Prediction in Practice Workshop Report
June 20-21, 2019

Powerful institutions increasingly base their decisions on AI-driven predictions of human behavior: whether a job applicant will be a successful employee, whether an offender will recidivate, how likely a recipient of social services is to get back on her feet, or how a student will perform in school. In these settings, the impact of AI depends on how its predictions are integrated into existing decision-making processes, policies, and institutions.

Research on how humans interact with technology suggests that there are several psychological, sociological, and political complications to the integration of AI into decision-making. In some settings, tools seem to receive more deference than their architects intend—but in others, humans completely ignore the predictions. Decision processes may change significantly when a system is introduced, but over time revert to previous ways of doing things; humans may struggle to integrate their own expertise with that of the system, and may learn how to manipulate inputs to reach desired outcomes.

We need to learn more about these high-stakes human encounters with AI. How can system architects best monitor and understand the frequently unanticipated ways that human decision-makers interpret and respond to a system's predictions? How do such predictions intersect with or challenge traditional forms of expertise? How and why are systems procured, used, resisted, and manipulated? Insight into patterns in the conditions of AI's integration into real-life contexts is key to ensuring the responsible use of these systems in making high-impact decisions. Understanding these processes should also help us to build tools that effectively capture both human and machine expertise.

In order to address these questions, we convened a workshop at Cornell Tech's campus on Roosevelt Island in New York City. The workshop was organized by Cornell's AI, Policy, and Practice Initiative (AIPP), Upturn, and Cornell Tech's Digital Life Initiative (DLI), with support from AI100, the MacArthur Foundation, and Cornell's Institute for the Social Sciences.

Over the course of two days, we explored a set of recent examples in which algorithms have been used to pursue public goals in the United States. We used four case studies as anchors for discussion: pretrial risk assessments in criminal justice; screening algorithms used in child protection; an algorithm that proposed new school start times in a large urban school district; and an algorithm that predicted high-risk infections in healthcare. The discussion involved practitioners who worked with each of these systems as well as scholars from a wide range of disciplinary perspectives, ranging from computer science to anthropology, sociology, social work, psychology, philosophy, law, public policy, design, and beyond. We aimed to build shared understanding of the challenges that have recurred across different real-world experiences of high-stakes predictive systems in the public sector.

These were not cases in which organizations simply followed long-established practice. Instead, they were sites of innovation, where practitioners were exploring new ways to leverage the tools of data science to pursue organizational goals. A major focus in each case was investigating how the tools of data science might best be leveraged to improve upon an existing system—realizing that no alternative is likely to be perfect.

Problem discovery and definition was a primary area of shared concern across the different cases. Defining the problem is a high-impact, upstream step in the practice of automated prediction. However, this part of the process is often not explicitly documented or studied. This represents an opportunity for further inquiry and, potentially, improved governance. In training new data scientists—and potentially leaders in other fields—attention might be given to the skills involved in framing a problem, so that the tools of data science can best be used to reach beneficial results. At the time of development, the question could be: "how can this problem (best) be formalized?"' In hindsight: "How did we arrive at this definition of the problem?" Or more broadly: "Are there other problems we might have solved instead?" Qualitative methods can help shed light on specific cases. This is a necessarily collaborative process that must draw on both deep contextual knowledge and technical know-how.

Integration, rather than deployment, was a second major theme. In practice, the degree of success or failure of a predictive system depends on a complex mix of factors including not only the technical properties of a system, but also the social and organizational context that surrounds the technical artifact. What are the incentives and concerns that drive personal and organizational behavior, and will inform and shape the response to a new piece of technology? What kind(s) of visible and invisible labor will the system depend on? What supports can ease an organization's transition to a new system and allow for course-corrections if necessary? Understanding these dynamics—and collaborating with stakeholders so that the resulting system succeeds within its context—is vitally important.

Recent efforts have sought to develop thorough documentation for the data used to train machine learning models and the models themselves (e.g., where the data comes from and the conditions under which a model is expected to perform well), but similar efforts may be valuable in establishing a record of the organizational and institutional processes that led to the model's development, and the social contexts in which it might ideally be used. This might make it easier to identify when models are likely to port well from one setting to another. There is a tension between designing systems that are purpose-built for a specific context and designing systems that are portable across domains, particularly in light of resource constraints.

Governance practices were a third theme across the workshop. Rather than following long-established governance practices, many organizations that begin to embrace predictive analytics today are innovating on two levels—building a new artifact, and also a new suite of practices and relationships to manage the system. These may include participatory governance, expert advisory bodies, or specific constraints—sometimes termed "non-negotiables"—such as a requirement that the model itself be owned by the public institution that is developing the

system. Ownership brings greater control over a model, but building and managing it internally may require a level of organizational maturity that few public bodies possess. As a result, relationships with vendors or consultants are an important site of contestation.

A second recurrent governance concern is managing change. How does the organization and its model detect, and respond to, changing ground realities? One concept proposed for exploration was having "expiration dates" for models.

Predictive systems are developed to respond to difficult public problems—allocating scarce resources, supporting high-stakes decision-making, and other contentious issues with complex politics. The organizations that deploy them may be under-resourced, and the practitioners on the front lines of building them may themselves be operating under constrained discretion. Successful integration of AI into these contexts requires difficult work, a deep understanding of the problem being addressed and the context of use, cultivation of long-term relationships with practitioners and affected communities, and a nuanced understanding of what different technical approaches can and cannot achieve. It also requires sensitivity to the politics that surround these high-stakes applications, as AI increasingly finds itself a mediator of competing political interests and moral commitments. What counts as a "better" prediction or more "optimal" strategy is often as much a political judgment as a technical property.

Solon Barocas, Miranda Bogen, Jon Kleinberg, Karen Levy, Helen Nissenbaum, and David Robinson (co-organizers)