PUBLIC SECTOR DATA ANALYTICS

CASE STUDY: 6

The development of artificial intelligence (AI) systems and their deployment in society gives rise to ethical dilemmas and hard questions. This is one of a set of fictional case studies that are designed to elucidate and prompt discussion about issues in the intersection of AI and Ethics. As educational materials, the case studies were developed out of an interdisciplinary workshop series at Princeton University that began in 2017-18. They are the product of a research collaboration between the University Center for Human Values (UCHV) and the Center for Information Technology Policy (CITP) at Princeton.

For more information, see **http://www.aiethics.princeton.edu**

DIALOGUES ON
AI AND ETHICS

The once-prosperous midwestern American city of New Leviathan has faced major difficulties in recent decades, including deindustrialization, rising racial tensions and a growing budgetary deficit. By the turn of the century, many with the means to emigrate had relocated to wealthier cities in the region. This urban flight exacerbated the city's financial woes, culminating in significant cutbacks in spending on schools, law enforcement and other essential public services. Residents who stayed behind worried for the ongoing health of their community, as well as their own personal safety.

As social conditions in New Leviathan deteriorated, the rate of violent crime began to rise. These troubling statistics and the atmosphere of fear they produced came to a head in 2014 when an officer from the New Leviathan Police Department (NLPD) fatally shot an unarmed high school student on his way home from class. NLPD union representatives explained that the officer felt his life was in danger at the time of the shooting, but many members of the community disbelieved these claims. Even those who supported the officer's actions wondered quietly amongst themselves whether this tragedy—and recent others like it—could have been avoided if the NLPD had been given sufficient resources such that officers were not chronically overworked and exhausted. Discussions on both sides grew heated and resulted in violent protests that drew national attention to the dual problems of crime and policing in New Leviathan. Wishing to de-escalate the situation and preserve her job, the city mayor, Thalia Hobbes, believed she would need to initiate drastic change in the form of a new violence reduction program.

Mayor Hobbes had lived in New Leviathan all her life and cared deeply for the city and its inhabitants. She wanted to help her hometown, but her options for handling the city's recent surge of violence and the mutual distrust it had fostered between citizens and police were constrained by a tight municipal budget and a tumultuous political climate. Furthermore, she was wary of getting too involved—at least publicly—in the particularly sensitive issue of law enforcement. There had been a long history of friction between the NLPD and the mayor's office, and tensions were then at an all-time high. She would need to be creative if she was going to find policy solutions that were both effective and efficient at addressing crime, and also didn't make it seem as if she was attacking law enforcement directly.

A potential solution to Mayor Hobbes' predicament was presented to her at dinner one evening with a group of childhood friends. One of the guests, Charles Prince, had recently been made CEO of the prestigious management consulting firm, Wales Consulting Group, or WCG. While WCG mostly deals in corporate problem-solving, the firm also has a niche data analytics group in their Public-Sector Division dedicated to helping national, state and local institutions better serve their constituents by "leveraging tech solutions to improve existing services and deliver new ones." Mr. Prince thought that this group might be the answer to New Leviathan's crime problems. He proposed the idea of a collaboration to Mayor Hobbes, who was intrigued but hesitant, noting concerns about price and privacy. Undeterred, he prepared to pitch the project to representatives from WCG's Public-Sector Division the following Monday.

WCG has strict protocols for determining whether or not it will take on a potential client, which meant that, despite Mr. Prince's authority and influence as CEO, he did not have the power to commit WCG to the New Leviathan project on his own. Instead, as per procedure, a meeting was convened with members of several relevant groups—including the Social Responsibility Team and the senior partners of the Public-Sector Division—in which they jointly considered three sets of questions regarding the project proposal:

1. What is the problem being addressed and can a data-driven approach be used to solve it?
2. Can WCG achieve the proposed mission and what technological capabilities would be needed?
3. Does WCG approve of the mission's goal? Do the individuals working on the project approve?

