While we expect your response to address at least some of these, you should treat these as guidelines rather than a checklist. a. Is your recommendation feasible? Who would have the power to implement it? Who might have the right to object? b. How would the impacts of your recommendations vary between individuals and groups? Who stands to benefit? Explain how you would respond to an individual who might reasonably oppose your recommendation. c. What are the implicit values or ethics that guide your call to action? Why are these important to you in this setting? How are these choices related to your social identities, educations, or cultural experiences? If you had any disagreements in your group around the answers to this question, please discuss.

Figure 1. Sociotechnical Components of Data 102 Final Project Specification.

Lecture Revisions

The team created a new “Case Studies and Context” unit for the end of course. The new unit included an existing case studies lecture, and two new lectures on HCE. The unit was designed to help students complete their final project, in particular, to critically reflect on the analyses they performed and the recommendations they provided. It was also designed to step back from the technical framing used throughout the course to discuss important issues “outside the frame” that require both technical and contextual consideration to handle effectively. The two new lectures replaced a mini-unit on differential privacy, and were supported by a new final homework

assignment, revisions to a causal inference lab, and revisions to the first homework assignment.”

The two new lectures covered algorithmic fairness at a high level, then focused on the Allegheny Family Screening Tool (AFST) algorithm as a detailed case-study [19, plus a significant body of other studies that we do not cite here]. The algorithmic fairness lecture used algorithmic fairness as a problem context to broaden students’ appreciation of the ethical impacts of technical choices made in high-stakes algorithms that administer social policies or recommend choices regarding individuals. We began by following the standard course narrative: pose a broad decision or inference problem, then operationalize it by fixing definitions and objectives. In most units, we would then focus the majority of our effort as a class on algorithms that solve the problem implied by the chosen definitions and objectives, and, in some units, on comparing and evaluating answers produced by separate problem frames. Typically, we evaluated methods based on accuracy or generalization standards.

The algorithmic fairness lecture proceeded in this vein, but aimed to intentionally break the traditional course narrative. Instead of spending the majority of the course on the *how* questions implied by a particular frame, we adopted a frame, then worked outwards to critically evaluate it. This exercise aimed to shift the students focus onto the importance of model and problem definitions, and their ethical import when the problem framed involves people. It also aimed to de-normalize standard framing choices that, if left undiscussed, may be left unconsidered, even if they merit consideration. In particular, we argued that technical choices made when framing the problem, are, in context, not purely administrative, but may encode political values or, in practice, define policy.

The lecture began by posing a mission statement, “We want our tool to use data to perform its assigned task while treating protected classes fairly.” After translating the mission statement into a technical problem by fixing a task, an accuracy metric, a data set, a model, a set of classes, a fairness measure, and a method for compromising objectives, we worked outward systematically to question each choice needed to operationalize the mission statement. For each choice, we demonstrated at least one plausible alternative, discussed the significance of the choice made and its implicit motivations, and how that choice encodes an ethical or positional stance in a human context. For example, we discussed the difference between individual and group fairness measures, and how the choice to adopt a group fairness measure is better justified from the position of the tool than from the perspective of the individual. For each example choice we briefly highlighted case studies that emphasized the real-world impact of the technical choices made when framing a problem.

We ordered the sequence of choices critiqued from the most “internal” or “technical” to the most contextually impactful or controversial. In this manner, the lecture provided a broad overview of examples and debates in algorithmic fairness, while providing a pointed critique of a technical philosophy that focuses on problems within the frame without debating the frame. We concluded the lecture by raising critical questions “outside” the frame. These focused on aspects of tool validity, including robustness, non-maleficence, beneficence, transparency, explainability, and accountability. In particular, we challenged students to think about what an algorithm *does* in the world once applied as its social function, and argued that the social function depends on framing choices that, thus far in the course, had been made based on other considerations.

