URBAN AI GUIDE

AUTHORS:
Sarah Popelka
Laura Narvaez Zertuche
Hubert Beroche

Illustrations and layout by Alizée Sire

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

We would like to express our deep gratitude for everyone who helped make this guide happen.

First, we would like to thank the urban AI experts and municipal employees who took the time to be interviewed and/or otherwise provided valuable insights for the guide: Patrick Agin (City of Montreal, Canada), Yossef Bourimech (City of Montreal, Canada), Emri Brickner (formerly City of Be'er Sheva, Israel), Tomas Carranza (City of Los Angeles, United States), Maxime Dusart (UPCITI), Caroline Goulard (Dataveyes), Robert Heinecke (Breeze), Ernest Kwan (City of Winnipeg, Canada), Katya Letunovsky (Habidatum), Maksym Nikolayev (BuilderZ), Sanit Sananikone (City of Montreal, Canada), and Simas Sodys (City of Vilnius, Lithuania).

Second, we would like to thank the members of the Urban AI community, who reviewed the draft and provided thoughtful guidance on the content and structure: Alain Chiaradia, Ruth Nelson, Ariel Noyman, and Iacopo Testi.

Finally, we would like to thank The GovLab, which served as a valuable project partner and aided us in making connections with some of the municipalities interviewed.

EXECUTIVE SUMMARY

More and more, urban practitioners seek informational and technological solutions to the enormous challenges that cities face. The proliferation of big data, coupled with massive advancements in the field of artificial intelligence, has propelled municipalities into the tech development space, often with limited resources and knowledge. The result, so-called "urban artificial intelligence" (referring to any system that incorporates data derived from the urban environment, which is then processed by algorithms, the result of which has useful applications in the socio-spatial nexus of the city), has three distinguishing features which necessitate a deeper understanding of such systems: the complexity of the city, the specific policy contexts, in which urban artificial intelligence operates, and the hybridity of urban artificial intelligence. We at Urban AI, a Paris-based think tank, designed this guide for city leaders and urban technologists (academic, public, private, and community-focused) to better understand how artificial intelligence operates in urban contexts, as well as considerations for better evaluating AI-based project implementations and maintaining human oversight in the process.

This guide dissects an urban AI system, exposing its eight core components. In doing so, it presents a framework for understanding the steps and considerations that go into implementing an artificial intelligence project in an urban context. Without *urban infrastructures*, i.e. the intertwined layers of the built and physical environment, flows, governance, and people, there would be no city. The multiplicity of stakeholders operating in metropolitan areas requires the identification and consideration of many actors. *Sensors and data collection infrastructures* allow for the capture of information emitted and transmitted by urban systems and processes. These can range from small, community-led efforts to massive city-wide sensor networks. *Network infrastructures* transfer information to *data storage infrastructures*, which allow for the organization and retrieval of data. Data privacy, security, and size drive decisions around the storage of data. Due to the unpredictability of a real-world environment, it is rare that incoming data will have an adequate format and no errors; data processing facilitates the assessment and cleaning of raw data. *Data visualization aids* analysts in better identifying patterns and communicating information. *Artificial intelligence/ machine learning,* can be utilized to enhance the completion of any of these steps, as well as for the implementation of analysis and predictions. Finally, *decision-making/ adaptation* transforms the insights gained in the process into action, in a continuously evaluated process.

Through the use of detailed case studies from three different cities, we demonstrate what urban artificial intelligence can look like, practically, and how it can be applied in different disciplines and urban contexts. The water meter lifecycle project in Winnipeg, Canada outlines the considerations necessary to build out an internal artificial intelligence team and shows how municipalities can take a phased approach to project implementation while they build capacity. The curb digitization for planning and asset management project in Los Angeles, United States highlights the role that a public private partnership can play in expediting project timelines, and underscores the importance of scaling the solution to the specific problem and urban context. The air quality sensing project in Vilnius, Lithuania shows how platform/ package-based AI service providers can provide AI capabilities with little overhead on the part of the city. Future iterations of this guide will further expand the geographic and sectoral scope of case studies presented.

The literature review, urban AI anatomy, and case studies all fundamentally highlight the role of human beings in implementing, overseeing, and evaluating urban AI projects. Since urban AI projects utilize human-based data, the physical components of urban AI systems interact with the human-built environment, and the decision-making that amounts from an urban AI project affects humans, people must remain actively involved in each component of the project. In this sense, communication plays a key role: engaging with the community at key moments, building intelligibility mechanisms such that algorithms do not exist as black boxes, and conveying an understanding of the system, holistically, to residents and decision-makers alike. Effective and responsible change does not arise from reacting to model outputs at face value. Instead, urban practitioners ought to remain actively involved in crafting, evaluating, and re-designing urban AI processes. Urban artificial intelligence holds its greatest value not as a means of *making* decisions, but as a means of *informing* them.

WHAT IS URBAN AI (THE THINK TANK)?

Urban AI is a Paris-based think tank dedicated to the emerging field of "Urban Artificial Intelligence." Given the highly complex and human-centered nature of urban systems, scholarship and practical implementation of artificial intelligence technologies in cities requires a uniquely multidisciplinary lens. For this reason, Urban AI seeks to generate a holistic body of knowledge on urban artificial intelligence by federating and collaborating with a growing, global community of researchers, public servants, start-ups, and urban subject matter experts, who work at the intersection of cities and technology. Together, we carry out multidisciplinary projects to better understand and assess the impacts of artificial intelligence on urban life and vice versa. Urban AI's work follows three primary streams: research, events & content, and education.

Research: The Urban AI research program carries out long term projects oriented toward implementing our Manifesto and urbanizing Artificial Intelligence. Urban AI's research outputs mainly focus on the intersection of localized technologies, open cities, decentralized technologies, frictional design, meaningful technologies, ecological technologies.

Events & Content: Urban AI fosters a space for experts in the field of urban artificial intelligence to share their insights, debate and exchange. We regularly organize events and publish content about considerations for and breakthroughs in Urban Artificial Intelligence. These activities aim to democratize this emerging field and to actively contribute to its development through a multidisciplinary and intercultural dialogue.

Education: The Urban AI educational program develops and supports scholars and professionals working in the emerging field of Urban Artificial Intelligence. The Emerging Leaders Program seeks to provide a theoretical basis, practical experience, and mentorship for young people who wish to pursue work or research in urban-oriented artificial intelligence fields. Additionally, we train municipalities and other urban stakeholders on how to think about and evaluate the utility and implementation of artificial intelligence in their operations. Through these programs, we aim to share the expertise and knowledge that we continue to accumulate as an institution.

WHY DID WE DECIDE TO MAKE THIS GUIDE?

As some of the most renowned international organization reports have predicted, more than fifty percent of the population globally are living in urban areas (UN DESA 2021). By 2050, more than two-thirds of the population will live in cities, providing big investment opportunities for tech development companies. In a more recent prediction from IDC, by 2021, the spending on urban technology is likely to reach more than $130 billion, with the intention of making urban areas and cities more livable by implementing advanced infrastructural facilities and society management systems (Jyoti et al. 2021). In this sense, cities worldwide are not just growing, but also trying to reconfigure themselves for a more sustainable and 'smarter' future, with an increased emphasis on providing a higher quality of life for every citizen. Cities are, in essence, home to many of the key requirements for digital transformation, including digital innovators and ecosystems, data, and infrastructure.

In parallel, artificial intelligence (AI) has generated a staggering amount of hype in the past several years. Just as in ancient Rome in the first century AD the invention of aqueducts was critical for population growth, AI is now emerging as an essential part of how cities work. It offers the potential for creating environments that are more responsive to their users' needs at all scales, including the urban scale.

Cities have been collecting vast amounts of data, and companies like Google have been providing functions like Street View and Earth for years -and more recently 3D immersive view- to create a rich, digital model of the world. AI can harness all this information, analyze it, and use it to help make cities work better. Such data are the raw material that urbanists like Kevin Lynch and William H. Whyte had to generate periodically and painstakingly to develop their ideas; now, these data are continually updated and available in real time.

Digital technologies can be used as levers to develop urban spaces while protecting the environment and ensuring a high quality of life. Scientists assume that cities around the world will be among the largest beneficiaries of these new technologies (Allam and Dhunny, 2019), because it is in cities that large amounts of data, computing power, and digital innovation ecosystems are concentrated alongside the kinds of problems these technologies can help to solve.

However, with the advance of the current digital revolution, a large proportion of practitioners and scholars have increased their belief

that smart urban technologies can mark a major turning point in the history of humankind (Tomitsch and Haeusler, 2015). This technocentric approach to solving urban problems has increased the theories and application of the notion of 'intelligent cities', more popularly known as 'smart cities'. These cities, also known as 'geographies of disruption' (Yigitcanlar and Kamruzzaman, 2018, p.51) harness digital technologies to offer new opportunities, shape the urban fabric, and improve the quality and performance of urban areas (Yigitcanlar and Inkinen, 2019).

The prospects of smart urban technologies range from expanding infrastructure capacity to generating new services, improving decision-making, and supporting the performance of businesses and cities. The most popular technologies of intelligent cities include autonomous vehicles, Internet of Things (IoT), virtual reality, digital twins, robotics, big data, blockchain, and artificial intelligence (AI). AI has been considered the most disruptive technology of all (King et al., 2017; Tegmark, 2017). Whilst all these technologies have been critical in transforming our cities into smarter ones, AI has significant potential to address the urbanization challenges we face in cities.

With the growing interest in the application of AI for urban innovation, this article explores how urban technologies, largely embedded in the development of AI, intersect with the development of cities. Essentially, what is proposed as *urban artificial intelligence* is what AI brings to cities and what cities bring to the advancement of AI.

For artificial intelligence to make a positive contribution to cities, it is vitally important to understand what AI means in an urban context, as well as review its evolving applications to date. However, many urban stakeholders, including local government officials, find themselves in a position to weigh in on the use of artificial intelligence in a given project, yet lack the knowledge to properly weigh in.

This guide dissects an urban AI system, exposing its core components. In doing so, it presents a framework for understanding the steps and considerations that go into implementing an artificial intelligence project in an urban context. Through the use of case studies from three different cities, we demonstrate what urban artificial intelligence can look like practically and how it can be applied in different disciplines, as well as how it can impact people and cities. This guide represents a first-of-its-kind effort to better understand what implementing artificial intelligence in cities really means.

WHO IS THIS GUIDE FOR AND HOW DO WE ENVISION IT BEING USED?

Cities today are facing enormous challenges, including an aging society, population growth and rapid urbanization, environmental stresses, infrastructural changes, and increasingly limited resources. It falls on all practitioners operating in the built environment to help shape the response to such challenges. At the same time, we have only just begun to explore the benefits that modern technologies (e.g. sensors, pervasive communications, robotics, cloud computing) and new business models can bring to cities, and to take advantage of the opportunities available through information (e.g. real-time data, analytics, visualization, personalization and social engagement). These benefits can only be fully realized by ensuring that technologies move beyond the limits of conventional service functions towards a more active and open collaborative working culture.

We designed this guide with city leaders and urban technologists (academic, public, private, and community-focused) in mind, but it is for anyone who wishes to understand how artificial intelligence operates in urban contexts. This guide provides an overview of the components that make up an artificial intelligence implementation, as well as how AI can be (and has been) applied in cities. It highlights the elements that local governments must consider when embarking on an AI project by showcasing three different case studies. We hope that it helps provide practical recommendations and highlights the ways in which the application of AI can help eliminate risks, lower costs, and/or reduce the effort required to manage cities effectively, while putting its citizens at the core of the process.

LITERATURE REVIEW

Although much has been written on artificial intelligence and urban governance separately, comparatively little has been written on urban artificial intelligence, which is to say the intersection between the two, beyond the concept of the "smart city." Much of the literature that covers this topic revolves around either high-level discussions of use cases for artificial intelligence implementations in urban contexts or highly technical descriptions of specific algorithms used (Woetzel, et al, 2018). While such papers prove indispensable in providing inspiration for projects and guiding technical development, they don't necessarily provide a blueprint for cities that aim to implement their own artificial intelligence-based solutions.

However, a handful of papers, which do provide more of a mid-level framing of artificial intelligence in urban settings, have informed the production of this guide in various ways. On the topic of artificial intelligence more broadly, Crawford and Joler (2018) trace the complete anatomy of the Amazon Alexa as an AI system, from the initial resource extraction to produce the materials for the device to the final decomposition at the end of life. This guide takes inspiration from their approach in mapping out an anatomy of urban AI project implementation, although their anatomy does not focus specifically on operationalizing AI, nor does it focus on urban contexts. Cugurullo (2020) explores the continuum of automation to autonomy and clearly lays out definitions of the main artificial intelligence concepts and components that contribute to automated vehicles, robotics, and automated decision making. Taking a more specific approach, Burgess (2018), in his book, "The Executive Guide to Artificial Intelligence," lays out a taxonomy of AI techniques, which provides a digestible basis for understanding the broad purposes and uses of such tools. To this end, he groups algorithms into three main functions, with each comprised of a set of capabilities: capturing information (which uses image recognition, speech recognition, search, and/or clustering), working out what is happening (which uses natural language understanding, optimization, and/or prediction), and understanding why something is happening (which uses understanding) (ibid). Wu and Silva (2010) completed a similar survey exercise based on algorithms used in urban land dynamics analyses, in which they found that any broad categories of artificial intelligence capabilities contain many different algorithms with distinct specificities, the appropriate selection of which for any given case can require deep technical knowledge (Wu and Silva 2010).

Beyond technological considerations, the AI literature also provides for

ethical contemplations, albeit to a lesser extent. Floridi, Cowls, King, and Taddeo (2020) describe seven essential factors to design "artificial intelligence for social good," which include: falsifiability and incremental deployment, safeguards against the manipulation of predictors, receiver-contextualized intervention, receiver-contextualized explanation and transparent processes, privacy protection and data subject consent, situational fairness, and human-friendly semanticization; additionally, they provide a case study-illustrated guide for best practices that AI designers can utilize to implement these factors (Floridi et al. 2020). In a guide on ethical artificial intelligence implementations in museum settings, Murphy and Villaespesa (2020) present an "AI Ethics Workflow," which consists of checks and ethical considerations that should be taken into account at various steps in the implementation process. Given the particularities of the ethical considerations of AI, Verhulst, Young, and Sloane (2020) point out the power that cities, uniquely, have in regulating artificial intelligence technology - an approach that they deem "AI Localism." Additionally, Gordon and Guarna (2022) explain the interplay between technology and government trust, detailing actions that cities can take to foster trust from constituents when implementing AI and other technologies.

Numerous papers discuss smart cities, both through the lens of artificial intelligence and with a more specific focus on internet-of-things (IoT) networked sensor systems. Yigitcanlar, Desouza, Butler, and Roozkhosh (2020) present a seminal literature review of 93 articles on the ways that artificial intelligence contributes to smart cities. They find that artificial intelligence has a strong possibility to make a positive impact, but that the existing literature on the intersections between AI and smart cities focuses primarily on artificial intelligence technologies, with little written about social implications (ibid). Yigitcanlar et al. (2020) contribute the best practices of transparency, adaptability, accountability, and diversity. Allam and Dhunny (2019) create a framework, which imagines smart cities as sitting at the intersection of culture, metabolism, and governance, built on an infrastructure of IoT technology generating big data, which in turn feeds artificial intelligence. Macrorie, Marvin, and While (2019) specifically explore the implications of robotics and automation in urban contexts, identifying two main categories that operate together in the urban domain: automated system management and robotization of urban services. They urge practitioners to utilize a critical lens when considering the utility and implications of automated devices and processes (ibid). In line with this perception, Cugurullo (2020) utilizes a case study of Masdar City in Abu Dhabi to highlight the embodied nature of artificial intelligence, as manifested through autonomous vehicles, robots, and the "city brain" (i.e. automated decision making and governance). Fernandez-Anez, Fernandez-Guell, and Giffinger (2018) examine smart city project governance and implementation through the case study of Vienna, proposing a framework for understanding differences between implementation and discourses surrounding "stakeholders' & governance core role," "smart city projects and dimensions," and "urban challenges

and global trends." Koseki et al. (2022), in a UN Habitat report on AI in Cities, provide a highly detailed guide on applications of artificial intelligence in the context of various urban systems and issues.

However, as Yigitcanlar et al. (2021) note, urban artificial intelligence implementations have extended beyond the concept of the smart city, being of interest to a number of municipalities regardless of their "smart city aspirations." In this context, they identify key issues that governments need to address to implement AI more responsibly: identifying unique challenges around data at the local level; determining the areas where AI could have the biggest impact; understanding the challenges that local governments face around data privacy, transparency, and security; developing an AI-ready local government workforce; and having a responsive lens when deploying and managing AI technologies" (ibid). However, their framework for responsible innovation is conceptual, rather than operational. Eckroth (2018) provides a comprehensive guide to building and deploying AI in the business sector, which contains many workflow components and considerations that hold relevance in the public sector as well. Similarly, Burgess (2018) enumerates many of the necessary components that a firm must incorporate in order to successfully implement an artificial intelligence-related project, as well as the considerations and challenges associated with each. Saltz and Heckman (2015) examine big data science projects more broadly, but in doing so provide a more generalizable set of project steps. Muller et al. (2019) utilized a semi-structured interview approach to better understand how data scientists at IBM work with data, focusing on descriptions, specific completed project, a methodological approach which partially informed the manner of inclusion of case studies in this guide.

Finally, Beroche et al. (2019) synthesize a massive report on urban artificial intelligence, which consolidates insights and case studies from cities around the world, which use AI to improve their operations. The report incorporates viewpoints from firms and start-ups, which build or use AI solutions in urban contexts as well, providing an additional layer of insight beyond the typical municipal survey (ibid). Among the breadth of knowledge that Beroche et al. provide in their Urban AI Report, they highlight the co-production of cities and artificial intelligence (i.e. the fact that cities provide the basis for urban AI implementations, which in turn provide the basis for city innovations), an interpretation echoed by Cugurullo (2020) and Macrorie et al. (2019). The intertwined relationship certainly requires a deep understanding of the nuances of complex urban spaces and artificial intelligence considerations, which this guide aims to provide for a diverse set of urban stakeholders from a wide array of technical backgrounds, in order to better facilitate the implementation of artificial intelligence technologies in urban settings.