WCG has opted to walk away from several large contracts in the past because they did not pass this initial review. In the New Leviathan case, however, WCG was able to answer each question in the affirmative. The reviewers determined that Mayor Hobbes' crime reduction goal was in keeping with WCG's values and believed they would be able to successfully leverage an algorithmic, data-driven approach in order to achieve that end. Indeed, the group was so strongly in favor of the project that they agreed to offer the firm's services to the city on a pro bono basis, meaning that they would not receive payment for their work. This was a huge boon for Mayor Hobbes, as WCG's services would typically have been too expensive for New Leviathan's already-stretched municipal budget.

Discussion Question #1:

Agents from WCG believe they ask all the right questions during the initial review process for potential new collaborations. Do you agree? Are there other questions they should be asking at this preliminary stage? Who should be involved in the review proceedings?

Discussion Question #2:

It is said there is no such thing as a free lunch. These days, consumers are often warned that if they aren't paying for a product, they are the product. WCG framed its proposal in philanthropic terms, but can you think of non-monetary ways in which they might expect to be compensated for their labor? Data? Experience? Access? Reputational advantages? On the other side, what would it mean for a city government to receive an AI service for free? What kind of obligations might be placed upon them? Can you anticipate any potential downsides to New Leviathan accepting the offer of "free" help?

The next step was for Mayor Hobbes to discuss expectations with WCG. WCG consultants explained to her that the firm was offering to send out a team—consisting of engineers, analysts and lawyers—which would develop software that could be used to trace an individual's ties to criminal offenders and analyze social media accounts in order to gauge the likelihood of that person being involved in a crime, either as a perpetrator or a victim. In technical parlance, the software would perform "social network analysis" (SNA) on the citizens of New Leviathan. WCG would not collect, sell or analyze any data themselves, nor would they store any of the data collected or analyzed during their engagement. Instead, WCG consultants would work with data already being collected by New Leviathan city agencies and teach mayoral office staffers how to perform SNA using advanced AI and machine learning techniques. Rather than entirely ceding policy decisions to intelligent machines, they explained, the goal of this program would be to empower human analysts to work with the models.

WCG's assurances of oversight and control eased the mayor's initial hesitations about allowing an outside contractor to handle sensitive information about the city's inhabitants. After having performed a cost-benefit analysis, Mayor Hobbes was convinced that the advantages of using a customized model to help determine how to efficiently allocate the city's limited resources—especially during these politically and financially troubled times—outweighed any residual potential threats to individual privacy. Thus, using the unusually strong executive powers of New Leviathan's mayor's office, Mayor Hobbes unilaterally decided to retain the firm's services. She signed a contract, granting WCG access to New Leviathan's databases, which consist of millions of searchable public records, court filings, licenses, addresses, phone numbers and social media data. She also gave WCG permission to view the city's criminal databases for information about

ballistics, gangs, probation and parole; jailhouse telephone records; the central case management system and the NLPD's field interview records. WCG would use this data to train its crime forecasting algorithms. Once the algorithms were ready, information from the same databases could be input in to identify New Leviathan residents at risk of being involved in a crime.

Discussion Question #3:

Mayor Hobbes made the decision to accept WCG's proposal without having consulted the citizenry or other officials through democratic processes. In this case, the decision to take independent executive action was legal, but was it justified? If not, how should she have gone about approaching the decision? Who are the relevant stakeholders that might have been consulted?

Discussion Question #4:

A data-driven approach to crime reduction necessarily requires data. To the extent that such approaches are effective at increasing overall safety, some people might be willing to grant outside contractors access to their personal information. Some, however, may prefer not to have their data shared at all. In the New Leviathan case, how should individual privacy interests be weighed against WCG and Mayor Hobbes' expectations that use of individual data would decrease crime and increase safety for all? Are there data collection, use and storage practices that WCG and Mayor Hobbes could employ to reduce potential privacy concerns?

Mayor Hobbes initially chose not to publicize the agreement she had entered into with WCG, but both parties expected the program to eventually be made public (ideally, after it had proven a success). Thus, it was important that their efforts were framed in the right light. Critically, neither Mayor Hobbes nor WCG wanted their project to be billed as "predictive policing." Predictive policing, or the use of algorithmic models to assess the likelihood of individuals or places being involved in a crime in order allocate police resources efficiently, had become an extremely controversial practice. Several American law enforcement agencies had recently experimented with these programs only to be attacked for civil liberties violations, racist/classist practices and allegations of ineffectiveness. In order to distance New Leviathan from these public relations disasters, WCG designed their system to be distinct from standard predictive policing in two ways. First, it would focus on identifying potential victims of crimes, rather than perpetrators. Second, the policy recommendations WCG suggested would be limited to mobilizing the city's social support services in aid of at-risk citizens (e.g. increasing welfare checks). The police wouldn't be involved at all.