The components of the lectures about the AFST, an algorithm meant to predict whether children are at risk of neglect to guide agency intervention, followed the typical structure of a case study, but with a focus on both technical and ethical considerations. For example, in addition to a question about the real-world impact of a false positive and false negative in a binary decision, which closely mirrors prior lecture discussion questions, students were also asked to evaluate the disparate impact of those decisions across socioeconomic and racial groups.

Existing Assignment Revisions

To support the algorithmic fairness lecture, we added a fairness question to the first homework assignment, which focuses on binary decision-making. The problem asked students to prove that three standard fairness objectives (matched false positive and negative rates between classes and calibration of predictions) cannot be simultaneously satisfied. The problem required students to apply prerequisite probability skills, extract a mathematical frame from a generic problem, and to summarize the components of a binary decision making model at a high level. Finally, we asked students to reason with their conclusion with an open-ended question: “What does it mean if three objectives cannot be satisfied simultaneously, in general cases? Is fairness hopeless? Can all three be satisfied within a reasonable approximation? If not, then what values would guide you to prioritize some objectives over others?” Similar to the CCAO project in Data 100, this helped students recognize the fundamentally contested and political character of considerations of fairness in relation to modeling decisions.

Reading Assignment

We adapted the final homework assignment to pair with the new unit on case studies and context. The new homework assignment asked students to read an academic paper on robustness, “Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty” [20], then a chapter from Virginia Eubanks’s book, *Automating Inequality*, on the use of algorithmic risk prediction in Child Protective Services in Allegheny County, Pennsylvania [19]. Students had two weeks to complete the assignment.

The first reading was chosen to highlight the impact of method on conclusions. In a large survey experiment, the authors showed that, separate teams of analysts given the same data and question, produced a wide range of answers while using the technical methods developed in the class, and despite a number of methodological controls (e.g., preregistration). Importantly, much of the spread in answers could not be explained by documented differences in methodology. This paper was chosen to motivate students to carefully inspect and document their framing choices. It served to both support the emphasis on framing in the case studies unit, and the critical evaluation component of student final projects.

The second reading was chosen to provide human context on the final case study offered in the case study unit. In particular, it was chosen for its narrative power, focusing on individual perspectives that include designers, users (call workers), and families impacted by the tool. It was also chosen to highlight the questions: “Is it fair to make decisions about individuals based on their predicted behavior?” and, more pointedly, “Is it fair to make decisions about individuals

based on their predicted behavior, when prediction is rooted in the statistical behavior of groups they are assumed to be members of?”

After completing the reading, students were asked to complete a discussion activity on an online discussion board. We broke the class, at random, into ten groups of roughly 25 students each. Students were asked to post at least two comments, questions, or reflect on two quotes from the reading, then post at least two replies to other student posts. All posts had to exceed a minimum length corresponding, roughly, to two thoughtful sentences. Students were asked to complete at least two posts before the first context lecture, then, to submit the full four by the end of the unit. Grading was performed automatically using an autograding tool [reference redacted for review]. A TA revised the tool for our use, then designed a grading notebook to handle exceptions and partial credit. In general, class participation was high. Three quarters of students completed the full assignment, while 90 percent completed the full assignment with minor formatting errors (at a 2 out of 3 points standard). In discussion and in course reviews, individual students gave positive feedback on the reading. The instructors did not complete or pose an experimental evaluation of the assignment.

Discussion/Reflection/Evaluation of Data 102

The results varied across the different modalities of integration. While student response to the lecture was very positive from students who attended, attendance was low compared to other lectures at the end of the semester. Many students appreciated the content, as shown by this representative remark from our course evaluations: *“Enjoyed the non-technical lectures as well, HCE concepts were integrated very well, which I think is important.”* This attitude was reflected in the responses to the reading assignment, which were generally sincere and thoughtful. Students engaged critically with the material, with several expressing surprise at the findings on researcher degrees of freedom [20].

Student responses to the project changes were unfortunately less well-formed. All groups of students found papers that could be described as relevant prior work, but most of them were only tangentially related to their problem, with many using approaches and ideas orthogonal to the methods that students had learned. Student responses to the technical critical evaluation were mostly thoughtful. This reflected a small improvement over their responses to similar questions asked in previous semesters.