DEFINITION OF URBAN ARTIFICIAL INTELLIGENCE

Some scholars argue that the field of artificial intelligence was founded at a workshop held on the campus of Dartmouth College during the summer of 1956. At that time, it was predicted that a machine as intelligent as a human being could exist with the purpose of performing certain tasks. This definition is often ascribed to Marvin Minsky and John McCarthy, who are considered as the founders of AI along with Alan Turing, Allen Newell, and Herbert A. Simon (Anyoha, 2017). Turing suggested that if humans use information and reason to solve problems and make decisions, then why can this not be done by machines?

AI, in a general sense, is a way to describe the efforts by scientists to teach machines how to learn independently (Bostrom, 2017), or in other words, to automate cognitive processes such as pattern recognition, planning, language, and text or speech recognition. Therefore, AI is not a single technology (Cave, S., Dihal, K., and Dillon, S., 2020) but a range of technical processes, most of which use variations of machine learning (ML) (Bishop, 2006). In these technical processes, data and learning algorithms, or code, are incorporated into a software model, from which probability-based conclusions are then drawn, supported by powerful hardware. Let us break this down even further. *Artificial* means that it is an *artifact*; it is not made by natural processes. It is the outcome of an artificial process. *Intelligence* can be considered the *ability to acquire and apply knowledge and skills*. It is an outcome of a learning process as information is acquired within a certain environment. This leads eventually to making decisions autonomously, which might resonate with what we normally refer to as 'thinking' (Bostrom, 2017; Warwick and Shah, 2016), a perspective in the AI field.

In the modern world, we are surrounded by AI. From assistants such as Amazon's Alexa to the internet predicting what we might like to buy next, AI is found everywhere. However, cities are complex systems; they are centers of human interactions and economic transactions as well as the innovations that arise from them (Batty, 2009). The development of AI in urban development has often been focused on the development of the technologies themselves and how they can be applied to cities. But the real impact of AI in cities is not on the technology but on its implementation in urban planning and design. It is in the plan-making process of cities that AI, in the form of machine learning, has its major impact (Batty, 2018). This is where urban artificial intelligence is born.

An extensive report by Hubert Beroche (2021) sought to derive a definition and understanding of urban artificial intelligence. Exploring 12 cities and with more than 130 actors involved (urbanai.fr), the research pioneered by Beroche represented a preliminary exploration into the role of AI in helping to build sustainable cities and in affecting urban identities. From this report, a definition of "urban artificial intelligence" arose.

The term "urban artificial intelligence" refers to any system that incorporates data derived from the urban environment, which is then processed by algorithms, the result of which has useful applications in the socio-spatial nexus of the city. Three important elements distinguish urban artificial intelligence from other forms of artificial intelligence, necessitating its own theoretical treatment.

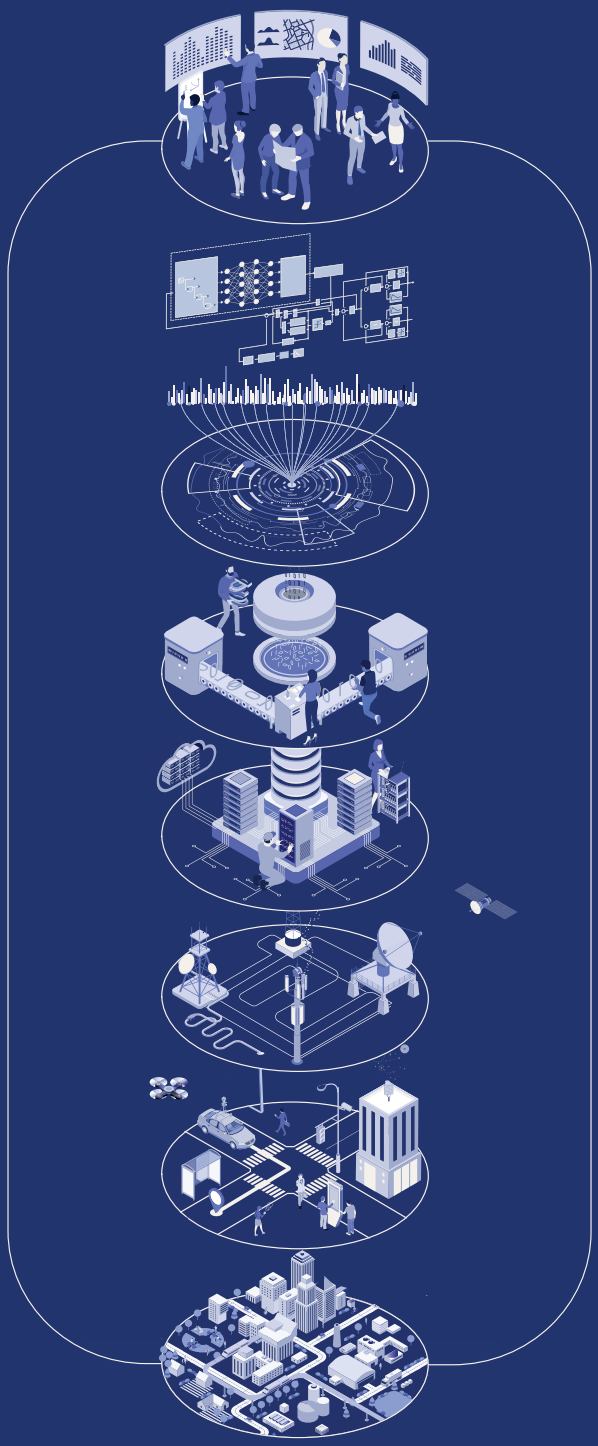
First, the complexity of the city. Due to their multifaceted nature, cities have been described as "complex systems of systems" (Batty 2016). Numerous, interwoven sectors contribute to the functioning of a metropolis, and a multiplicity of stakeholders participate in the resulting institutions. As such, urban artificial intelligence implementation necessitates a multidisciplinary approach. Additionally, when algorithms interact with the urban environment, they not only have to cope with the complexities of the city, but they themselves inherently become a complex system.

Second, the specific policy contexts, in which urban artificial intelligence operates. Since the city serves as its own political arena, in which local governments often have the authority to make decisions that affect the health, safety, and wellbeing of the millions of people that live, work, and recreate within their jurisdiction, urban artificial intelligence implementations carry a particular political potency. This manifests both in terms of public interest and participation in the inputs and outputs of urban AI systems, but also in terms of the ways in which governments can utilize urban AI to serve broader aims (Chubinidze et al., in press).

Third, the hybridity of urban artificial intelligence. Whereas other applications of artificial intelligence can exist entirely in the digital sphere, urban AI has a unique materiality and infrastructural component. Digital urban AI systems have intrinsic linkages with physical urban systems, be it through the implementation of autonomous robotics that operate in urban space or through the infrastructure of the urban sensor systems that feed data to many urban AI platforms or by virtue of the fact that the results of urban AI platforms play out in urban space, affecting residents, as well as the built and natural environment. In this guide, we elucidate a framework for understanding the components of an urban AI system, as well as the specific considerations that come into play when implementing artificial intelligence in urban contexts.

PART 1

URBAN AI ANATOMY



URBAN INFRASTRUCTURES

Urban challenges can lead to technological solutions, but it can also be the case that innovations in artificial intelligence cause entrepreneurs and data scientists to seek out urban applications for their technologies. Due to this duality, it remains crucial to critically evaluate a proposal for a technology-based solution, to ensure that such a solution serves a city's needs and population. Regardless of the problem definition pathway, the city, as a composite of built forms that shape and are shaped by complex social dynamics, provides the infrastructural and informational basis for urban AI implementations. Urban AI, then, should arise as a result of co-production: with the city and the data that it outputs fueling artificial intelligence systems, while artificial intelligence and its outcomes serve to shape and alter the city at the same time.

The city itself is composed of many layered and intertwined components, including the urban environment, flows, governance dynamics, and citizens. The urban environment, as conceived in the context of the city, refers to both the biogeophysical environment (atmosphere and ecology) and the built environment (buildings and infrastructure). Flows refers to both human flows (transportation/mobility) and metabolic flows (water, waste, energy, and logistics systems systems), which move in and among the various environments. Governance refers to the mediatory space between the environment, the flows, and the people. Often, this governance layer takes the form of planning (ensuring that the components are well-integrated with each other), maintenance/operations (ensuring that each component functions as it should), and public health and safety (ensuring that humans are not harmed the urban environments and flows with which they interact, nor other humans).

Perhaps the most important structural element of the city is the people. The city is built by and for agglomerations of people. Citizens sit at the core of urban AI. The problem definition phase begins with urban residents (melding solutions to fit the perceived or actual needs of the population). The components of Urban AI are designed, built, and overseen by people. Humans and human-built systems produce the vast amounts of data necessary to drive urban AI implementations. Indeed, the outcomes of Urban AI solutions directly or indirectly affect urban citizens in one way or another. Since people have such a core involvement in urban AI projects, they can understandably feel uneasy with, and potentially fall victim to, unintended consequences if not involved in decision-making processes. Such a pattern is true with nearly all governance and policy interventions, but public fear can be more pronounced in the context of artificial intelligence, due to the novelty of those systems, a general lack of understanding of the involved technologies and processes, and stories of negative outcomes arising from poorly designed implementations. Similarly, since urban governments are made up of people, public decision-makers and urban implementers commonly experience similar misgivings around digital integration.

To serve this end, at least from the internal government perspective,

the City of Montreal has created an AI ambassador program, in which members of the AI team interact with other departments within the city to both increase civil servants' comfort levels around AI technologies, as well as to help identify problem areas that could lend themselves well to AI solutions. The Montreal team accomplishes this largely through a combination of listening and education. By listening to the concerns of their colleagues, the AI team can better understand the reasons why other employees might face hesitancy at the prospect of utilizing artificial intelligence and other digital technology solutions. Then, following this, they can provide information to help their colleagues better understand the technological systems, as well as which concerns can easily be addressed. Similarly, by listening to and learning about the processes that their colleagues go through on a daily basis, the AI team can better identify areas for automation or enhancement, as offered by AI. At that point, the conversation shifts to cooperatively deciding how AI can fit in and what a potential project design and solution might look like. This guide seeks to provide similar information and guidance to those who read it.

Due to the complexity of cities and the many components involved, the start of any urban AI requires carefully taking stock of a number of elements. First, there should be a clearly defined urban problem, with a clear connection to social or environmental ills. The problem should lend itself to being solved by artificial intelligence (sometimes other, simpler solutions are more appropriate for the problem at hand), and the role of artificial intelligence in solving the problem should also be clearly defined. The infrastructures and flows that contribute to both the creation of the problem and the proposed solution should be identified, as part of scoping out the data elements. Relevant stakeholders (including public officials, civil servants, urban operators, technology vendors, and city residents) and their various interests should be laid out, to ensure that each party is engaged meaningfully and appropriately throughout the process (note: sometimes this engagement takes the form of participatory decision-making, but sometimes it merely takes the form of information provision in either or both directions). Only once these physical and social infrastructures have been identified and accounted for, can the anatomy of urban artificial intelligence project implementation truly begin to take form. In this regard, the first step of the urban AI anatomy (sensors and data collection infrastructures) builds off of the social and physical infrastructures identified in this foundational layer. Each of the constituent components of the city (urban environment, flows, governance, and people) has elements that constantly undergo temporal and spatial changes, giving off signals that can be captured and understood as data. These data often serve as the basis for the problem definition, the fuel for the artificial intelligence solution, and the canvas upon which the solutions are implemented and tested.

SENSORS AND DATA COLLECTION INFRASTRUCTURES

Data collection captures the knowledge that fuels Urban AI and ultimately guides urban decision making. Every second of the day, in all corners of any given city, urban elements emit signals laden with information about their current state. For example, cell phone location pings convey the number of people in a given place at a given time; the visual characteristics and roughness of road pavement cracks and inconsistencies convey the health of that infrastructure; infections and deaths in the community convey the speed and nature of the spread of disease.

The "Smart City" model of urban sensing, which frequently forms the basis for urban tech, revolves around the concept of an Internet of Things (IoT): an urban lattice of digital sensing systems that continually measure urban emissions and convert the product into usable digital information. The array of available urban sensing possibilities can be broken down into five main functional categories:

- Sensing location/mobility (these can include GPS/bluetooth positioning systems used for cell phone location services, fleet trackers, intersection counters, etc)
- Sensing the environment (these can include weather sensing instruments, Lidar/Radar, satellite/aerial imaging, water quality monitors, air quality monitors)
- Sensing consumption (these can include water/electricity/gas meters, financial record systems)
- Sensing societal health and safety (surveillance systems, censuses, opinion surveys, contact tracing applications)
- Sensing infrastructural health (these can include pavement monitoring systems, vibrometers, building monitoring systems)

However, not all urban data-use needs require the coverage and infrastructural overhead of the traditional "smart city," to achieve their aims. There exist many vendors that will run short-term surveys of relevant variables within a city, rather than deploying a permanent data collection system. For example, as an alternative to the costly practice of installing pavement sensors throughout a road network for the purpose of continuous condition sampling, vendors can conduct periodic road surveys using vehicle-mounted equipment. In the case of datasets that can be harder for cities to collect or access, such as cell phone location data, third-party data brokerage firms will purchase those data at scale, aggregate the information, and sell the packaged output to cities or other interested parties. Additionally, not all data collection relies upon digital sensors to capture information. Some data collection efforts will either take the form of, or be supplemented by, manual sampling. In these cases, people will go into the field to record information about what they see (often in a structured manner, like in a form). These surveys often require more time and a higher cost to complete when compared to digital data collection efforts, but they can also provide information not easily captured through other means. When paired with digital/

automated data collection, manually collected data samples will be compared to the information captured by the digital sensing systems, in order to validate and calculate the accuracy of those efforts. Manual data collection and low-tech and/or build-it-yourself sensors have featured prominently in a recent turn toward crowd-sourcing urban data (Gallo et al. 2014). In some cases, crowd-sourcing efforts originate from public agencies, as an effort to engage citizens and fill gaps left by conventional sensing equipment, providing a more localized insight. For example, the Boston Mayor's Office launched an application called "Street Bump," which collected data on road conditions while Bostonians drove around the city (City of Boston 2019). In other cases, crowd-sourcing efforts originate from citizen groups, as a means of reclaiming citizen rights and centering human interests in the face of data-driven decision making. As an example of this more bottom-up approach, the DEcentralised Citizen-owned Data Ecosystems (DECODE) project in the EU invited residents of Amsterdam and Barcelona to share their personal data to an open data portal, both giving individuals the ability to choose what information they want to share (if any) and allowing NGOs, local communities, and nonprofits access to data that are normally restricted to the wealthy tech firms that can afford to purchase it (DECODE 2023).

When considering crowd-sourced and citizen-centered data collection, qualitative data (such as social media posts and opinion surveys) play a major role in giving people a voice in the operations of their city. Although much of the focus on data collection surrounds quantitative data, machine learning and AI techniques also provide the ability to derive quantitative data from qualitative data (like written text) at scale. Natural language processing and natural language understanding methods can turn text and speech into data, too. These tools can derive categories of meaning and detect speaker/writer sentiment from text-based matter. In these cases, and in cases where urban AI projects use data sets collected for other purposes external to the city, the data obtention considerations revolve less around sensing, but rather around building data sharing and data governance frameworks, with an emphasis on assessing the data sources currently available within the city, as well as the owners, formats, ages, and update frequencies of those datasets.

Even if an Urban AI project does not inherently involve a data collection component (i.e. if the project utilizes pre-existing datasets), understanding the considerations associated with data collection allow for a more clear understanding of how to read and interpret the dataset being used. A number of factors can drive data collection implementation considerations. Data collection can also be continuous or on demand. With continuous data collection, data points are sampled (collected) at a regular frequency, without human intervention. Continuous data collection is often used to build a time history of an observed phenomenon. For example, continuous data collection is often employed in environmental sensing, when day-over-day data provides more information than a

single, point-in-time data collection effort. When performing some types of continuous data collection, one must decide on an appropriate collection rate by considering how often the phenomenon is likely to change in a meaningful way. If capturing the location of people as they move about the city, one would use a higher sampling rate for vehicles than pedestrians, as vehicles often displace their position by a much greater distance than pedestrians would in the same amount of time, so it requires more frequently sampled data points to maintain the same degree of spatial information for both modes. Similarly, if a project aims to assess a long term process, such as land use change over a decadal timeframe, a hyper-frequent sampling rate would not make sense in that context.We could also notice that the introduction of digital services in government operations and citizen interaction (e.g. online 311 systems, digital utility billing systems, online point-of-sale platforms for parks and recreation services) allow for data to be captured about and during a city's daily operations.

Maxime Dusart from UPCITI, a company that specializes in privacy-by-design multi-use sensors, shared insights on how his team accounts for multiple considerations of sensor implementation in urban environments: implementation scoping, installation design, device health and maintenance, and data privacy.