Initially, the program showed some success in crime reduction. Two years after the collaboration with WCG began, statistics revealed a modest decrease in gun violence and murders in New Leviathan. This downward trend didn't last long, however, and crime stats began to slowly creep back up. Some staffers suggested that this might be due to the mayor's conservative use policies regarding the AI system. With an election looming, Mayor Hobbes was convinced to take a more aggressive approach. In order to push crime reduction along more quickly, the city would begin targeting all those that WCG's algorithm deemed likely to be involved in a crime – now including potential perpetrators. And rather than focusing exclusively on community outreach and social services, as the program had originally been designed, she opted to involve law enforcement as well. Mayor Hobbes proposed that a list of all individuals determined to be at high risk of committing a crime be assembled and made available to the NLPD. The NLPD could summon potential offenders from this list to police stations for interrogation. Officers could draw from the familiar toolbox of carrots and sticks in order to discourage these individuals from committing future crimes.

Discussion Question #5:

What do you think of Mayor Hobbes and WCG's initial efforts to distance themselves from the term "predictive policing"? Did the original program differ meaningfully from predictive policing programs in the past? What about the new plan?

Within months of instituting these changes, news about the program leaked. An online investigative reporter, J. Wallis, published a scathing exposé on Mayor Hobbes and WCG's handling of data about New Leviathan and its citizens, in which she did not shy away from calling the program "predictive policing." Wallis was critical of the AI system overall but reserved particular venom for the secrecy surrounding its origins. Outside of the mayor's office, she couldn't find a single public actor who admitted prior knowledge. The mayor's office did not deny that they had kept their dealings with WCG quiet, and many locals and government officials were outraged at having been kept in the dark. Wallis quoted a popular city councilman, John Bramhall, Jr., saying, "I'm all for adapting to the times. I would gladly embrace a responsible data-driven approach to crime reduction if that's what it takes to get our city back on track, but I'm deeply uncomfortable about the level of secrecy used in this instance. City officials have a right to know when policies are being changed. The **people** have a right to know how their government is making decisions about them!"

Many New Leviathan citizens and their allies agreed with Councilman Bramhall's sentiments, adding several ethical objections of their own. None of these voices were louder or more powerful than those, which came from a group of WCG employees that had rallied to the side of city residents. Ranging from engineers and product developers to lawyers and consultants, this group claimed to be shocked at the company's willingness to bypass consent in furtherance of a program that could inflict meaningful harms on the people of New Leviathan and beyond. Many threatened to resign over the scandal, and together, they released a public letter demanding WCG immediately cease its work with New Leviathan, at least until the ethical issues at stake could be fully considered.

Ethical Objection #1: Government Secrecy and Individual Privacy

Both residents and city officials were angry that the mayor's office hadn't informed or consulted them about its actions. Because the WCG deal wasn't public knowledge, the people of New Leviathan had not been given the opportunity to ask questions about the resulting algorithm's basic functions, risk of bias and overall appropriateness. Many believed Mayor Hobbes' choice to insulate the program from public debate undermined the notion of popular sovereignty, or the idea that governments are responsible to the people from whom they derive their authority to act. Some citizens also pointed out that this secrecy was hypocritical, given how much of their own personal information had been shared with WCG without their explicit consent. Privacy was beginning to look like a luxury reserved only for the political elite.