Most students neglected to engage with the ethical implications of their work, omitting that part of their conclusions. Others gave rote and formulaic responses that reflected a limited level of engagement. Only a handful of groups critically engaged with the guiding questions shown at the end of Figure 1. Many groups gave calls to action of the form “gather additional data” or “conduct additional research,” rather than thinking through recommendations with tangible or immediate impacts. We concluded that this was likely due to a combination of (a) an overly steep learning curve (i.e., insufficient practice with simpler versions of similar questions); (b) insufficiently precise instructions, which left open the possibility of rote answers; and (c) students putting off these questions until the technical work was complete, and not leaving sufficient time to engage with them.

We plan to address this in the coming semester by adding more guided versions of these questions to each homework assignment, giving students examples of calls to action that follow from various ethical principles, and giving them more guided practice with critical thinking around the impacts resulting from, as well as the values underlying each recommendation.

Conclusion

Sociotechnical systems are consistently and visibly at work across data science and data analytics. At UC Berkeley, this circumstance has become a pedagogically inviting opportunity to blur the boundaries of the box that is conventionally placed around a “technical system” (for instance, an algorithm that otherwise would be viewed as separate from “everything else”). In Data 100 and Data 102, we seek to give students different kinds of practice in crossing between domains they may initially view as “technical core” and “social context.” This experience gives them opportunities to recognize how these domains are entangled and in fact mutually constitutive. Moreover, course exercises give students scaffolded opportunities to learn to work across the domains that are no longer so clearly separated. We see this as an example of sociotechnical integration from the ground up.

This work of integration is challenging across multiple dimensions. With all the appreciation that enthusiastic students may express, it is manifestly challenging for many students in our courses. We have observed this particularly in Data 102, where sociotechnical integration runs against habits of thought and practice that have been ingrained in their previous studies. We expect a similar learning curve in Data 102 to what it has taken to advance integration in Data 100. It is also challenging for instructors. The literature on sociotechnical integration in engineering education and related topics has stressed the need for instructors who engage in interdisciplinary integrative work to have supportive environments that allow intellectual and pedagogical exploration, and to find interlocutors in other disciplines who are open to mutual learning.

Finally, it is challenging for an institution to foreground integrative approaches under the demands of ongoing teaching loads, including classes that must operate every semester at significant scale. However, the new interdisciplinary field of data science is a striking opportunity to bring STS and related approaches to bear within engineering education. The work presented in this paper was supported in various forms: the explicit inclusion of HCE in the core degree requirements of the Data Science undergraduate program, the external funding that supported the instructional resources for integrative coursework development, and the existing collaborative relations between the instructors. At a higher level, the process of sociotechnical integration that emerges in this work is inherently collaborative and community-oriented. By embodying these values through our curricular co-development, we have more opportunities to model effective sociotechnical integration to students in their learning. The educational outcomes that we are aiming at have the potential to redefine our students’ practice of data science.

Acknowledgments

The project is funded through the Mozilla Foundation Responsible Computer Science Challenge in partnership with Omidyar Network, Schmidt Futures, Craig Newmark Philanthropies, USAID, the Mellon Foundation, and the Rockefeller Brothers Fund.