In scoping a sensor implementation, he recommends scaling the number of sensors to fit a city's size and needs. For cities that have use cases for urban sensing but are hesitant to invest large sums of money or engage in lengthy bidding processes, Maxime Dusart suggests device rental can be offered as an all-in-one package with platform subscriptions, in order to lower a city's barrier to entry. This practice highlights one of the main conclusions of this report : urban AI does not belong solely to large metropolises - *with more flexible scoping options and the cost reductions that come from hardware and software innovations, small- and medium-sized cities can also benefit from the insights captured through sensing systems.*

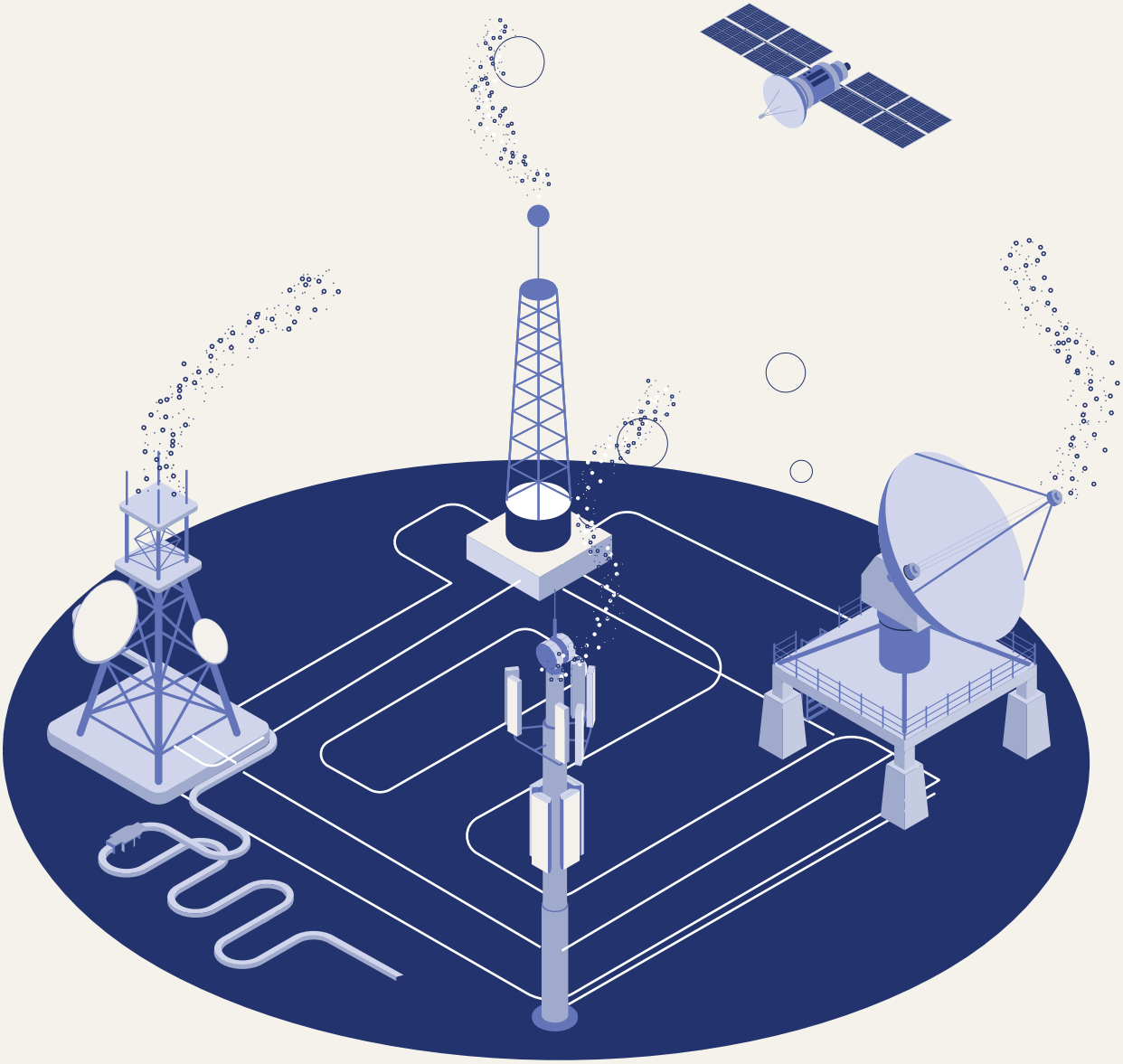
Although the type of data being collected determines the installation location to some degree, Maxime Dusart shared some general things to keep in mind when planning to mount sensors in urban environments. Depending on whether the device is self-powered (as with a battery or other power sources like solar panels) or requires external power, the mounting location may need to coincide with an available power source. In the case of UPCITI sensors, light poles serve as the ideal mounting environment, since the sensors rely on height to adequately collect their intended data points and since the sensors can tap into the light pole's electricity in order to power the devices without the need for humans to change a battery. As an added bonus, mounting the sensors high on the light pole mast protects the devices from both vandalism and accidental disruption. Dusart does note that utility companies do have particular

regulations related to light poles and often require that their own teams perform the installation.

As with any electronics, sensor devices require some degree of maintenance and upkeep to ensure their proper functioning. Such efforts can have significant time and monetary costs if a deployment contains many devices, the sensors are deployed over a large area, or the sensors are mounted in less-accessible locations. On top of being multi-use (offering five use cases simultaneously), and In order to reduce the need for frequent manual interventions, the UPCITI sensors, as well as similar data collection instruments, have the ability to sense aspects of their own system and hardware health, which can be used to prioritize maintenance efforts and proactively respond before a sensor fails.

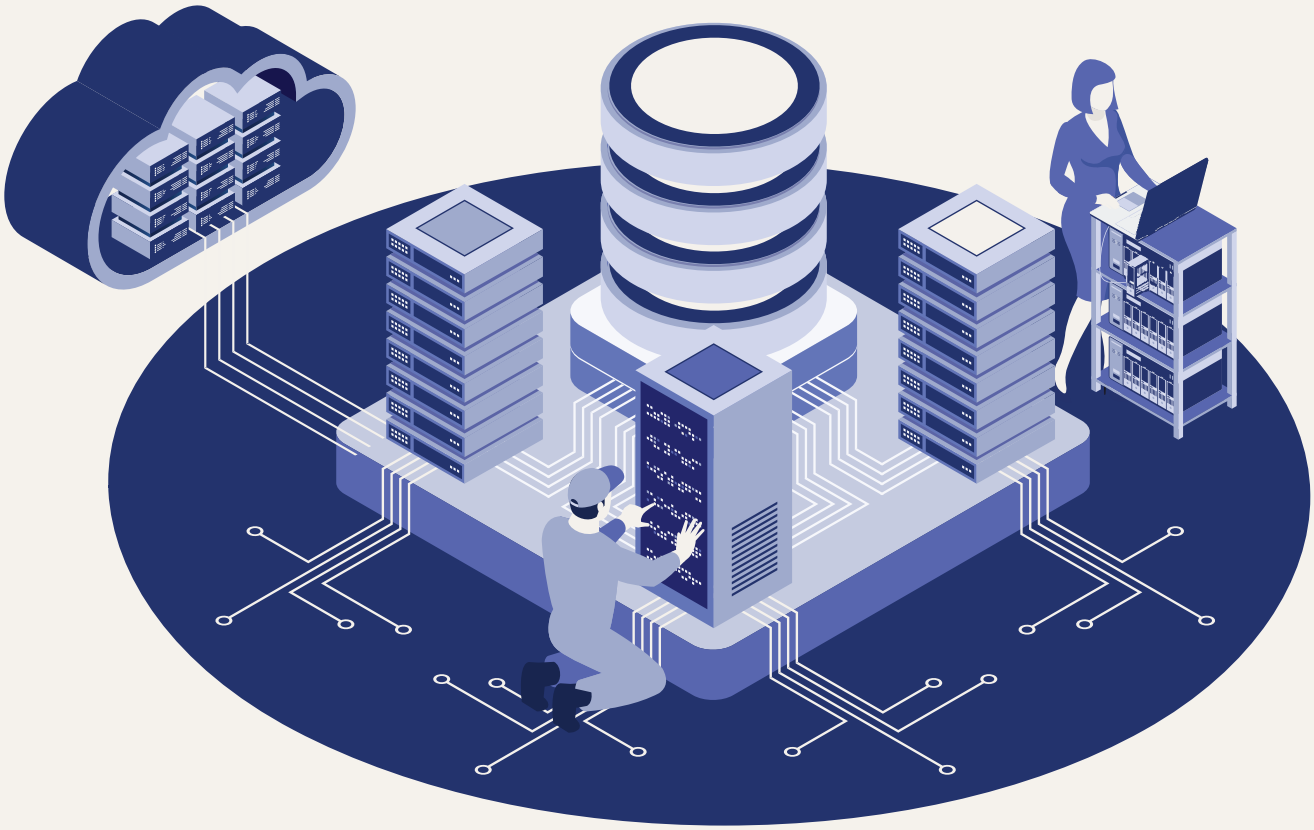
Finally, depending on the nature of the sensor and the type of information being collected, sensor systems have the potential to invade the privacy of passersby. Data privacy laws, such as the EU Global Data Protection Regulation, often place limits on the type of information that can be collected, as well as how it can be stored and processed. In light of these regulations, and in order to protect the public's privacy, Dusart recommends collecting data in an already anonymized fashion, rather than stripping it of identifiable features after the fact. To do so, UPCITI uses a very low-resolution lens for video recording, which allows for the identification of a recorded object as a person, for example, but with a low enough quality that the object cannot be traced back to a specific person. Similarly, in capturing sound, UPCITI uses recording techniques that can measure sound levels without capturing actual, spoken words. As an additional safeguard, UPCITI performs all of its data processing within its sensor devices (rather than storing image or video recordings), so that humans are only ever interacting with the processed data and not raw footage. In order to ensure that citizens feel comfortable with and aware of the data collected by the sensors, UPCITI recently included QR codes on every sensor-mounted light post in various cities like Angers (one of the largest EU smart city projects), which directed the public to an information page about the project with key data privacy information and examples of the insights obtained. Similarly, an exhibition in France displayed examples of various urban sensors and encouraged participants to interact with those technologies to better understand how their data are captured in urban spaces (Consortium DATA PUBLICA - KPMG 2021). In encouraging public awareness of and participation in data collection, cities can design as constructive interfaces, rather than invisible surveillance (Beroche 2022; Schmoeker and Lu 2021).

NETWORK INFRASTRUCTURES



Once data have been collected, particularly in the case of sensors, network infrastructure can be used to transmit the data from the collection device to other locations, such as data stores. Types of network infrastructure include, but are not limited to: radio frequency networks (e.g. 5G, 4G, Bluetooth, LPWAN, etc.), fiber-optic cables, and broadband. Cybersecurity and data protection become pertinent issues when considering the design and implementation of network infrastructures. From a physical security standpoint, extreme weather events pose a risk to the integrity of network infrastructure, as well as the effectiveness of wireless connectivity. These issues become especially pertinent if a city relies too heavily on networked systems for critical decision making. From an information security standpoint, transmitting data carries the risk of malicious interception between point A and point B. In order to safeguard against this risk, a process called encryption is often used to conceal the contents of the data being transmitted (requiring a special code to unscramble it for example), so that they will be incomprehensible and unusable if intercepted. Additionally, network infrastructure design must consider interoperability. Some data transmission systems allow the data to be ingested by multiple other systems, while others remain specific to a single use with a particular origin and destination. Open data standards can promote data sharing and transparency, allowing the public to more easily interact with urban datasets, while propriety or privacy considerations might cause a company or public entity to opt to lock down their data transmissions.

DATA STORAGE INFRASTRUCTURES



Urban sensors and other data collection efforts can produce a vast amount of data, which need to be stored, in order to facilitate their use at a later time. There exist many forms of data storage infrastructure, depending on the volume, format, complexity, and sensitivity of the data. Data storage infrastructure requires "hosting" (I.e. selecting the physical space where the data storage infrastructure will be located). Generally, two options will be available for hosting: on premises and in the cloud. With on premises hosting, the servers that house the data storage infrastructure exist on site in the city and the city maintains sole control, security, and maintenance of the systems. With cloud-based hosting, hundreds of servers in multiple different locations house the data storage infrastructure. In that case, a provider will manage the maintenance and security of the systems, and grant the city some control.

Data storage can range in capacity and complexity. Beyond simple file storage, data can be stored in a single table, at the smallest and most basic level. An Excel spreadsheet is a common example of this. Tables require the least amount of technical knowledge to view and manipulate, but also have limitations to the amount of data that one can store within them, as well as less ability to structure and ensure the integrity of the data stored within them. Databases (which largely take the form of relational or non-relational databases), allow for multiple tables to be stored together. In a relational database, data exist in multiple thematic tables, which have rules specifying how they fit together based on shared characteristics. Relational databases allow for programmers to specify rules, which dictate precisely what data can be stored and how, preventing undesirable data from entering the database. They also can be arranged in a particularly lean manner, referred to as "normalized," which streamlines data updates by allowing an information change in one table to be reflected in all other tables, by virtue of their relationships. In a non-relational database, the data exist all in one bucket. Whereas relational data follow specific rules, non-relational databases offer more flexibility in terms of the structure of the data being stored, as well as the amount of data being stored. Often, non-relational databases will be used for real-time data, as their lack of structure allows for quicker data insertion, while relational databases will be used for other forms of data. In the same way that a database is a collection of tables, a data warehouse is a collection of databases. The main purpose of a data warehouse is to integrate data from multiple different sources. For example, if an urban AI project requires data from multiple sensor systems as well as non-sensed data, a data warehouse can be used to bring the various data stores together into one storage infrastructure. Data warehouses still maintain rules and relationships between the various databases, similar to those between tables within a relational database. Finally, data lakes represent the largest and most complex form of data storage. Data lakes have the ability to store vast amounts of data from multiple different data sources, regardless of how the data are structured. With the proliferation of IoT networks and the participation of individual devices within those systems, researchers and technologists have also

been exploring opposite approaches - favoring data decentralization over agglomeration. In one such project, deemed "Federated Learning," researchers at Missouri S&T in the United States are improving methods of running machine learning models using data from IoT devices, while keeping the data on those devices, in an effort to increase data security and privacy (Ehrhard 2020). While these technologies have not yet been fully developed, it's quite possible that urban data storage could follow that trend in the future.

Recent innovations, like blockchain technology, have introduced a new aspect to data storage, in which data are stored in a distributed, decentralized manner (meaning that no one person, organization, or collection of servers maintains the control over and responsibility for stored data). Additionally, cryptographic verification techniques allow unique codes to be assigned to datasets each time they are added to the blockchain (or similar database system), allowing users to periodically verify that the current code matches the original code for each element. Matching codes means that the given record contains the same values as the original, ensuring that no data have been changed or deleted. In this way, blockchains and similar technologies create an infallible record of each observation in the database, ensuring that each committed entry remains in its original state in perpetuity. For this reason, many processes that require precise knowledge of an item's history (like logistics tracking) or determinate information of financial transactions (like with cryptocurrencies) utilize blockchain technology to store their data in an ironclad manner.

Beyond selecting the format in which data will be stored, a number of considerations inform data storage. When storing data, one must decide how long to store the data. Sometimes, data retention policies or data use regulations dictate the duration of storage for a given dataset, based on the type of information being stored. Additionally, keeping adequate metadata and documentation is crucial for ensuring that others will be able to locate data within the system that you implement. Often, performing the actual data access itself will require the use of a query language. Querying allows for the retrieval of only the relevant data, without accessing the entire table, database, or data lake (unless necessary). Structured databases use Structured Query Language (SQL) to access data, while unstructured databases commonly use Mongo.

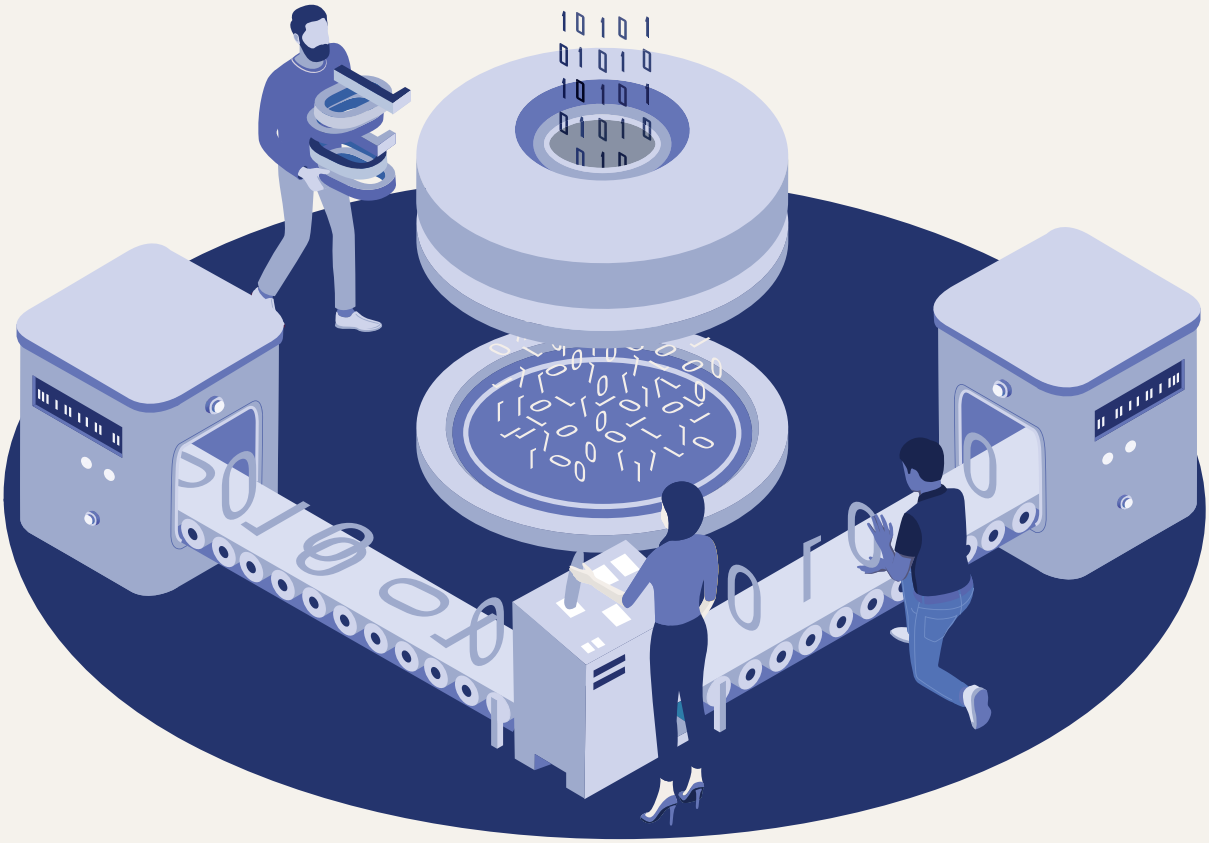
Additionally, data can be stored raw or aggregated. Raw data offers greater analytical possibilities, but can require a lot more storage space than aggregated data. According to Katya Letunovsky, Vice President and Co-founder of Habidatum, which utilizes AI to inform real estate location intelligence, using a database that allows you to slice and dice data at various levels of aggregation offers the greatest possibility for flexible analysis. She mentions that databases can be designed to facilitate data storage and retrieval along different axes (e.g. temporal databases which allow data to be stored in relation to a time element and/

or spatial databases which allow data to be stored in relation to latitude and longitude). Additionally, she mentions the importance of ensuring that your data storage schema allows for data to be harmonized across multiple different sources, in the event that a project requires the use of multiple data sources. In other words, the schema should support a data organization that contains comparable aspects in each data source and which outlines the relationship between them.