Ethical Objection #2: Inequality, Injustice and Ineffectiveness

Mayor Hobbes and WCG maintained that they were not engaged in predictive policing, but the public wasn't so sure. Due to the lack of information about the New Leviathan program, it was difficult for residents to see how it may have differed from those predictive policing programs that had recently been in the news. And research into those programs had produced a generally dim view of them. Some studies found that predictive policing may have a disparate negative impact on poor and minority

communities, while others called into question their efficacy. To the extent that the New Leviathan program might disproportionately target poor and minority residents, many citizens thought it ought to be discontinued. Such practices were not only unfair and unjust in an unequal society, they argued, but were also likely to exacerbate the already high social tensions in New Leviathan. Even those who were skeptical of the claim that the program was inegalitarian in practice argued that the risk of harming poor and minority residents would be unjustified if the program failed to achieve its stated aim of making the city a safer place. So far, evidence of the program's success was weak at best.

Ethical Objection #3: Civil Liberties and Autonomy Infringements

The ACLU came to New Leviathan to join the fight. Lawyers from the organization reminded New Leviathan's political elite that, in the United States, citizens must be treated as innocent until proven guilty. They argued that the use of algorithms to determine who is likely to be involved in a crime—especially when accompanied by policies that target those individuals for special treatment—undermines this essential tenet of the American legal system, as well as the underlying notions of institutional fallibility and equal respect for all. Many of the locals agreed, saying that the algorithms designed and implemented by WCG and Mayor Hobbes, respectively, had no place in the American criminal justice system, which must protect civil rights and civil liberties. To this constitutional claim, some more philosophically-minded critics added the argument that to treat individuals according to their statistical probabilities erodes their status as autonomous agents with free will. In other words, it treats human life as deterministic. Many protestors, including Mayor Hobbes' college-age daughter, were seen wearing t-shirts emblazoned with the words, "I am not my probabilities!"

Mayor Hobbes responded to this criticism by insisting that the AI system had been necessary to secure residents' safety in the wake of the 2014 protests and pointed to the post-implementation dip in violent crime as proof of its success. She defended her decision not to disclose information about the WCG collaboration by citing the tense political climate of that time. Had she been forced to "play politics" under such conditions, she argued, she would have been unable to adequately serve the public interest. Furthermore, in an unguarded interview, Mayor Hobbes questioned the very value of public disclosure. She doubted that many of her constituents would have understood the complex AI system, even had she shared it with them. Those who did understand the algorithm posed a threat in that they might try to game the system. Thus, Mayor Hobbes believed her best option to give the program a fair chance was to act independently and quietly. And, as she pointed out, New Leviathan's political institutions were on her side. New Leviathan had long ago embraced the model of a powerful executive. People may not have liked the solutions she adopted or the secrecy with which she did so, but she acted within the legal bounds of her position and in furtherance of what she believed to be the ultimate good of the people she served.

As to the claim that the program developed with WCG might have had an unequal impact on different members of the community, Mayor Hobbes pointed out that this was an empirical question, and therefore one that could not be answered until the city had done a proper accounting of the effects of WCG's proposed interventions. That could take years. However, she strongly refuted any and all attempts to categorize the project as predictive policing. Even after the original program had been revised to target potential criminal offenders (not just likely victims) and refer them to law enforcement (not just social services), she continued to maintain that her office was merely engaging in data analytics, which were necessary for the efficient resource allocation demanded by shrinking budgets and rising crime stats.

WCG was also compelled to defend its participation in the New Leviathan project. The Wallis exposé revealed that WCG had been using its experience in New Leviathan to market its crime reduction capacities to other cities. Whether the New Leviathan pilot program had been successful or not, the algorithms trained on that

city's data were growing more accurate every day and were now quite valuable. It was discovered that a certain South American nation had already signed a contract with WCG using the technology developed with New Leviathan as part of its anti-terrorism program. To the extent that WCG financially benefitted from the collaboration with Mayor Hobbes, many residents of New Leviathan felt the firm owed them explanations and justifications for how their data had been used.

Representatives from WCG responded to these disclosure requests by reiterating Mayor Hobbes' claim that they were not engaged in predictive policing. And echoing Mayor Hobbes's comments about the complexity of the system's design, they argued that they were unable to explain to citizens exactly how their data had been used. However, these WCG representatives insisted that everything possible had been done to keep the training data anonymous in order to protect individual privacy.