REFERENCES

- [1] E. A. Reddy, M. S. Kleine, M. Parsons, and D. Nieusma. “Sociotechnical integration: What is it? Why do we need it? How do we do it?” Paper presented at 2023 ASEE Annual Conference & Exposition, Baltimore, Maryland. 10.18260/1-2--44239.
- [2] D. MacKenzie and J. Wajcman. Eds. *The social shaping of technology*. 2nd ed. Open University Press, 2002.
- [3] W. E. Bijker and T. Pinch. Eds. *The social construction of technological systems: New directions in the sociology and history of technology*. MIT Press, 1987.
- [4] A. Y. Patrick, M. H. Wisnioski, L. McNair, D.S. Ozkan, D. Reaping, T. L. Martin, L. Lester, S. Dunning, B. Knapp, L. B. Walker, and C. E. Haines. “In it for the long haul: The groundwork of interdisciplinary culture change in engineering education reform.” *Engineering Studies* 15:2, 144-167 (2023). 10.1080/19378629.2023.2243608
- [5] B. E. Seabrook, K. A. Neeley, K. Zacharias, B. R. Caron. “Teaching STS to engineers: A comparative study of embedded STS programs.” Paper presented at 2020 ASEE Virtual Annual Conference Content Access, Virtual On line. 10.18260/1-2--35281.
- [6] G. L. Downey and T. Zuiderent-Jerak. Eds. *Making and doing: Activating STS through knowledge expression and travel*. Cambridge, MA, USA: MIT Press, 2021.
- [7] N. Brown, B. Xie, E. Sarder, C. Fiesler, E. S. Wiese. “Teaching ethics in computing: A systematic literature review of ACM computer science education publications.” *ACM Trans. Comput. Educ.* 24, 1, Article 6 (January 2024).
- [8] Mozilla Foundation. Teaching Responsible Computing playbook.
<https://foundation.mozilla.org/en/responsible-computing-challenge-playbook/curriculum/>.
- [9] D. A. Nolan. Case studies in the mathematical statistics course. Lecture Notes-Monograph Series, 40, 165–176. 2003. <http://www.jstor.org/stable/435618>.
- [10] Academic Data Science Alliance. The data science ethos.
<https://ethos.academicdatascience.org/>.
- [11] T. M. Porter. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, 1995.
- [12] D. Haraway. “Situated knowledges: The science question in feminism and the privilege of partial perspective.” *Feminist studies* 14, No. 3 (1988), 575-599.
- [13] Cook County Government Open Data. “Cook County Assessor Model & Valuation Data Release.” <https://datacatalog.cookcountyil.gov/stories/s/p2kt-hk36>.

- [14] R. Ross. “The Impact of Property Tax Appeals on Vertical Equity in Cook County, IL.” University of Chicago working paper, May 2017.
<https://apps.chicagotribune.com/news/watchdog/cook-county-property-tax-divide/data/harris-study.pdf>.
- [15] J. Grotto, “An Unfair Burden: Cook County failed to value homes accurately for years. The result: a property tax system that harmed the poor and helped the rich.” *Chicago Tribune*, June 10, 2017.
<https://apps.chicagotribune.com/news/watchdog/cook-county-property-tax-divide/assessments.html>.
- [16] Cook County Assessor. “More transparent, predictable assessments are creating a better economic environment for all of Cook County. Here’s the proof.” *Medium*, November 6, 2019.
<https://medium.com/@AssessorCook/more-transparent-assessments-are-creating-a-better-economic-environment-for-cook-county-629ba45ba38f>.
- [17] A.L. Hoffman. “Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse.” *Information, Communication & Society* 22, 7 (2019), 900-915.
- [18] A. Hanna, E. Denton, A. Smart, J. Smith-Loud. “Towards a critical race methodology in algorithmic fairness.” *Conference on Fairness, Accountability, and Transparency (FAT* ’20)*, January 27-30, 2020, Barcelona, Spain. ACM, New York, NY, 10.1145/3351095.3372826
- [19] V. Eubanks. *Automating inequality. How high-tech tools profile, police, and punish the poor*. St. Martin’s Press, 2018. Chapter 4, “The Allegheny Algorithm.”
- [20] N. Breznau, E.M. Rinke, A. Wuttke, and T. Żółtak. “Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty.” *PNAS* Vol. 119 No. 44 (2022) e2203150119.
- [21] L. Winner. “Engineering Ethics and the Political Imagination.” *Broad and Narrow Interpretations of Philosophy of Technology*, ed. P. T. Durbin, 53–64. Dordrecht: Springer, 1990.
- [22] K. Palmås, “Engineering Judgment and Education: An Arendtian Account.” *Engineering Studies*, 16(3), (2024) 184–205. 10.1080/19378629.2024.2333239