Aside from technical considerations, data governance decisions also have an effect on decisions and outcomes related to data storage. Policies around which data to store, who can access them, and how they can be accessed drive the selection of data storage infrastructure, as well as the design of data management frameworks. Storing data in an open manner (i.e. making them publicly accessible) allows for greater transparency and a higher degree of citizen participation. On the other hand, since much urban data contains signals from humans, ensuring data privacy by only storing necessary data and limiting access to sensitive data to only those users who require access builds public trust and represents ethical handling of citizen information (Saran 2021).

DATA PROCESSING

(CONSOLIDATION, ORGANIZATION, AND QUALITY ASSURANCE)



An old and commonly held adage in data science states "garbage in, garbage out." Under this principle, any product of analyses, data visualizations, and artificial intelligence models holds little value unless the underlying data themselves are of a high quality. Since urban data come from a variety of sources, they often vary in cohesiveness, which can affect their composite accuracy. Inherently, urban data tend to mimic the messiness of the hectic, real-world urban conditions, which generate them. Data processing techniques and principles allow data practitioners to ensure the cleanliness and integrity of their data, prior to implementing a project, in turn ensuring the value of the insights produced through urban AI and other output analyses or visualizations.

Data processing encompasses a wide array of tools and techniques. Broadly, data processing can take the form of data cleansing, aggregation, de-identification, extrapolation, and merging datasets from different sources. Data cleansing focuses on identifying and correcting errors or anomalies within a dataset. Often, datasets may contain errors or missing data points that result from the collection. For example, text-based data may contain typos, sensor-collected data may have dropout values where the sensor didn't make a proper reading, and location-based data might have records with varying degrees of accuracy depending on the cellular network. Data cleansing processes use algorithms to identify values that don't fit an expected pattern and update the values according to a set of rules specified by the data engineer. . Sometimes problematic data points will merely be deleted from the dataset, but sometimes interpolation techniques will take nearby values into account in order to estimate a more plausible value for the dubious one.

Similar to interpolation, which allows for missing data points to be filled in between two or more known data points, extrapolation allows for the creation of missing data points outside of the collection boundaries. Extrapolation uses algorithms to identify patterns within the collected data and extend the pattern beyond the known information. Frequently extrapolation is used in temporal analysis, to predict future values, based on past trends.

Aggregation focuses on grouping individual observations and reducing multiple values into a single representative value. Various aggregation measures are used to perform the reduction, including sum (adding all the values to get a total), mean (averaging all the values), median (finding the middle value), mode (finding the most occurring value), minimum (finding the lowest value), and maximum (finding the highest value). Data aggregation will happen along temporal, spatial, or categorical lines. Temporal aggregation often occurs when data are collected at a greater frequency than necessary for the analysis. For example, point-in-time transaction data might be summed by day, in order to analyze a daily total instead of each individual transaction. Spatial aggregation often occurs when data are collected as individual point locations, but the analysis requires information over a given space. For example, it

might be useful to understand the dynamics of a variable at the specific collection sites, but depending on the use case, having values at the neighborhood or city level might better facilitate planning. In those instances, geometric calculation techniques would be used to identify which data points exist in the given area, and an aggregation measure would be applied to the corresponding values. In the case of categorical aggregation, aggregation measures will be applied according to a shared attribute characteristic (e.g. individuals might be grouped by age group, gender, occupation, etc).

Commonly, aggregation is used as a de-identification technique, in order to protect the privacy of individual people. In these cases, it becomes much harder to make out a specific person from the aggregated data. Other de-identification techniques include encrypting personal identifiers (i.e. applying algorithms that obscure the true values, making it harder to trace them back to a given person), randomizing identification elements (e.g. applying algorithms that might recalculate a location point within a given radius of where the data point was actually collected, making it harder to identify the true location in question), and separating personal information from the other data points being used in the analysis.

Often, merging multiple datasets will be necessary in order to gain a complete picture of a phenomenon. Faircloth, Connock, Welch, Elsworth, and Escott (2022) refer to the concept of a "data-mosaic," in which "designers" (the concept of a designer can be extended to urban practitioners) meld multidisciplinary data from fields such as "urban ecology, environmental management, and public health," in order to have a more holistic understanding of urban environments and dynamics. However, this practice of combining data from different sources requires additional data processing to ensure that the data will be compatible. Commonly, different data sources will handle date or number formatting, geographic coordinates, or data structuring differently. In those cases, data processing pipelines will recalculate values and restructure rows and columns as needed, to ensure a shared format between the different datasets, which allows them to be merged correctly. Even if a project only uses data from a singular source, data might require restructuring or reformatting to mimic the input requirements of a given machine learning model.

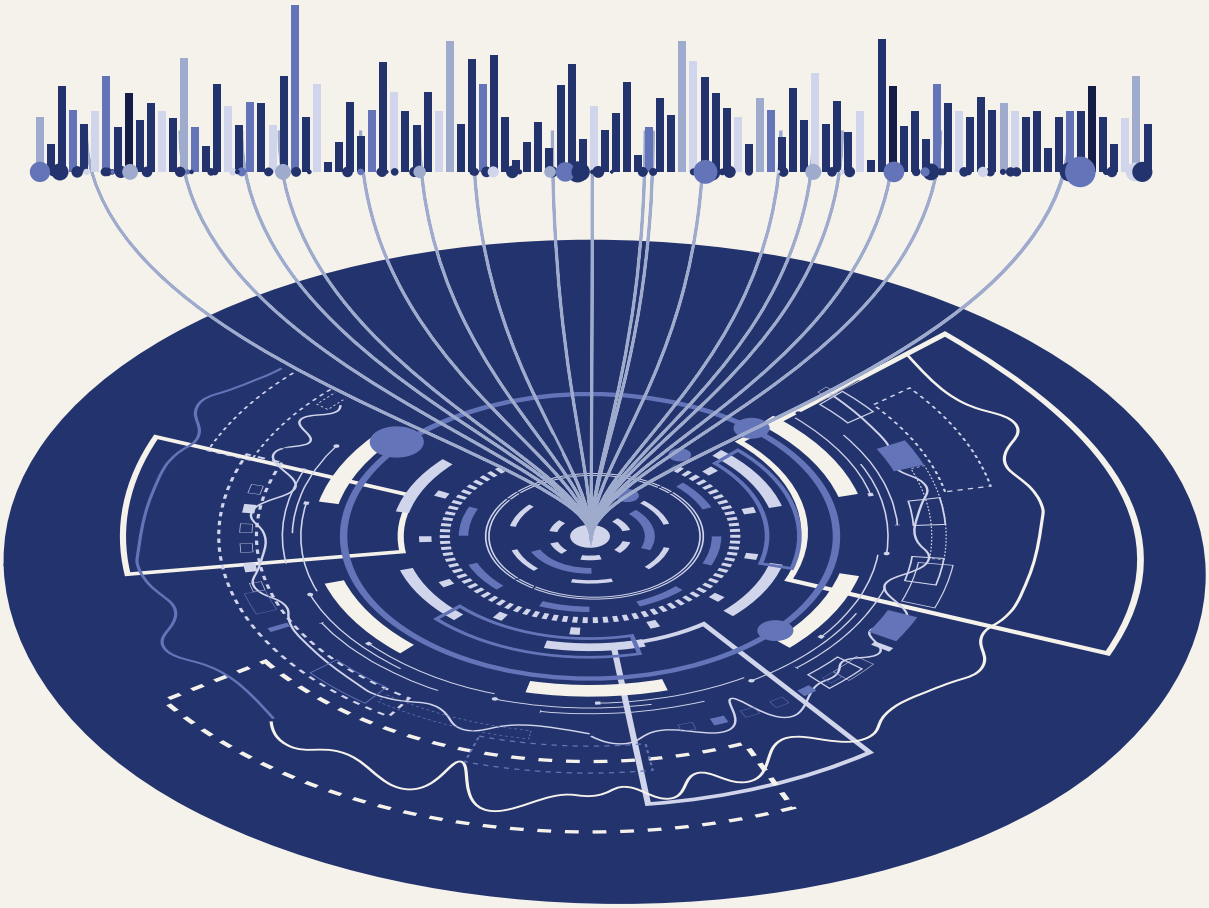
Data processing can occur in a manual or automated manner. Manual data processing refers to a person performing the data alterations, while automated data processing utilizes a set of instructions, calculations, and algorithms to complete the data processing tasks with little to no ongoing human intervention. In some cases, as with poor data quality or extreme variations that make it difficult for computers to identify a pattern and respond accordingly, manual processing might be the only way to achieve a desired level of accuracy among the data. However, manual data processing can incur massive time and monetary costs, which can quickly become prohibitive with large or frequently collected

datasets. On the other hand, as long as the appropriate algorithms and conditions are applied, automated data processing can occur almost instantaneously and more consistently. Depending on the nature of the automated processing, there may be costs associated with the initial setup and ongoing maintenance of the processes, and potentially costs associated with the computing resources needed (especially if the automated processing tasks run in the cloud).

New technologies continue to improve the accuracy and capabilities of automated data processing. According to Robert Heinecke, CEO of Breeze Technologies, which specializes in air quality monitoring systems, artificial intelligence has revolutionized the field of environmental sensing, particularly with regard to enhancing data collection and processing capabilities. Whereas older sensor technologies required costly and lengthy manual data retrieval and processing procedures, contemporary devices use algorithms to automate a lot of that work. With such innovations, authoritative, ready-to-use sensor data can be available in near-real time, compared to previous months-long delays between data capture and a viable data product.

The accuracy of sensor readings is sensitive to both variations in the environment immediately surrounding the instrument and slight anomalies inherent to the instrument itself. Any raw data feed taken from a sensor will contain a variable amount of extraneous signals and outlier data points that need to be identified and corrected for in order to obtain an accurate picture of the target variable in the absence of recording artifacts. Breeze developed algorithms to accomplish two main goals: anomaly detection (i.e. identifying data points that have a high likelihood of occurring as a result sensing mishaps) and smoothing/ interpolation (i.e. generating plausible values to fill the gaps held by anomalous or missing data points). Their data cleanup methods utilize contextual data about the sensors and immediate vicinity to consider the sensed values in context, performing cleansing specific to the characteristics of each device. In order to accomplish such specified automation, Heinecke's team consisted of environmental scientists, who could lend subject matter expertise about the specificities of air quality data, and data scientists/ machine learning experts, who could build the models necessary to rapidly and confidently process the data. Integrating both sets of knowledge - a familiarity with the data content as well as a familiarity with data structures and techniques - allows for more meaningful data processing outputs.

DATA VISUALIZATION



Even in the absence of artificial intelligence and machine learning outputs, data visualization turns urban data into meaningful information. When coupled with AI/ML or other data analyses, data visualization becomes an even more powerful tool that allows users to understand the meaning of the results, identify areas for action, and communicate about the decisions that will result from the project. Data visualization can be used to highlight patterns and anomalies, to make sense of the past, and to give an impression of what might come in the future. Often, data visualizations will consist of one or more graphs, charts, or maps. These elements can be designed to convey information individually, or they can be combined in a way that allows them to collectively tell a story (often referred to as a dashboard).

Data visualizations can be static or interactive, depending on the intended use and how they will be distributed to users/viewers. With static data visualizations, the graphs, charts, and maps will show a specific snapshot of the data, as designed by the person who created the data visualization. Static visualizations allow the information to be communicated across a variety of different platforms: online, as a still image or pdf, or on a printed sheet of paper. They also allow designers to highlight, specifically, the patterns or trends that they want to show. This can be quite useful for concise communication efforts, or to display results in situations where resource limitations prevent the implementation of interactivity. With interactive data visualizations, on the other hand, the graphs and charts update according to user input (e.g. if the user wants to highlight a different variable, assess a custom timeframe, or zoom to a specific location on a map) or in response to changes in other graphs. This gives a lot more flexibility for users to draw their own conclusions and interpretations, and it can have the effect of engaging users far more than a static layout. One notable platform, CityScope, takes interactivity into three dimensions, allowing stakeholders to physically change elements of a modeled city, in order to observe how outcomes to various political or built interventions might play out. By incorporating physicality into data visualization, citizens take more ownership with regard to providing input processes (Popelka 2022). Across the board, interactive visualizations can prove particularly useful at the start of the project, as they allow greater freedom to explore the data, and also during the AI decision-making phase, because they allow for testing out various policy changes or viewing the results of scenario modeling. However, interactive visualizations require more infrastructural overhead, such as specialized software or web development tools in order to build them and web hosting infrastructure in order to share them in an interactive manner. Additionally, depending on their complexity. interactive data visualizations might require additional training in order for lay people to understand how to navigate them.

Regardless of whether data visualizations are static or interactive, a good visualization should be able to convey key information about the dataset, analysis, and outcomes in a concise and not too technical manner. In this

regard, striking a balance between descriptiveness and simplicity is key. If a chart is overloaded with information and visual cues, the takeaways will be lost and a viewer might get confused. If a chart has too little information, then it, too, will not effectively communicate the results, as it might lack necessary context. When it comes to walking this line, careful usage of color, symbol sizing, and labeling can have a big impact. Changing the color or increasing the size of an element will draw attention to it and highlight its importance. Labeling key aspects and including a legend will clarify which variables are being shown and what is meant to be conveyed through certain colors and sizing differences. Certain design conventions (like using cool or lighter colors for lower values or "better" values and using warm or darker colors for higher values or "better" values; the same goes for using larger sizes for larger values and smaller sizes for smaller values) convey default meanings. Deviations from these conventions often warrant explanations to emphasize the new meaning being attributed to standard representations.

In many data visualizations, information is grouped based on categories, in order to translate a collection of data points into a cleaner representation. Sometimes, as with categorical data, the groups will be predetermined. Geospatial or temporal data require the designer to decide on a level of specificity to display. For example, temporal data collected at one-second intervals can be shown by second, minute, hour, day, week, etc. with each level of specificity conveying a different meaning. Similarly, geospatial data collected at singular points can be shown at the block level, neighborhood level, city level, etc. with each also conveying different meanings. Numeric data can use different cutoffs in order to be put into groups, with some cutoff calculation methods taking into account the distribution of the data and others using a designer's prior knowledge about the data. Given the broad array of choices that a designer can make, it becomes crucial for a designer to understand the data and analyses that underlie the visualizations, in order to make the appropriate choices about what aspects to highlight and how to represent the information.

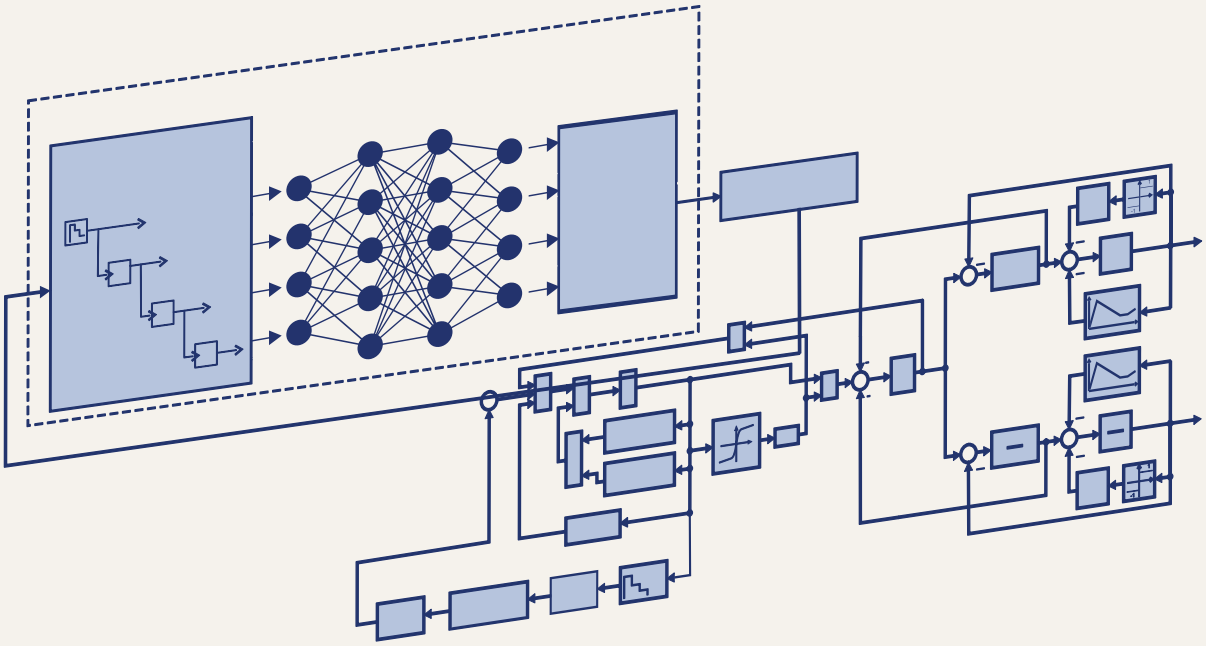
Caroline Goulard, the CEO and Co-Founder of Dataveyes, which enables enhanced human-data interaction through custom data visualization tools, champions the role of data as an important and powerful communication tool. The process of preparing well-designed graphs and charts makes information visible - unlocking new insights for decision makers and allowing governments to inform and involve their citizens in meaningful ways. In order to make effective use of data visualization tools, Goulard highlights the need to consider the specificities of the data being ingested and the questions being asked. She starts every project by engaging with the client to understand the kind of information they seek to learn from their data and how they intend to manipulate that information. Having a clear understanding of the problem allows her to help design the appropriate ways of representing the information, designing graphs and charts that enhance the meaning of the data

rather than manipulating data to fit visualizations. To this end, her team uses tools like Javascript, React, Webgl, D3, and Node.js to create custom refined, context specific, interactive graphs and charts, but she also makes it clear that more generalized, commercial software can still provide the same level of depth and aesthetic across generalized use cases.