Responding to members of its employee "uprising," WCG's public relations team defended the ethicality of the New Leviathan collaboration. As they reminded the aggrieved employees, beyond the preliminary ethics review, WCG requires project teams to assess their assignments and their impacts regularly over the course of the engagement in order to determine whether or not the relationship should continue. At the end of each year, or when a major change to the program has occurred, members of the team must meet to discuss four questions:

1. Has the broader context changed, such that WCG's services are no longer needed or appropriate?
2. Have the nature of the institutions evolved such that WCG no longer wishes to support them (e.g. a change in political leadership, widening of the original mandate)?
3. Has there has been any unacceptable or "repugnant" use of their products?
4. Does the team still support the project?

This procedure builds WCG's confidence that its collaborations are and remain ethically sound. The New Leviathan project passed not only the initial ethics review, but all subsequent reviews as well. And in fact, WCG team members had just recently performed an ethics audit of the New Leviathan project following Mayor Hobbes' decision to expand the program to target potential offenders and involve law enforcement. While the team had some hesitations about the way their AI products were now being used, the review ultimately concluded that the project remained ethically sound and that they wished to support it.

Representatives from WCG admitted that the firm rarely walks away from a project after one of these interim reviews. However, they argued that that is only because the initial weeding out process is so rigorous that it almost always catches potential ethical problems before entering into a contract. WCG wished to add that the firm remains proud of its ethics protocols and plans to do even more going forward. Members of their Social Responsibility Team recently developed a framework for an internal ethics process that is transparent and simple enough so that all members of the organization are able to use it. In the future, they hope to institute a formal ethics educational program within the company – ideally one that could be scaled and exported to other consultancies addressing similar political and ethical dilemmas.

Discussion Question #6:

Should the New Leviathan collaboration have passed WCG's interim ethics reviews? How did the most recent review differ from the first? What would need to be included in the firm's ethics protocols to make a sound ethical review at all stages? More broadly, can a corporation's internal ethics review provide sufficient evidence that a project is "ethically sound"? What other procedures might be needed to make such a determination?

Discussion Question #7:

WCG plans to teach ethics to its employees. How can one effectively operationalize values in a company like WCG and the IT systems it produces? Why is this necessary? Is it necessary?

Reflection & Discussion Questions

Democracy: Like the broader United States in which it is situated, New Leviathan has a democratic system of governance with checks and balances. However, the office of mayor in New Leviathan has been vested with an unusual degree of authority to make decisions apart from her constituents and the other branches of government. Some support this distribution of executive power on the belief that a strong central authority is the best way to protect the people and keep them safe. Others are more skeptical. While a strong executive may make decisions in the best interest of the people, the people's role in determining what those interests are (i.e., the ends they wish to pursue) and the means for achieving them is diminished under an authoritarian leadership model. We often see this debate paralleled in the tech world, in which developers and proprietors of AI systems must determine the appropriate balance between top-down and bottom-up decision-making procedures.

- In teaming up with New Leviathan, WCG ostensibly aimed to make the city a better place. Do companies that claim to be developing AI products to improve public welfare have a responsibility to consult with the people they purport to serve – either by securing their approval or actively endeavoring to better understand the community in order to improve their products?
- Do democratic governments have special responsibilities to involve the people, as well as existing government officials and processes, in decision-making surrounding the use of AI that go beyond those of private corporations? (See, for example, calls for the use of privacy commissions to asses AI.) Do democratic governments have special responsibilities to be transparent about their use of AI that go beyond their obligations to inform citizens of their non-AI practices?
- American democratic institutions are designed to safeguard civil rights and liberties against threats from both powerful leaders and populist impulses. Some critics of the New Leviathan program argued that, by granting WCG access to the city's databases, Mayor Hobbes violated their right to privacy and made them vulnerable to corporate influence. Do you agree? If so, can you think of examples where the government sharing such information might be appropriate? Would the tradeoffs be any different if, for example, WCG's algorithms were used to target potential terrorists who are not US citizens, and therefore, not entitled to the same legal protections? What about the South American nation that recently contracted with WCG to perform the same services as in New Leviathan?