Goulard seeks to incorporate interactivity into her visualizations as much as possible, as she thinks that the most meaningful results and decisions come from hands-on approaches to understanding the dynamics and implications of the data collected. A good data visualization, in her mind, should allow users to understand the situation, test ideas, and see the impact of those ideas on the data. From the perspective of understanding the situation, visualization can highlight phenomena and trends that non-graphical statistical methods might not reveal. Very often, when working with extremely granular data, visualization can allow aggregate, overarching patterns to emerge beyond the insights afforded by individual data points. For example, a table containing information about relationships between members of a community might not provide much information about the dynamics of those relationships, but mapping out a network diagram immediately uncovers the varying degrees of connectedness and centrality of each individual.

Beyond aiding in decision making, Goulard also highlights the role of data visualization as a critical citizen communication tool. Inhabitants have an intimate involvement with any AI implementation, either as the impetus for the project, the producers of the data that feed the system, and/or the recipients of the outcomes. For this reason, Goulard argues that citizens must be taken into account in the elaboration of algorithms and in the prioritization of decision making efforts. Presenting critical information about their AI implementations through data visualization allows cities to be transparent with inhabitants about the process and results, in a more easily understandable manner. Further, building that kind of sharing through data visualization allows for the creation of a relationship between the city and its citizens, built on mutual understanding and shared feedback.

ALGORITHMS AND MACHINE LEARNING



As highlighted throughout this guide, algorithms and machine learning can enhance multiple aspects of the Urban AI project implementation process. However, in this section, we explore artificial intelligence as the meat of the project (i.e. as the main implementation that all the preceding steps support). Artificial intelligence, used in this way, itself becomes the basis for urban decision-making and intervention, allowing for automated alerts, future predictions, scenario modeling, and the identification of non-visible patterns. These capabilities either inform governance and public service delivery, automate aspects of the urban environment, or promote sustainability (Castillo, Karaiskou, & Gandhi. 2022).

Burgess (2018) defines a typology of artificial intelligence algorithms consisting of buckets of tools related to "capturing information," "working out what is happening," and "understanding what something is happening." The "capturing information" category consists of algorithms oriented toward applying meaning to otherwise less-meaningful data. This can take the form of algorithms that perform image recognition (as is the case for urban AI implementations that identify behaviors or people from security camera footage, to automate urban security monitoring), speech recognition (as is the case for automated customer service agents to route citizen calls appropriately), search (as might be the case to aid in document management), and clustering (as is the case for public health disease tracking). The "working out what is happening" category consists of drawing conclusions from available data. This can take the form of natural language understanding (as is the case for automation-aided permit application review), optimization (as is the case for prioritizing municipal budget allocations), and prediction (as is the case for estimating infrastructure deterioration). The "understanding why something is happening" category refers to a theoretical, not-yet-available capability of AI systems to choose the correct operational approach according to contextual information (understanding cues), without following the specific algorithms that comprise the other categories.

When incorporating AI into a project, there are three main options: off-the-shelf AI software, AI platforms, and building bespoke applications (Burgess 2018). Off-the shelf AI software is the easiest to implement, but the least flexible. With off-the-shelf AI software, a vendor handles the technical work associated with the AI models; the user is only responsible for the data preparation aspects. However, since the models are ready-made, they might not align perfectly with the intended use. When using an off-the-shelf product, it is important to ensure that the AI offering is vetted for efficacy and that it is an appropriate tool for the use case. Off the shelf options will usually be priced as a subscription to the service (allowing unlimited use during the designated time frame) or charged per use of the service. If considering an AI solution oriented toward optimizing trash pick-up routes, for example, an off-the-shelf AI software would allow the city to upload its geospatial data and adjust a handful

of configuration settings and receive routes as an output, without ever interacting with technology beyond the software interface. AI Platforms offer a middle ground between flexibility and ease of implementation. AI platforms provide access to AI tools and algorithms, but require the user to put the building blocks together in a way that will work for their project. As such, using an API platform requires some time and technical knowledge beyond an off-the-shelf solution, but it also allows for greater customization of the AI implementation. In this case, pricing will commonly be per run of the model (ibid.). In the trash route optimization example, the city might build out a system that performs certain data processing and manipulation tasks, but relies on a route optimization API to perform the heavy calculations. Building bespoke applications provides the greatest flexibility, but also the most overhead in terms of time and money. Bespoke applications involve building and training artificial intelligence models from the ground up. As such, it's possible to ensure that the model is perfectly tailored to the use case and that all components of the algorithms are known (as opposed to the commonly referred to "black box" nature of artificial intelligence algorithms), but it requires a highly skilled technical team and/or an AI consultant. In this case, pricing will be based on staff time and infrastructure costs. A bespoke trash route optimization application would use a custom built route optimization algorithm, tailored to the specificities of the city and its data, in addition to other custom-built data processing and manipulation tasks.

A number of technical considerations drive the selection of which algorithms to use and in what combination, as well as how to properly train the selected models. Hiring an AI expert to advise on a project can help steer the AI implementation in the right direction, especially if the implementing team lacks the technical capacity to evaluate the options in-house. An AI expert can also be helpful to assess the ethical aspects of urban artificial intelligence, as the algorithm-based decision making power of AI implementations can have unintended social consequences if externalities are not properly understood and corrected for. Since many models use historical data, they can perpetuate institutionalized biases. In the United Kingdom, this unfortunate reality has led nearly 50% of governmental data experts to estimate that their datasets are subject to bias (Scammell 2020). To safeguard against such pitfalls, Fjeld, Achten, Hilligoss, Nagy, and Srikumar, M. (2020) outline eight key themes for principled artificial intelligence implementations: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and the promotion of human values. Privacy refers primarily to limiting and/or safeguarding any personal information that might be collected or processed as part of the implementation. Accountability refers to ensuring that the models undergo initial and continuous assessments and that the public has recourse in the case of model-induced harm. To serve this end, some jurisdictions have implemented an AI register, which lists all elements of decision-making that utilize algorithms, as

well as the variables that those models take into account. Safety and security refers to ensuring that the models and outputs do not cause harm. Transparency and explainability refers to the act of disclosing to the public when AI is being used in services that they interact with and how the decisions are being made by the algorithms. Fairness and non-discrimination refers to ensuring that models are adequately trained, such that they do not favor or harm one social group over another. Human control of technology refers to both maintaining human operator oversight on model operations/decisions and allowing individuals to opt-out of AI-based decision making. Professional responsibility refers to holding the model design to high standards (e.g. ensuring high accuracy and involving sufficient stakeholders in the design). Finally, the promotion of human values refers to ensuring that artificial intelligence is used for social good, in a manner that benefits humanity. Taking these principles into consideration throughout the implementation process can help ensure a more successful and responsible AI project.

DECISION MAKING AND ADAPTATION

Decision-making and adaptation is often the main reason why governments and private companies opt to implement AI solutions. The decision-making and adaptation step is crucial, in that it allows urban stakeholders to operationalize the results of the preceding processes, thus affecting the environment, flows, governance, and/or people with the intention of solving the originally stated problem. While this step appears at the end in this anatomy, after the artificial intelligence step, it can also occur directly after the data visualization step. In fact, those three really represent a constant continuum: the data visualization step unlocks insights, the AI and machine learning step allows for conclusions or predictions to be drawn from those insights, and the decision-making and adaptation step allows for real world change to arise, which in turn alters the data collected. This leads to new visualizations and insights, which can generate different predictions, and in turn inspire alternative decisions. For this reason, decision-making and adaptation ought not exist as a one-off step, but rather as part of a repetitive process of continuous evaluation. Having a strong evaluation framework and adaptation strategy helps protect both people and resources within a city.

In order to adequately encourage adaptations in response to the decision-making processes, city leaders must take time to thoughtfully lay out their deployment processes. In this context, deploying an urban AI project equates to translating model results from insights to on-the-ground change. If deployed hastily or without care, AI projects have a much higher likelihood of failing due to lack of adoption. As part of a sound deployment strategy, urban stakeholders should ensure to thoroughly document their inputs, outputs, and algorithmic processes for the sake of transparency and interoperability; they should encourage adoption by demonstrating the value of the project/tool (however the users define value). Having a strong deployment and change management strategy helps to ensure better adoption of a project by decision-makers, target groups, and citizens. Especially if the urban AI project is being run as a pilot, stakeholder buy-in is key to helping ensure the success of the pilot and thus moving it from pilot to fully-implemented project.

Even if a project deploys successfully, changing external conditions can cause it to fail eventually; in order to encourage behavioral or environmental adaptation, the project itself must have the capacity to adapt. **Building in feedback loops and other continuous evaluation methods can allow decision-makers to keep a pulse on the outcomes of a project and make adjustments as necessary.** Some helpful continuous evaluation techniques include: defining performance metrics (i.e. quantitative values that measure a project's progress toward target outcomes), frequently test the AI system to ensure that it continues to produce expected results, and periodically collect user feedback, updating the system as needed in response (Eckroth 2018). Continuous evaluation introduces an iterative design approach, which can aid in identifying and correcting for pitfalls as soon as they may arise, rather than when they become a

bigger issue. As discussed in the algorithms/machine learning section of this guide, the utilization of artificial intelligence can lead to biased results, depending on the training data used and the degree of human monitoring and oversight involved in the artificial intelligence design step. Evaluation processes serve to audit and refine AI models, ensuring that they operate in a fair and unbiased manner. When decision-making is taken out of human hands, even partially, maintaining human oversight on the processes and outcomes remains even more critical.

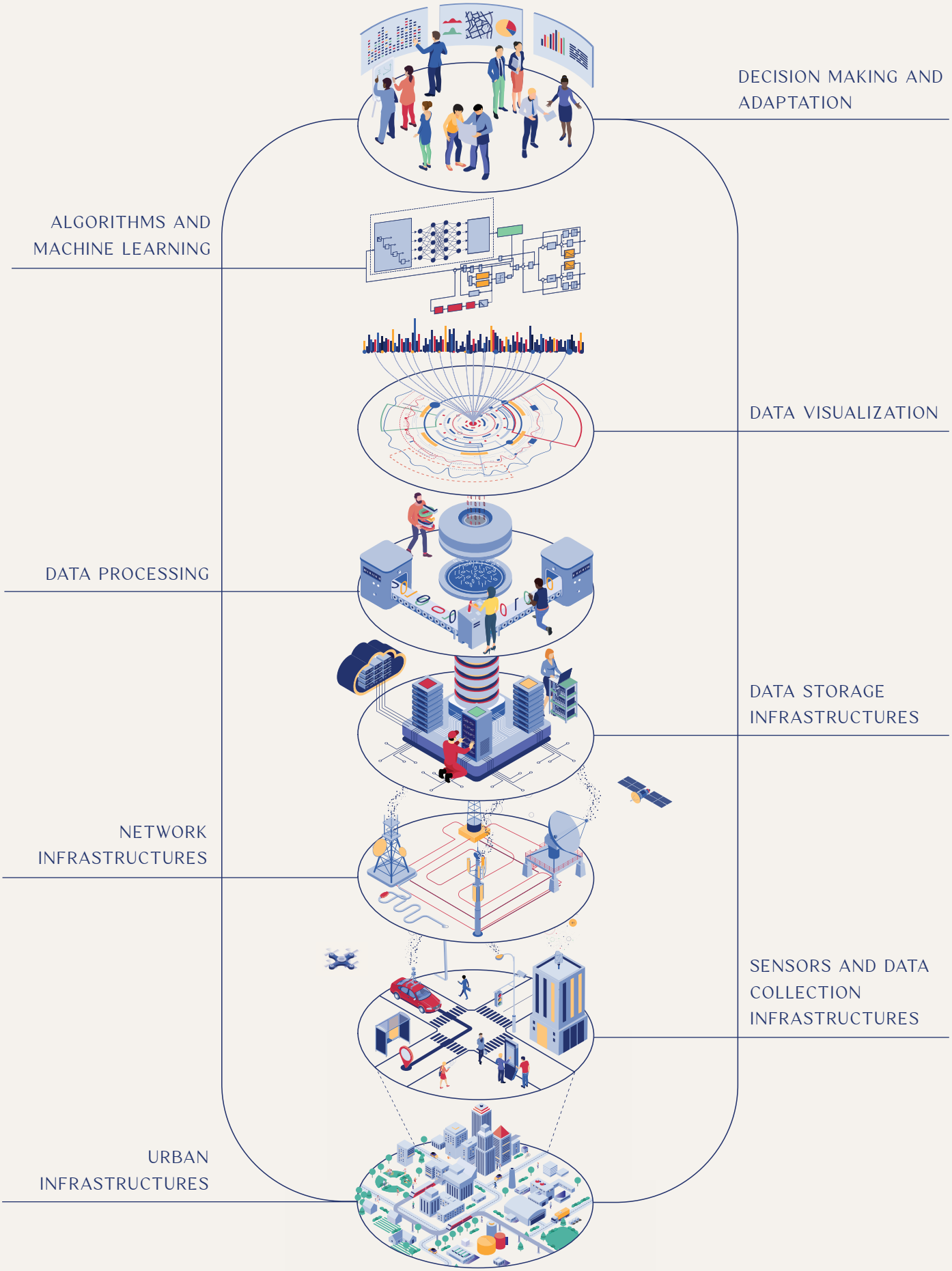
In the decision-making and adaptation step, transparent, collaborative communication mechanisms play a critically important role, especially since the people who design and build the AI project are rarely the subjects of its outcomes. Additionally, those who have the position of deciding how urban processes will change as a result of the AI solution rarely have intimate knowledge of how those systems work. Thus, communication has two purposes: first to convey the assumptions/ underlying components of the analysis, as well as the results and their implications to decision-makers; second to convey the transformation of data to model results to actionable change in a way that is understandable, agreeable, and accessible to citizens. Decision-makers need a sufficient understanding of the project, in order to use project outputs to make informed decisions. Citizens deserve a sufficient understanding of the project, in order to grasp the influence of AI on their daily lives. To better serve both of these ends, researchers in Boston have explored the role of gamification to engage diverse stakeholders in the AI implementation process, especially in cases where the tacit introduction of algorithms failed due to public discomfort (Popelka 2022). Eric Gordon and Gabriel Mugar (2020) offer that "Meaningful Inefficiencies," i.e. the practice of intentionally introducing play and deliberation into otherwise streamlined algorithm-based organizational change processes, can increase trust and ultimately provide a greater assurance of project success. In this way, evaluation can center around citizen needs, by collaborating with citizens to define success and using direct resident feedback as data sources to measure progress toward goals.

Emri Brickner, the former Smart CIty Department Manager for the city of Be'er Sheva in Israel, echoes Eckroth's suggestions, offering many insights on the practicalities of the decision-making and adaptation process, based on his time in that position. First he highlights the importance of being able to look to the past in order to make predictions about the future. If solutions are only presented, without the context of the past, then decision-makers cannot truly understand the nature of the solution and cannot make informed decisions. He also emphasizes the importance of selecting relevant key performance indicators (KPIs) and developing a strong change management and evaluation plan. He suggests hosting a workshop with all the stakeholders, in order to lay out the existing procedures, make sure everyone Is on the same page about how the proposed solutions will affect the existing processes, and come up with measurements to ensure that they have the intended effect.

Brickner also mentioned that he and his former team often wrangled with the question "Once an implementation has been successful, now what?" In this sense, they would implement a technology or service that generated new data or new analyses or new insights, but they did not always have the capacity to effectively make use of the products. Having a strong sense of how the data or analysis products will be used from the start of the project can help avoid this pitfall.

All the considerations detailed in the decision-making and adaptation step amount to ensuring that humans maintain governance of urban artificial intelligence systems, rather than merely subjecting themselves to those systems. Effective and responsible change does not arise from reacting to model outputs at face value. Instead, urban practitioners ought to remain actively involved in crafting, engaging with, understanding, and re-designing urban AI processes, such that they can *interpret* analysis results. Urban artificial intelligence holds its greatest value not as a means of making decisions, but as a means of *informing* them. As such, we offer this anatomy, and indeed the guide as a whole, so that decision-makers can learn about and engage more critically with the process.

URBAN AI ANATOMY



PART 2

CASE STUDIES

In this section, we present a number of case studies in a step-by-step manner, to show actual urban AI projects that have been implemented in cities. Specifically, we highlight the steps and considerations that went into each case. Of the three projects we describe in the following subsections – water meter lifecycle in Winnipeg, Canada, curb digitization for planning and asset management in Los Angeles, United States, and air quality sensing in Vilnius Lithuania – each covers a different urban sector, utilizes artificial intelligence in a different way/ different stage of the process, and each takes a different approach for integrating the artificial intelligence component. Additionally, we aimed to highlight cities of different sizes in different geographic regions, as examples of the diversity of artificial intelligence implementations (with not all urban AI projects originating in large smart cities). Future iterations of this guide will further expand the geographic scope of case studies presented.

WATER METER LIFECYCLE

IN WINNIPEG, CANADA

PART 2 | CASE STUDIES

In Canada, the City of Winnipeg's Water and Waste Department (hereinafter referred to as the Utility), is determined to leverage, develop, and expand artificial intelligence capabilities to better address challenges presented by growing urban population and footprint.

This journey began in 2018 and has since transformed into a multi-year program approved by the City to target real world opportunities and challenges faced by the Utility - prioritizing initiatives based on value and complexity.This was made possible by listening and attending to the needs of the stakeholders which guided the rapid development of three proof of concepts related to safety, customer accounts and billing, and water meter lifecycle to showcase the art of the possible.

This case study focuses on the water meter lifecycle proof of concept. In this context, Ernest Kwan and his team aim to build an in-house artificial intelligence solution, to be implemented as a strategic decision-making tool for key stakeholders. Ernest leads the the Data and Analytics Branch at the utility and is responsible for delivering the multi-year program that will develop the governance, methodologies, people, processes, architectures and technologies that will transform the Department's data assets into accurate, trusted, and assured insights to enable evidence-based decision-making.