Secrecy: The criticism Mayor Hobbes faced often had less to do with the fact that she acted alone, and more to do with the secrecy surrounding her actions. Given the unique position of law enforcement to impact the lives of civilians, some could argue that the city was morally and socially obligated to reveal the terms of the WCG contract and the scope of its mission (if not also the system's technical details, such as the algorithm itself). Such would be the requirements for procedural justice. But while openness may sound like it's always a good idea, there are reasons for secrecy in the policy world. For example, Mayor Hobbes explained that she did not wish to divulge information about the new AI system lest residents discover how to subvert WCG's algorithms and skew results.

- What do you make of the claim that AI systems must be opaque to function effectively? Can you think of other legitimate reasons why secrecy might be appropriate regarding New Leviathan's collaboration with WCG? If you believe secrecy is never appropriate, defend that view.
- Some New Leviathan protestors claimed that government expectations of secrecy are hypocritical in matters where the state shares private data about its citizens without their consent and which may be used against them. How would you engage with this view? Are secrecy and privacy the same thing? How might the concepts differ?

Inequality: One criticism against predictive policing (and programs like it) is that it disproportionately impacts poor and minority neighborhoods. In part, this is a consequence of skewed data collection. Police departments tend to have good data about the communities they already patrol, but they may have little information about communities with a lighter police presence. For a variety of reasons, the former neighborhoods tend to be poor and minority, while the latter tend to be affluent and white. When this unbalanced data is input into an algorithm for assessing risk, the results may encourage the allocation of more law enforcement resources to some neighborhoods over others. This can lead to increased arrests in those neighborhoods, as well as hostility from individuals who feel they're being unfairly targeted.

- How might a crime prediction algorithm be designed to minimize inegalitarian outputs based on biased data? If tech solutions are unavailable or insufficient, can you imagine public policy solutions that could mitigate against unjust treatment of poor and minority neighborhoods?

- If algorithms predict that people in poor and minority communities are more likely to be involved in a crime, and if targeting interventions at members of these communities is proven effective at reducing crime overall, would the state be justified in doing so? What countervailing values might you consider? For example, how would this approach fare against traditional notions of justice, which insist that punishments fit the crime and not an individual's potential for crime?

Fallibility: Mayor Hobbes was impressed by the high predictive accuracy of WCG's algorithm, which promised to save the city money by enabling her to focus crime reduction efforts on high-risk individuals. However, it is important to keep in mind the limits of certainty in even highly advanced AI systems. As with traditional statistics, the probabilities produced by algorithmic models are just that – probable outcomes. They are not certs. And while they may tell us much about populations, they reveal less about individuals. Even a 99 percent chance that someone will be involved in a crime leaves a one percent chance that he will not, as well as some margin of error. And this is assuming the model itself is flawless and accounts for all variables. This is rarely (if ever) the case. However, the uncertainty inherent in predictive models is not always clear to clients, who may accept algorithmic outputs as truths.

- Why is it important that people using AI systems understand their fallibility? What are some things AI developers and proprietors could do to make the limitations of their models clearer?

- The supposed infallibility of scoring algorithms may encourage people to substitute their results for qualitative judgment and human responsibility. What are the implications of deferring to an algorithm's outputs, especially in areas as important as law enforcement? In answering this question, think in both the long- and short-term.

Determinism: Underlying the "I am not my probabilities!" movement was the belief that humans are autonomous agents with free will. According to this view, people are not destined to be or do any one thing. While risk assessment algorithms do not necessarily contradict the idea of free will, in practice, they may undermine autonomy. Labeling an individual "at risk" encourages others to think of her in those terms, increasing the likelihood that she will live up to the label she's been given. In the case of New Leviathan, they city may not have prosecuted citizens deemed likely to commit a crime on basis of that prediction alone, but it did treat them differently (i.e., sending in social services, calling them in to police stations). These interventions may then have influenced the way that such individuals behaved going forward – perhaps nudging them towards riskier behaviors.

- In cases where interventions based on algorithmic predictions still result in negative outcomes, what, if any, moral responsibility do the New Leviathan program and the various relevant actors (Mayor Hobbes, NLPD, WCG) bear for those outcomes?

- Humans engage in evaluative judgments all the time, naming some people "bad seeds" and steering clear. Is AI labeling meaningfully different, or is this just more of the same?

AI Ethics Themes:

Democracy

Secrecy

Inequality

Fallibility

Determinism