STEP 1: PROBLEM IDENTIFICATION

The Utility observed a series of issues related to the health of water meters used to capture water consumption data by residential, commercial, and industrial customers . They frequently found significant discrepancies between the amount of revenue the City should have earned based on the volume of water it procured and the amount of revenue the City actually earned. This discrepancy suggested the presence of either uncounted water in the system, either unintentionally (as from a leak), intentionally (as in with theft), or as a result of faulty/failing water meter hardware. Identifying wastage and theft at the system-wide scale is critical in not only preventing revenue leakage, but also in accounting for the full life cycle of water procured in order to manage water resources in a sustainable manner.

Alongside day-to-day reconciling of such discrepancies on a case-by-case basis as they are discovered (if at all), the City routinely performs replacement of aging water meters. Traditionally, this has been done on

the basis of acute water meter failure or volume replacement by street or neighborhood for logistical and workforce driven reasons.

Ernest and his team decided to tackle the issue from an asset management and revenue leakage perspective. In total, the City employs approximately 220,000 water meters to service its growing population of approximately 800,000, most of which consists of direct meters (dumb meters). Approximately 5000 direct meters are replaced annually. In recent years, the City has been piloting and introducing new AMR/AMI-based water meters (smart meters) to new urban developments and renewal areas, which can total up to 1500/year.

Replacement of water meters is generally planned around available budget, and logistical and workforce convenience, rather than meter health - as such, some healthy meters may be replaced prematurely, while many failing meters persist in the meter population. Ernest and his team saw that other jurisdictions had used machine learning as a way of estimating water meter health and prioritizing replacement efforts, so they decided they would try the same.

Time

In 2018, the Utility began the process of identifying ways that they could leverage data and analytics to improve the efficiency and cost-effectiveness of their operations. The initial problem identification and development of a functional POC took 4 months.

Costs

New installations are typically an expense to the City, while damaged/lost water meters as a result of customer actions are at the customer's expense, although typically at a discounted rate for the hardware offset by some monetary recovery from salvage. Currently, damaged or lost dumb meters are procured at a discounted rate of $150-300 each to customers (i.e., this covers some of the cost of the dumb meters but does not cover any of the work or labor to procure). Each smart meter costs approximately $200 exclusive of work or labor. Annual revenue leakage from underperforming water meters can amount to up to a $1M. Thus developing an AI solution to help guide decisions around the health of the water meter population, water lifecycle, and fee schedules is of utmost importance in ensuring accountability and sustainability in water and wastewater services to residents, businesses, and industry.

Considerations

When undergoing the problem identification process, Ernest used two main factors in order to prioritize projects and assess feasibility: value to the organization and ease of implementation. When considering the value

to the organization, Ernest attempted to answer "why" a project should be implemented, focusing on what success would look like and how the end result would benefit the organization. When considering the ease of implementation, Ernest attempted to answer "how" a project would be implemented, the functional and technical complexity of the problem, the availability and quality of data and its sources (e.g. the format, quantity, and quality of the relevant data), the relevant stakeholders, and real-world precedent (case studies and lessons learned from other cities that had attempted similar projects in the past). When prioritizing projects, high-ease, high-value projects were given higher ranks than low-ease, low-value projects. During the process, Ernest faced some resistance to new technologies from stakeholders. In these cases, he found that highlighting the success of other cities' real-life use cases was an effective tool to get buy-in.

Skills

Ernest found that the following skills were assets during the problem identification phase:

- The ability to solicit project ideas from different business stakeholders.
- The ability to understand what the burning issues of the different business entities are.
- Having a holistic problem-solving attitude.
- Creating a range of implementation scenarios: building a solution in-house vs. buying a premade solution vs. outsourcing the custom development.

Stakeholders

Director of the Utility, Division Managers, Branch Heads

STEP 2: ANALYTICS CANVAS EXTRACT DATA

Since the water meter lifecycle use case focuses on identifying water meters that are incorrectly measuring water usage, water consumption is the main dataset that is of concern. The team utilized an analytics canvas to map and plan out the various facets of the problem, the opportunities and challenges, the functional and technical requirements to address each problem, and perform a high-level assessment of the quality of data available to satisfy the solution. This data needs to be consumed and presented in a manner that is meaningful and useful to stakeholders so that they can make better decisions around water meter

replacement strategies and authoring policies to prevent undesirable water consumption behaviors.

By ingesting water meter data, machine learning algorithms can analyze patterns in water use over time, detecting changes that might be consistent with deteriorating meter health (i.e., anomaly detection).

Furthermore, the dumb meters required manual reading by either a customer or by a technician dispatched to a premise. Recently, the Utility began shifting towards "smart meters," which would transmit the water meter readings digitally over wireless or broadband networks.

Time

Data collection from the dumb meters are extremely time-intensive, due to the manual nature of reading and logging the data. Due to these constraints, actual readings are only collected every few months (typically quarterly). In the absence of an actual water meter read, the Utility's main billing system 'estimates' the water usage at a given time. This presents limitations to the performance and reliability of machine learning modeling. However, the introduction of smart meters will substantially improve the sampling frequency, thus improving model performance.

Technologies

In the context of dumb meters, data is collected by customers or technicians reading the value on the dial. Once the values are read, the customer or technician would report the data either online or by phone, using an IVR system. The readings are centralized in a system called the ITRON Field Collection System (FCS). Furthermore, usage data from FCS is also ingested by Oracle Utilities Customer Care and Billing (CCB), the Utility's core customers accounts and billing system.

In the context of the smart meters, the City has been piloting several AMI and AMR solutions with plans to formally adopt them into the replacement standard in the near future. The AMI and AMR options either allow direct streaming of usage information to the utility's central office or cloud, and/or allows scanning of meter information by walking or driving near a meter via wireless communications. In either case, the adoption of smart meters would generally improve the accuracy, relevance and quality of data points collected, and open up new opportunities to expand the City's current quarterly billing schedule to monthly, and more quickly catch anomalies in the City's water consumption patterns.

Costs

Data collection for this project came with a cost trade-off, depending on the type of meter employed. For dumb meters, the cost of data collection equates to the cost of hiring technicians to go out and physically read the meters and record the data (a cost which increases with higher sampling frequency, as collecting data more often means spending more hours in the field). For smart meters, the cost of data collection equates to the one-time cost of purchasing the meters, installation, and any vendor costs to use/maintain the service. When comparing the two, the smart meters present a higher up-front cost, but save money in the long run by cutting down on the ongoing costs of sending staff into the field.

Considerations

Data collection considerations largely centered around cataloging existing data assets and identifying gaps. In doing so, the team performed a preliminary canvas/scan of the owner, format, volume, and quality of existing data sources, as well as whether the data sources were structured or unstructured. Equipped with this information, they dove deeper into the opportunities and challenges, and focused on the data that could be useful for the project and who to contact in order to gain access to those datasets.

Skills

Data collection for this project required the following skills:

- The ability to access various sources of data (e.g., databases, applications, systems, digital and paper documents, etc.).
- The ability to perform rudimentary data engineering tasks to prepare the data in order to support a Minimal Viable Product (MVP) for the purposes of the POC design and development.
- The ability to define and communicate preliminary data governance practices to optimize data quality and ensure proper usage of data.
- The ability to catalog, document and manage different datasets.

Stakeholders

Division Managers, Branch Heads, Superintendents, and Supervisors.

STEP 3: DATA PREPARATION, TRANSPARENCY, AND LIMITATIONS

The nature of the water meter data collection practices, particularly the human error associated with manually reading and entering meter values could result in a number of data quality issues. For the purposes of the POC development, the team had to prioritize data quality tasks that were most critical to the solution. Deficiencies and limitations, including any assumptions were assessed and documented. Stakeholders were made aware of these as well as further actions that would need to be taken beyond the proof of concept to make it the solution fully operational.

Time

During data preparations, the largest amount of time was with data wrangling and resolving meter data to service points and addresses. Examples included verifying and excluding meters that were no longer active (but somehow remained in the application) - this was most apparent when a service point had multiple meters (normally it is 1:1). In the production version of the solution, AutoML would be used to perform the bulk of the initial data preparations. The added benefit of this is that it would guide the correction of duplicate or erroneously entered meter assets associated with each service point in the source as well.

Technologies

For the purposes of the POC development, a Microsoft SQL Database was created to perform most of the data ingestion, preparation, and output activities. The dashboard was developed on both Microsoft PowerBI and Tableau to compare performance in showcasing visualizations that leveraged millions of data points.

For the production version of the solution, the team has implemented an Azure cloud-based technology stack, namely comprised of Azure Data Lake Storage (ADLS) Gen 2, and Azure Synapse for Data Warehousing and Compute to minimize system management overhead with an on-premise system. Azure Data Factory (ADF) and Talend Cloud are being used for data integration and transformations. The end-user experience is agnostic to visualization tools, however, for the purposes of generating the first set of critical use cases, the team will be using Tableau.

Costs

In this step, the primary costs were mostly attributed to staff and consultant time. The development effort was a one-time upfront cost (for the POC).

Considerations

Since the datasets came from different sources and inputs, they needed to be standardized and merged together. When data came from manual collection processes, human errors needed to be identified and corrected for. When data came from automated collection processes, device errors needed to be identified and accounted for.

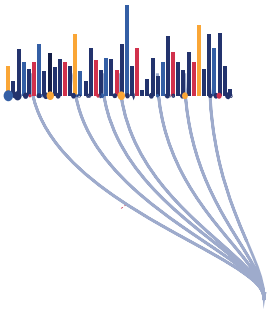
Skills

Data quality control and clean-up, in the context of this project, required a number of technical skills, including:

- Coding various scripting languages: SQL, Perl, Python, etc., to cleanse raw data and bring it into a usable form where relevant.
- Appropriately vectorizing (transform, convert and model) water meter reading data.
- Performing data sampling and data aggregation

Stakeholders

Superintendents and Supervisors



STEP 4: BUILD A PROOF OF CONCEPT (DATA VISUALIZATION)

Context

Ernest and his team prepared a data visualization dashboard as a “Proof of Concept,” to highlight the possibilities that the water meter data could be leveraged to help provide insights to water meter health, identify water usage trends, and organize and compare meter replacement cohorts to optimize rate of return for recovering revenue leakage and reducing water abuse.By presenting the information visually, stakeholders were able to

better understand the available data and current/past trends, assess the gaps in the existing datasets, and show the power of utilizing data and analytical methods to drive decision making. Notably, the proof-of-concept data visualization was intended as a means of achieving buy-in to receive the funding necessary for the integration of the AI component.

Time

The data visualization dashboard required 4 months to build. This POC was made portable to allow for easy sharing with interest parties. The full solution which includes operationalization of the POC dashboard and enhancement of the solution to include rate forecasting and anomaly detection utilizing Machine Learning is scheduled to be completed in 2023.

Technologies

The POC was developed as a lightweight, portable, self-contained application that used Tableau Reader to access. The future state is to operationalize this using the cloud technology stack mentioned previously in step 3 (this is a work in progress and is slated to be completed in Q4 2023 - Q1 2024).

Costs

Aside from the staff and consultant costs of building out the dashboard, there are also licensing fees associated with the Tableau software. There are no additional maintenance costs associated with using the POC, however, there will be capital and operating costs associated with the full solution that is currently a work in progress.

Considerations

For this project, Ernest and his team used proof of concepts utilizing data visualizations as a stepping stone to AI/ML solutions. Used in this way, data visualizations allowed the team to better understand data before running models or utilizing algorithms. When the graphs and trends were analyzed with the perspective of subject matter experts, who had a deep understanding of the underlying business context, Ernest and his team were able to identify discrepancies that could form the basis of what the models attempted to identify. They also found data errors and inconsistencies that could be corrected prior to AI implementation. However, beyond their exploratory capabilities, data visualizations can also be powerful communication tools. Most importantly, the proof of concept was used to gain support from the executive leadership to justify and secure funding to pursue modern analytics and AI/ML capabilities.

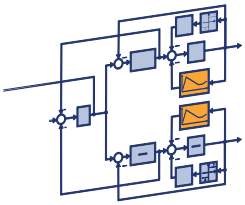
Skills

In the context of data visualization, Kwan found that the following skills were useful:

- General statistical literacy
- The ability to apply statistical analysis and visualization techniques
- Networking to understand which business mechanics produced the data
- Having significant curiosity to learn and understand a business process to its core
- The ability to generate smart hypotheses of the underlying mechanics of the business process
- The ability to use various statistical and visualization tools, such as Excel, R, Python, MATLAB and SAS programming language

Stakeholders

Director of the Utility, Division Managers, Branch Heads, Superintendents



STEP 5: INTEGRATE AI COMPONENT/ PREDICTIVE CAPABILITY

At the time of this guide, Ernest and his team have completed implementing foundational pieces of the technology stack to support the design, development, and operationalization of AI/ML solutions. The Utility has prioritized the development and operationalization of the Safety Analytics use case by the end of H1 2023, the Water Meter Lifecycle Analytics by the end of H2 2023, and Customer Accounts and Billing Analytics in 2024.

Time

For each analytics use case, the predictive analytics components will take approximately 3-6 months to build. Following their operationalization, there will be routine maintenance, monitoring, and tuning to ensure the quality and performance of the solutions are maintained. Additional projects may be created to enhance these solutions further, as requested by stakeholders.

Technologies

Ernest was not able to share details about the specific algorithms and techniques that will be used in the artificial intelligence components as they have not been formalized at the time of this guide. However, he was able to share that the predictive algorithms used will be oriented

toward estimating which meters should be prioritized for replacement or recalibration based on variables such as meter age, meter design, meter reading route, and water usage data. The solution will be designed, developed, and operationalized on the Microsoft Azure cloud technology stack described in the data processing section.

Costs

The AI components will be co-developed by internal and external resources (e.g., consultants, contractors, vendors, etc.), but Ernest hopes to build internal capability and capacity so that his team can become less reliant on external resources . However, Ernest recognizes that specific situations may call for a specific set of expertise or skills that may not come naturally from his team, in such cases, experts external from his team will be brought on to assist.

Considerations

The implementation of predictive AI capabilities allows assessment efforts to be run continuously (as often as data are collected), rather than as one-off on-demand efforts, providing greater insight and increased responsiveness to the situation on the ground.

Skills

The skills that Kwan will look for as he bolsters the internal artificial intelligence capabilities within his team include:

- Subject matter expertise to design and calibrate the models
- Data science skills to build and run the models
- The ability to apply a wide array of machine learning models: various regression and clustering techniques, ensemble techniques (random forests, XGBoost) and time-series analysis.
- Experience in real-world testing of ML models such as champion-challenger (A/B testing) and cross validation.

Stakeholders

Director of the Utility, Division Managers, Branch Heads.

STEP 6: UTILIZE AI PREDICTIONS FOR DECISION MAKING

Context

In order to ensure the success of the project, Ernest highlighted the need to identify key stakeholders and business champions. The resulting solution(s) should be designed with these people in mind, in order to ensure effective adoption and use. It is anticipated that early adopters of these solutions will become advocates, influencers and promoters for others to try and adopt AI/ML-driven solutions for their day to day decision making. This will help immensely with the organic growth of an evidence and insight-driven culture within the Utility and beyond.

Considerations

Understanding the value of utilizing AI solutions for ongoing decision-making, strategic planning, and monitoring of the Utility's health. The prioritization of the three use cases: Safety Analytics, Customer Accounts and Billing, and Water Meter Lifecycle addresses three very important themes of social, economics, and environment. Furthermore, the success of advanced analytics and AI/ML adoption is driven by a user/business-centric approach, as opposed to a technology-centric one. Additionally, building in strong review and evaluation processes can help ensure that the solution is working as intended and benefitting citizens equitably.

Skills

Ernest Hopes to build a team with the following skills, in order to in order to monitor and maintain the solution, so as to ensure its continued use and effectiveness:

- The ability to understand the data flows and sources at the business/functional level, and of the data necessary for the development and proper operation of models.
- The ability to integrate model performance management tools into the current business infrastructure.
- The ability to run champion-challenger (A/B tests) on production systems that analyze how the proposed ML model performs against the current methods for decision support and process automation.
- The ability to continuously monitor the execution and health of models within the process environment — or at the very least providing a strong integration with the IT team in charge of the process/app operations.
- Assist the user community in raising awareness, education/training, and troubleshooting the solution

Stakeholders

Director of the Utility, Division Managers, Branch Heads, Superintendents, Supervisors, Team Leads, and all other users.

KEY TAKEAWAYS

The water meter lifecycle use case in Winnipeg demonstrates the skills, technologies, and capacity-building efforts that are necessary when putting together an internal government team with artificial intelligence capabilities. This approach requires a lot more political buy-in and a higher degree of resources, as it necessitates creating positions for and bringing on technically-proficient staff members, as well as evaluating and purchasing new hardware and software systems. However, the effort can pay off in that an in-house team often leads to artificial intelligence projects that have a higher degree of sustainability in the long-term (i.e. which can outlast implementation contracts or platform subscriptions). Following the overhead of the initial start-up, building out an internal artificial intelligence team and technology stack can also allow for a quicker process and lower costs for starting new projects, as the municipality can use existing resources rather than engaging in an external bidding process for every new implementation. As shown in the Winnipeg example, utilizing the services of a consulting firm when scoping technical requirements and building out a team allows a municipality to leverage private sector technical expertise to ensure that the initial investments into the program are directed appropriately. The Winnipeg example also highlights the fact that building internal capacity is a continuum, and that municipalities can take a phased approach to artificial intelligence project implementation, expanding the scope and scale of the project as more human and technological resources become available.

CODE THE CURB: CURB DIGITIZATION AND PLANNING/ASSET MANAGEMENT

IN LOS ANGELES, CALIFORNIA

Due to its massive size and highly diverse citizenry, the City of Los Angeles, in the United States, is working on incorporating AI into its operations, as a means of improving service delivery at scale, as well as making the city more connected and reducing inequalities. With the City's SmartLA 2028 strategic plan, released in 2020, departments across the city are expected to advance the City's main technology goals of enhancing: infrastructure, data tools & practices, digital services, connectivity & digital inclusion, and governance. The Los Angeles Department of Transportation (LADOT) is among the city agencies involved in digital transformation. Their specific focus is on infrastructure and mobility. They recently embarked on a curb digitization and asset management project, which leverages AI to catalog various features of the street curb and signage, to support data-driven curb-use management. The data collection and technical aspects of the project are handled by the vendor, IBI Group. However, Tomas Carranza, a Principal Transportation Engineer for LADOT, oversaw the project and served as the department's main point-of-contact.

STEP 1: PROBLEM IDENTIFICATION AND PLANNING

Los Angeles is notorious for its complex parking regulations and curb use restrictions. Dynamic and more proactive curb management was a priority set forth by the Los Angeles City council, as well as a strategic goal outlined by LADOT. However, there was not a definitive repository of current curb uses and regulations. Without that knowledge, it was difficult to plan for changes that would improve circulation throughout the network. The City created a five-year roadmap for digitizing all curb assets and regulations, but, due to budget issues, ended up thinning it to a much more pared down pilot program focusing on three dense commercial areas as a starting point. The pilot sought to answer three questions: What are the benefits and challenges of the different asset collection methodologies that are out there? How does the land use context affect your ability to collect data? What is the optimal and scalable solution for curb digitization?

Time

Since LADOT needed to contract with a vendor for the project, the main time constraint was the bidding/solicitation process. The city has a lengthy series of administrative steps and checks that ensure fair purchasing practices. However, those procedures can also lengthen the timeline of completing a project. The city has a shortcut process

for pre-approved vendors, which can speed up the contract approval time to around 1 month, but none of the pre-approved vendors had the specific skills and technology to accomplish the curb cataloging task. Since LADOT went with a non-pre-approved vendor (IBI Group), the approval process would have taken 9-12 months. However, the City of Los Angeles had established a public-private mobility innovation accelerator (Urban Movement Labs), which was able to formally take on the project, with LADOT acting as an advisory arm. In this way, they were able to circumvent some of the purchasing procedures.

Costs

Costs were a major factor in determining the scope of the project. While the original goal was to digitize all of the curbs in the city over the span of five years, the COVID-19 pandemic and associated fiscal cutbacks caused a reduced budget for the project. In light of this, the city embarked on a reduced-cost, reduced-scope pilot program to digitize curbs in three highly dense commercial areas as both a starting point and as a way of honing in and refining the process and cost estimates.

Considerations

Since the Los Angeles City Council had set forth better curb management as a priority and LADOT had outlined it as a strategic goal, Carranza and his team had little trouble getting buy-in for the project. With the top-down approach, there was also more pressure for Carranza to ensure that the project got done quickly. In light of this and the COVID-19 budget constraints, the pilot program approach was a useful way to demonstrate meaningful progress and work out the kinks of the project, without spending too much money upfront. However, the City Attorney's Office generally avoids being directly involved in pilot programs (to avoid perceptions of unfair advantage), which was another advantage of running the project through Urban Movement Labs.

Skills

Since the problem had already been identified by the City Council and LADOT, most of the skills involved in this step centered around finding ways to make the project come to fruition. These included:

- The ability to find alternative sources of funding (e.g. partnerships, grants, regional collaborations, etc.)
- Project management/ scoping
- An understanding of procurement/ Legal processes

Stakeholders:

The primary stakeholders in this step of the project were the City Council/ Mayor, in terms of setting the priorities, LADOT, in terms of reinforcing

the priorities and finding a way to make it happen, Urban Movement Labs as the accelerator that allowed the project to happen with less red tape, and the City Attorney's Office and City Procurement Office, which set the constraints for the vendor solicitation/ contract processes.

STEP 2: DATA COLLECTION

For this project, the relevant data sources included curb use restrictions (length, location, and color of painted curbs), parking restrictions (location, timing, and parameters of parking regulations), and other assets in the public right-of-way (bike racks, traffic signal equipment, parking meters, etc.). Understanding how curb use and parking restrictions fit together and interact across blocks allows for a more holistic view of the streetscape and aids in scenario-based modeling planning. While information existed on signs and along the curb within the streetscape, there were no datasets that collected the information in a consolidated and machine-readable way. In order to collect the data, manually walking every foot of curb space in the City to collect information from signs and curbs would have taken a massive team of people years to accomplish. As such, LADOT decided to partner with a company called IBI Group to utilize their curb management platform, CurbIQ, which offered artificial intelligence-enabled curb data collection methods (mobile mapping), to conduct the pilot. IBI Group's data collection mostly consisted of imaging the curb and classifying the images. As an additional data source, citation data were also collected, for QA/QC purposes later in the project.

Time

The data collection time depended on the complexity of the curbscape as well as the collection method used. For CurbWheel-aided manual collection, data collection took between 1 and 1.4 hours per mile. AI innovations allow for IBI Group to collect equivalent data from imagery, such as car-mounted smartphone camera footage. When using those technologies, data collection time dropped to between 0.1 and 0.6 hours per mile. In the context of the car-mounted cameras, pre-survey route optimization can serve to further reduce data collection times. Since the curbscape goes through continual changes, Carranza estimates that a curb survey would have to be repeated every 3-4 years, with manual updates as-needed in the intervening years.

Technologies

The data collection technologies used were the SharedStreets CurbWheel and car-mounted smartphone cameras. The CurbWheel is a manually-operated data collection device that allows users to measure curb use regulations and log them in an accompanying smartphone app, while they push the device through a city. It has a high degree of accuracy, but takes more time to operate, as it requires a human to push it and log the data. Car-mounted smartphone cameras were also used to collect images of the streets, which AI data processing interpreted and converted into data. The smartphone camera + AI process has a lower accuracy rate in highly dense areas, but allows for faster data collection.

Costs

The specific costs were confidential, but Carranza did share that data collection via CurbWheel is 30-40% more costly than data collection by car-mounted smartphone, due to the fact that it takes human operators much longer to do the collection. However, he did mention that anyone can operate a CurbWheel with a bit of training. As such, he recommended tasking interns to do the CurbWheel data collection, rather than hiring a contractor, as a way of cutting the costs associated with conducting the CurbWheel surveys.

Considerations

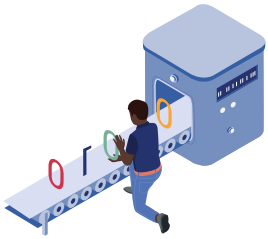
There are trade-offs in cost, accuracy, and time between the two approaches (with CurbWheel being the most accurate, but car-mounted smartphone cameras being the most cost- and time-effective). The busyness of the location being surveyed can dictate the optimal combination of the two technologies, as car-mounted smartphones are adequate in less-busy areas, but the CurbWheel is necessary in locations with a higher likelihood of objects and vehicles blocking the curb features. Routing and time of day can have a big effect on the timeliness and accuracy of the surveys as well. Data ownership presents another major consideration: under the terms of the pilot, IBI Group maintained ownership of all data collected and processed. In this agreement, LADOT could interact with the data, but could not remove it from the CurbIQ platform and would not maintain access following the termination of the pilot.

Skills

Aside from training for using the CurbWheel device and a knowledge of route optimization to improve the usage of car-mounted smartphone cameras, data collection for this project required few skills.

Stakeholders

The main stakeholder in this step was IBI Group and their curb data visualization tool CurbIQ, which provided the devices and knowledge about this type of data collection, but they were also guided by LADOT and Urban Movement Labs.



STEP 3: DATA PROCESSING (AND AI INTEGRATION)

Context

In this case study, rather than having AI as the final product that aids decision-making, AI was incorporated as an integral part of the data collection and processing step, which in turn allowed for the generation of useful information that could be used to better inform curbside management and planning. AI technologies were used for three main purposes: detection (identifying signs and objects within the images captured in step 2), comprehension (understanding the signs and objects that were identified), and geolocation (determining the precise location of the signs and objects). Without the processing step in this case, which turns streetscape images into curb data, the photos that were collected would be meaningless.

Since the AI/ML models were not 100% accurate, additional quality assurance methods were necessary to validate the output data. The QA/QC took two forms. First, IBI Group employees manually analyzed the images that the models could not confidently process and derived the necessary information. Following the automated and manual processing procedures, LADOT employees compared the image-derived regulations on each block to the parking citations issued on each block, to further confirm the accuracy of the data processing. After the data collection and processing efforts were underway, the Open Mobility Foundation released curb data specifications for better standardization and portability, so the data will eventually have to be fit to that new schema.

Time

IBI Group went through an R&D process prior to this project, in order to develop the algorithms that it uses for data processing, so upfront development times did not factor into this project. However, there was still time devoted to data processing, which happened to differ based on

data collection method and the busyness of the curb. For the CurbWheel data collection method, data processing took between 0.4 and 0.7 hours per mile. For the car-mounted smartphone cameras, data processing took between 0.1 and 0.2 hours per mile.

Technologies

From an artificial intelligence standpoint, the main technologies used were various machine-learning models. These included computer vision models, which were used to identify objects within the cameras and videos and text comprehension models, which were used to make sense of the signage. Geographic Information Systems were also used, in order to associate the footage and generated data with the data collection location and map the curb uses and restrictions. Since all of the processing was handled by the vendor, LADOT did not need to worry about building any of the processes nor maintaining any technological infrastructure.

Costs

Carranza couldn't share specific costs for the project. That said the data processing costs were included in the overall cost of the pilot program, as part of the partnership between UML and IBI Group and the use of their services.

Considerations

It's important that the data are processed in a complete and accurate manner, otherwise they hold little value for future analysis. If any signs are ignored, misinterpreted by the algorithms, or incorrectly located, then they do not reflect the true curb regulations and cannot be used for meaningful planning efforts. In different areas of the city, signs and objects were encountered, which weren't included in the original algorithm training data. In these instances, the new signs/objects had to be manually classified and incorporated into the training dataset. In this sense, economies of scale benefit AI classification algorithms, as they are able to account for more possibilities. High density areas posed more of a challenge for the identification, understanding, and geolocation models, as tall buildings reduced device GPS accuracy and busy/cluttered curbsides prevented the models from identifying some pertinent objects.

It can be helpful to have multiple checks in the system, using multiple verification methods, to ensure that models are performing in the intended ways and to the optimal levels. If the models are flawed, then any subsequent use of processed data or derived outputs will be useless, misleading, and even potentially dangerous, depending on the context.

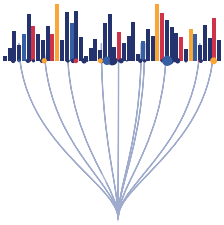
Skills

Although Carranza and his team did not need any technical skills for the data processing step, as the vendor handled the data processing, skills that would be useful if conducting this type of data collection, processing, and validation in-house are:

- Data science skills
- GIS skills
- AI/ML skills

Stakeholders

The main stakeholder in the data processing step was IBI Group, as they developed and performed the data processing. Additionally, LADOT performed secondary verification, to ensure the accuracy of the results.



STEP 4: DATA VISUALIZATION/ ANALYSIS

As part of their standard service, IBI Group provided LADOT with access to its Curb Viewer platform, CurbIQ. The platform was custom-built by IBI Group, but is designed for general use by anyone who uses CurbIQ's services (I.e. it's not customized to any one city). LADOT utilized the CurbIQ platform in order to visualize the derived curb regulations. Since IBI Group maintained ownership of the data as part of the pilot program agreement, LADOT was not able to build their own visualization platform nor was it able to incorporate the data into existing visualization/analysis tools.

Time

IBI Group developed the Curb Viewer platform prior to the City of Los Angeles project, so upfront development time did not factor into this case. There was minimal time to load the LA data into the platform and create a sectioned view that included only Los Angeles data.

Technologies

The only data visualization platform that LADOT used for this project was the CurbIQ platform, which was custom-built by IBI Group.

Costs

Carranza couldn't share specific costs for the project. That said costs associated with accessing the CurbIQ platform were included in the overall cost of the pilot program, as part of the partnership with IBI Group and the use of their services.

Considerations

When assessing the merit of the CurbIQ visualization platform, Carranza brought in the LADOT IT and GIS teams, in order to do a deeper dive into the ways the data were being received and visualized in the platform. Engaging these other sections within the department allowed the City to ask questions and better understand the technology and visualization tools beyond what Carranza, as a civil engineer and traffic engineer, might have considered on his own. In terms of the platform itself, it allows users to visualize the various curb regulations, uses, and parking restriction signs by time of day, in an interactive map-based interface. Additionally, the CurbIQ platform contains an analysis module, which allows users to better understand the distribution of curb regulations among surveyed curb segments, as well as model various restriction change and parking pricing scenarios.

Skills

Since the data visualization platform was already built and provided by the vendor, Carranza and his team only needed the skills to be able to interpret the visualizations presented within the platform. These included:

- Data literacy
- Traffic management understanding

Stakeholders

The main stakeholder in this step was IBI Group, which built the platform. The LADOT IT and GIS sections were also stakeholders in terms of evaluating the visualization platform from a technical perspective. LADOT was the primary stakeholder that was involved in the interpretation of the resulting CurbIQ charts and graphs, due to the LADOT employees' subject matter expertise.

STEP 5: DECISION MAKING

LADOT has three main analysis projects planned based on the data collected as part of the pilot (as well as what will be collected during the broader Code the Curb program). These include improving curbside asset management, determining the optimal locations for zero-emissions delivery zones within the city, and better enforcement of curbside restrictions. In order to accomplish the final piece, LA also needed to pass an enforcement ordinance, to allow it to perform enforcement of new curb regulations, like zero emission delivery zones. Although the pilot does not allow these actions to be undertaken for the city as a whole, it proved quite useful in supporting such analyses in the target neighborhoods, as well as inspiring confidence in the program among city council members and demonstrating material progress toward the strategic goals and initiatives set forth by the council and department.

Technologies

Although not a very technical step, the main technologies used in this step of the project were (and continue to be) the CurbIQ platform, which is used for analysis-based planning as well as communicating details about decisions that have been made, and an additional asset management platform, which allows for curb assets to be better-tracked and facilitates decision-making around their repair and replacement.

Considerations

Carranza highlighted the importance of having buy-in from the mayor and City Council members. Such support is quite helpful in terms of obtaining funding and pushing projects forward that enhance curb management from piloting to operational. In this regard, completing pilot programs can be a great way to show the utility of the technologies/ tools, demonstrate progress and capabilities, and secure support for the continuation of the project.

Skills

Since the technical components of the data analysis were pre-built and provided as part of the CurbIQ platform, Carranza and his team only needed the traffic engineering and planning skills to be able to know which planning modifications might be reasonable scenarios.

Additionally, they needed the skills to interpret the analysis results, make suggestions based off of the outputs, and communicate those suggestions to collaborators and decision-makers.

Stakeholders

The main stakeholders in this step were IBI Group, which built the analysis components of the data platform, LADOT, which ran the analyses and made suggestions based on their interpretation of the results, and LA City Council/ the mayor, who had the power to make decisions and changes based off of Carranza's and LADOT's suggestions.

KEY TAKEAWAYS

The Code the Curb use case in Los Angeles demonstrates the implementation of a customized (but not bespoke) artificial intelligence solution, which was co-produced by a municipality and a private technology company. Even though the LADOT did not oversee the full implementation of the project, they retained a high degree of collaboration in building the models, as well as oversight on the process, through the configuration of their relationship with the vendor. The Los Angeles case study also highlights the fact that urban artificial intelligence projects can return useful results, even without the implementation of large, continuously sensing data collection systems. The temporal context of the Code the Curb project (i.e. the fact that curb uses and signage change on a timescale of months or years) meant that data sampling could occur less frequently. As such, LADOT was able to tailor a solution to the specific problem definition, by focusing only on key locations and discrete data collection efforts. This specificity reduced the cost and project timeline, while still allowing the analysis results to have a meaningful impact with regard to solving the problem. Smaller municipalities, which lack the resources to install and maintain citywide urban sensors, can benefit from some of the pilot techniques used in Los Angeles, in order to still engage with artificial intelligence and design more tailored implementations.

AIR QUALITY MONITORING

IN VILNIUS, LITHUANIA

PART 2 | CASE STUDIES

The City of Vilnius, Lithuania has recently made a push to incorporate data-driven decision-making into its planning and operations. In order to do so, it has emphasized the use of key performance indicators (KPIs) as the primary method of evaluating proposed projects and programs. The planning authority of Vilnius, Vilniaus Planas, recently partnered with a platform-based air quality monitoring company, Breeze Technologies, to implement air quality sensors throughout the city. The air quality sensors and resulting data support planning efforts related to transportation, urban ecosystem planning, and healthier living. In this project, artificial intelligence is used both for data processing and for predictive analytics. On the city side, the project is overseen by Simas Sodys, the Lead of Innovation for Vilniaus Planas.

STEP 1: PROBLEM IDENTIFICATION

Vilnius, Lithuania has three main KPIs that it uses to improve city services: a happiness index, a travel time index, and life expectancy. Following residents' responses to surveys commissioned by the city, it became clear that air quality and pollution were major issues of concern citizens. These issues, in turn, were affecting Vilnius's performance in its KPIs. As such, the city decided to implement a sensor system to better understand and more quickly respond to air quality issues and hopefully improve its performance along the KPIs. Additionally, this initiative fit into Vilnius's broader goal of building an information ecosystem that crosses industries and sectors to connect all people within the city. The project aims included both supporting better planning to improve air quality and making air quality data available to citizens.

Time

In terms of identifying the problem, conducting resident surveys and developing key performance indicators took a few months. In terms of developing a solution to the problem, since Breeze Technologies approached Vilniaus Planas with an agreeable air quality monitoring solution, the scoping phase was relatively short. Since the solution started with the vendor, Sodys did not need to go through the process of figuring out potential options and researching a variety of vendors. Given that the Breeze Technologies was a unique offering, the City of Vilnius was able to procure the contract through direct tender, rather than a competitive business process. This fact also served to speed up the problem definition step substantially.

Costs

Breeze technologies designed a contract package that would serve the needs of Vilniaus Planas, while still being cost effective. For example, Breeze proposed the minimum amount of sensors needed to perform accurate air quality monitoring.

Considerations

In identifying a problem and determining an approach to tackle it, it's highly important to understand your citizens and their needs. As such, conducting regular resident surveys allowed the City of Vilnius to truly understand what aspects of living in the city its citizens found most concerning. Additionally, as through the key performance indicators, Sodys found it helpful to have clear benchline metrics, which allowed him and his team to identify operational and planning areas that needed improvement and aligned with citizen concerns. Having key performance indicators also allows them to track the performance of the project implementation. To ensure that the solution serves to improve the intended problem.

Skills

For this project, the main problem identification skills utilized were:
- The ability to conduct surveys and perform qualitative analysis
- Knowledge of stakeholder engagement strategies
- The ability to review contract proposals

Stakeholders

In the problem identification step, the citizens, who responded to the survey and voiced concerns with living in Vilnius, the City of Vilnius and Vilniaus Planas, which have the responsibility of responding to citizen requests and improving the quality of life within the city, and Breeze Technologies, which approached Sodys with a proposed solution to the air quality issues that citizens were experiencing.

STEP 2: DATA COLLECTION

Although a handful of antiquated air quality sensors exist within the City of Vilnius, the majority of air quality monitoring and response efforts in the recent past relied upon citizens to report air quality issues. As such, there was a temporal disconnect between when a citizen would report an issue and when a technician from the city would be able to come out to assess and respond to the issue. Often, the nature of the situation would often in the intervening time. For this reason, Vilnius partnered with Breeze technologies to implement a sensor system that would collect air quality metrics in real time. The sensors collect data on localized temperature and humidity, climate, and pollutants. Those data points provide an impression of the environment around the sensor and can indicate the presence of irritants, allergens, or pollutants in the air.

Time

The data collection step had a couple of up-front time investments including the initial planning and strategizing as to where to locate the sensors so that they would have the biggest impact, as well as the installation of the sensors at the proposed sites. Once the sensors were installed, they sampled, and continue to sample, data in near real-time. Occasional maintenance, such as replacing the battery, is required on a periodic basis.

Technologies

Breeze designed its own sensor system, which is what it provided to Vilnius Planas for the data collection step. Included in the system are outdoor air quality monitoring sensors (made with low-profile, cheap, lightweight materials), as well as data transmission technologies (in order to send the data from the sensor devices to the Breeze data platform. The sensors are flexible, in that they can operate over public WiFi, if available, or cellular data plans if no WiFi connection is available. The sensors can also operate over LoRa networks, which carry less data than traditional networks, but use less power.

Costs

The data collection step encompasses the cost of the sensors (which Vilnius purchased; by contrast some sensor vendors maintain ownership of the sensor devices and lease them to the participating city on a subscription basis), the cost of installing the sensors, and a subscription to the Breeze platform. Breeze takes the approach of catering an installation package to the needs of the city, based on geographic size, in order to reduce the costs.

Considerations

Traditional air quality monitoring sensors can be quite costly and as large as a car, which makes it difficult to get adequate coverage within a city (usually cities would only have one or at most a handful of these sensors). Additionally, traditional air quality monitoring stations relied on manual processes to calibrate the sensors and collect the data, which meant that it could take months before an authoritative dataset was ready. Contemporary technology has allowed sensors to be much smaller and much less expensive to deploy. The Breeze sensors require a vertical mounting surface (such as a utility pole or streetlight), a power supply (can be a standard outdoor power outlet), and a means of transmitting the data collected (e.g. public wifi, LORA networks, cellular data plans). Cellular data plans facilitate the quickest deployment, since the sensors can be ready to go as soon as they are mounted, without requiring additional configuration or placement considerations. Breeze worked with the City of Vilnius to determine the optimal number of sensors, as well as the best placements.

Skills

The following skills were utilized for air quality monitoring data collection, in the context of this project:

- Sensor Installation skills (which were handled by Breeze Technologies)
- Data analysis and urban knowledge to determine the optimal placement of the sensors

Stakeholders

The primary stakeholders in this step of the project were the City of Vilnius and Vilniaus Planas, which provided the institutionalized knowledge and urban subject matter expertise necessary to inform the optimal placement of the sensors, and Breeze Technologies, which collaborated on determining the placement of the sensors, in addition to installing the sensors and managing the data collection pipelines.

STEP 3: DATA PROCESSING AND QA/QC (AND AI INTEGRATION)

Context

In order to facilitate near real-time air quality monitoring, Breeze technologies utilizes cloud-based artificial intelligence methods to calibrate and process the data collected by the sensors. In this context, AI is primarily used for outlier detection (to determine and account for anomalous data points), sensor health monitoring (to allow cities to proactively replace the units before they fail), and to incorporate the inherent characteristics of the sensors (i.e. how the sensors behave under various conditions; this allows the algorithms to better interpret the raw data values, in the context of the sampling environment).

Time

The data processing algorithms and infrastructure were built by Breeze Technologies prior to their partnership with Vilniaus Planas. As such the upfront development time traditionally associated with building AI solutions did not factor into this project. The processing itself happened in near real-time, as soon as the data streamed into the Breeze Technologies platform.

Technologies

The main technologies utilized for this step of the project included AI algorithms (primarily outlier detection and clustering algorithms), cloud infrastructure (to support the AI algorithms), and big data storage systems. Since Breeze technologies built and maintains all of these technologies, offering the processed data product to clients through their platform, which operates under a software-as-a-service (SaaS) model, Vilniaus Planas did not need to interact with the technologies at all.

Costs

The cost of accessing the Breeze Platform, which processes and produces the finished data product, is incorporated in a subscription to the Breeze platform

Considerations

The use of artificial intelligence in data processing allows for data to be available in near real time (with AI-enabled integrated QA/QC checks), whereas alternative manual approaches require much longer and more timely processing and quality assurance. It can take a substantial amount of time to develop and refine the algorithms, but using a pre-built subscription service allows users like Vilnius Planas to take advantage of those models without having to invest time, money, or skills in their development. Additionally, the SaaS model makes data processing much cheaper, more scalable, and better facilitates the AI components. However, SaaS products often do not allow for customization, if a particular use case happens to not fit the mold, for which the solution was designed.

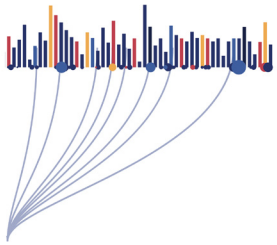
Skills

Since the final, processed data products were presented to Vilniaus Planas, Sodys and his team did not need any artificial intelligence or data science skills to perform the processing. However, the Breeze Technologies team utilized the following skills in developing the algorithms and data processing components of the platform:

- Data Science Skills
- AI model building and calibration
- An understanding of the sensor dynamics

Stakeholders

The sole stakeholder in the data processing step was Breeze Technologies, as they did all of the development work prior to the Vilnius Project, and they conduct all the data processing behind the scenes, with no involvement from the client.



STEP 4: DATA VISUALIZATION

Once the air quality data are collected by the sensors and processed by the AI technologies, they are fed into a data visualization platform built by Breeze technologies. The platform integrates the data from all the sensors into a unified map/visualization, which displays patterns and unlocks insights about spatiotemporal air pollution dynamics that the raw data on their own do not sufficiently show. Additionally, the City of

Vilnius wanted to incorporate the air quality data into its own integrated data visualization applications, in order to overlay air quality data with other municipal data like mobility data.

Time

The data visualization platform was built by Breeze Technologies prior to their partnership with Vilniaus Planas. As such, the upfront development time traditionally associated with building data visualizations and the associated infrastructure did not factor into this project. There was a short time investment in order to configure the visualizations to be specific to Vilnius, as well as a time investment of a couple of hours in order to train Sodys and his team on how to navigate the platform and interpret/understand the visualizations.

Technologies

The data visualization platform was built by Breeze Technologies prior to their partnership with Vilniaus Planas. As such, the upfront development time traditionally associated with building data visualizations and the associated infrastructure did not factor into this project. There was a short time investment in order to configure the visualizations to be specific to Vilnius, as well as a time investment of a couple of hours in order to train Sodys and his team on how to navigate the platform and interpret/understand the visualizations.

Costs

The cost is incorporated in a subscription to the Breeze SaaS platform. Since citizen data transparency is a primary goal of Breeze as a company, they will discount the subscription cost, if the government agrees to share their air quality data to Breeze's public platform.

Considerations

Integrated data visualization systems can promote transparency both within governments (allowing departmental silos to be broken down and providing a basis for collaborative planning) and between governments and citizens (allowing citizens to better understand the dynamics at play in their cities). Additionally, visualizing historical and spatial data can provide deeper insights and highlight patterns that were previously unseen in similar datasets. Using a pre-built data visualization platform can simplify the data visualization step quite substantially, but it can also limit the possibilities of exploring the data, if the visualizations are not customizable. To get around this, data portability supports better integration between multiple datasets from different sources and enhances a government's ability to utilize its own data outside of a vendor platform.

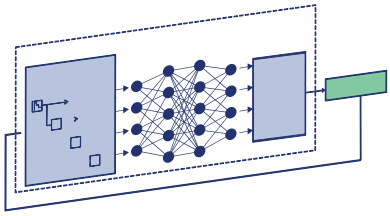
Skills

Utilizing the Breeze Technologies data visualization platform requires minimal skills beyond some cursory training. In terms of building the visualizations, the Breeze team made use of the following skills:

- Web development skills
- GIS skills
- Data engineering skills
- Data science skills

Stakeholders

The primary stakeholders in the data visualization step were Breeze Technologies, which developed the data visualization platform, and Vilniaus Planas/ the City of Vilnius, which were the users of the platform and which developed their own integrations to bring the Breeze data into internal data visualization applications.



STEP 5: AI COMPONENTS/ PREDICTIVE CAPABILITY

Beyond utilizing AI for data processing, the Breeze platform is integrated with AI technologies to support better analysis and planning efforts, in a number of ways. First, AI algorithms are used to interpolate between sensors, in order to fill data gaps where there might not be sensors, generating coverage of a greater area of the City of Vilnius. This allows the Vilniaus Planas to have a more complete picture of the state of the air quality within the city. Secondly, AI is used to inform clean actions, by modeling the parameters of various clean-air interventions, in addition to taking into account the context-based performance of previous air quality interventions. This allows the analysis components of the Breeze platform to identify the potential outcomes of various remediation scenarios. Additionally, although this has not yet been implemented, Vilniaus Planas intends to integrate the Breeze data with its MaaS platform, in order to perform AI-enabled transport network and traffic engineering modifications that best provide for efficient and clean mobility throughout the City.

Time

The analytical AI components were built by Breeze Technologies prior to their partnership with Vilniaus Planas. As such, the upfront development time traditionally associated with building analytical AI components did

not factor into this project. There was, however, a time investment of a few hours in order to train Sodys and his team on how to navigate the platform and interpret/understand the results of the analytical components.

Technologies

The main analytical technology used was Breeze Technology's own platform, which is what Sodys and his team interacted with when they wanted to explore various clean air interventions and potential outcomes. In developing the analytical components of the platform, Breeze Technologies employees utilized skills such as:

- Knowledge of and ability to implement AI algorithms (including interpolation and prediction algorithms)
- Knowledge of data integration technologies (to bring the air quality data into the analysis platform)

Costs

The cost is incorporated in a subscription to the Breeze SaaS platform

Considerations

Since the analytical components come pre-built, the main considerations for a jurisdiction revolve around training. Breeze provides users with instruction on how to use the platform, how to run the analyses, and how to understand and interpret the results. Since users do not necessarily have visibility into the assumptions made by the algorithms in pre-built SaaS solutions, developing an understanding of how the analyses are being conducted is important to ensure that any potential risks are identified and accounted for. Beyond the pre-built capabilities, as various jurisdictions interacted with the analytics components of the Breeze platform in various ways, they began to identify additional needs that the data and analytical components could potentially satisfy. Through additional development work, the team at Breeze Technologies was able to extend the sensors and algorithms to be used for additional purposes, such as wildfire detection and pollen monitoring.

Skills

Aside from fairly basic training, users do not need technical knowledge to run and interpret the results of the AI models. In the building of the analytical components, however, the Breeze Technologies team made use of the following skills:

- AI/ML skills
- Data science skills
- Environmental engineering/ planning subject matter expertise

Stakeholders

The primary stakeholders in the data visualization step were Breeze Technologies, which developed the analytical capabilities within the platform, and Vilniaus Planas/ the City of Vilnius, which were the users of the platform. Additionally, Vilnius was able to benefit from enhancements suggested by other jurisdictions, which identified needs beyond what had previously been accounted for in developing the analyses.

STEP 6: DECISION-MAKING

Context

The City of Vilnius largely intends to use the air quality sensor data, as well as data collected from similar efforts, to foster a more collaborative relationship between the departments within the city, as well as between the city and its citizens. The broader goal is to support the creation of an "Intelligent City Brain," which allows the city to better prioritize its efforts, in order to have the biggest impact on the quality of life of its citizens, based on communicated needs.

Time

This kind of human-centered governance relies upon an ongoing, continual feedback and improvement process, in which the implementation is continuously evaluated with respect to the citizens' needs, and adjustments are made accordingly.

Considerations

Having well-defined performance metrics is key to ensuring the biggest impact from the interventions implemented. Dialogs and surveys with citizens are crucial for understanding citizen priorities and needs. Data-facilitated interdepartmental collaboration lends itself well to defining the best solution. AI is extremely helpful for unlocking insights previously unseen. Although not explicitly mentioned by Sodys or Breeze Technologies, strong data sharing and transparency frameworks are also highly relevant to this kind of work, as evidenced by both Vilnius' and Breeze's commitment to publicly available data.

Skills

The main skills involved with data-mediated human-centered governance, as utilized in this project, include:

- Qualitative methods skills (i.e. the ability to design and conduct surveys, as well as analyze the results)
- Stakeholder engagement techniques
- Citizen engagement techniques
- Strategic planning capabilities
- Urban subject matter expertise

Stakeholders

The main stakeholders in this step of the process are Vilniaus Planas/City of Vilnius and the citizens of Vilnius, which all engage in data-mediated collaboration to settle on solutions that improve the lives of residents and continue to do so.

KEY TAKEAWAYS

The Air Quality Monitoring use case in Vilnius demonstrates a fully out-of-the-box artificial intelligence project implementation. In this case study, most aspects of the process fell under the purview of an external service provider, including sensor installation, data storage, data processing, data visualization, and artificial intelligence integration. Taking an out-of-the-box approach allows municipalities of all sizes, budgets, and technical maturities to engage in urban artificial intelligence. Since the service provider handles most of the steps, the municipality only needs to manage and supervise the project. However, as the Vilnius case study highlights, using an out-of-the-box solution does not have to mean forfeiting agency and oversight. Municipalities still can (and should) take a collaborative approach in melding the solution to their specific contexts and the needs of their citizens, especially if they understand what is happening during each of the steps. Additionally, as Sodys and his colleagues did, municipalities can integrate the data and model results into existing data visualization and decision-making platforms, to ensure that the insights and analyses align with the goals and activities of other departments and sectors within the city.

CONCLUSION

Through the presentation of a literature review representing the current "state of the art" of urban artificial intelligence, we have envisioned a diagram of an anatomy that outlines and describes the key components of an urban artificial intelligence implementation. We have done this by highlighting three case studies that detail specific considerations for carrying out an urban artificial intelligence project in different urban contexts. This guide, therefore, has synthesized the research and practical insights of scholars, start-ups, and municipalities working at the forefront of artificial intelligence in cities.

As shown in the literature review, artificial intelligence has become an increasingly popular approach employed in urban contexts. An increased urgency around issues of sustainability and population growth, coupled with augmented technological capability introduced by the widespread dissemination of sensor systems and cloud computing/ big data solutions has allowed cities and other urban practitioners to begin to tackle their challenges in new and innovative ways. In parallel, citizens and public officials remain uneasy about the externalities and implications of artificial intelligence technologies, as well as privacy and security concerns associated with massive personal data collection. Scholarship has made great strides in proposing ethical and human-centered approaches to data stewardship and algorithm-augmented decision making. Yet, no practical guide had previously laid out the components and procedural considerations associated with implementing artificial intelligence in urban contexts, in a detailed but digestible manner.

By developing an anatomy of urban artificial intelligence, we broke down the concept into its constituent parts: Urban Infrastructures, Sensors and Data Collection Infrastructures, Network Infrastructures, Data Storage Infrastructures, Data Processing, Data Visualization, AI/ Machine Learning, and Decision-making/ Adaptation. Each section highlighted the physical and technical components required to complete a full artificial intelligence project in an urban context, as well as social, governance, and practical suggestions to better facilitate decision-making within each of those steps. Expert perspectives and insights enriched the anatomical content, particularly highlighting the variations in which each step can be adapted depending on the specific contexts of a given city. The urban AI anatomy intentionally omits "outcomes" as a final step, as urban artificial intelligence systems contain feedback loops, in which the decision-making affects the urban and social infrastructures of the city,

which in turn affect the emitted data, and so on and so forth. Urban AI does not stop at a neatly packaged solution, but rather continuously evolves as the city does.

The case studies demonstrate the practical application of the components detailed in the guide, highlighting the financial, temporal, technical, and political necessities that can help or hinder an AI project. We intentionally chose three case studies in different countries, focusing on cities of different sizes, involving different sectors, and at various stages of implementation and project maturity to emphasize the modularity and portability of artificial intelligence, as well as different approaches for obtaining buy-in. Any city, regardless of population, budget, and the state of its infrastructure can find a method for implementing artificial intelligence that fits within its confines, should it identify a need for such a solution. Additionally, the case studies demonstrate the fact that even though artificial intelligence often occupies its own discrete step in an artificial project implementation (as modeled in the anatomy of urban AI), it can easily permeate the other steps as well. In doing so, tasks like data collection or data processing can be automated, reducing the need for staff capacity within a team and making a solution more cost effective. Yet, city governments must still retain oversight of each step of the process, in the interest of ensuring a responsible implementation.

In order to evaluate artificial intelligence solutions and successfully implement AI projects in urban contexts, cities and other urban stakeholders should develop an understanding of the constituent technologies and their implications. With a baseline knowledge, as provided in this guide, technologists and decision-makers can more meaningfully engage with citizens and vendors to coproduce and adequately scope AI solutions. Doing so builds trust and ensures that system components align with project goals, which in turn align with public need. By understanding the urban AI system as a whole, decision-makers and technologists can better build transparency, communicability, and visibility into their data collection, data processing, and decision-making efforts, whether those steps are performed internally or by external vendors. Finally, understanding the implications of the technologies allows decision-makers and technologists to implement a solid framework for evaluation and adaptation of their artificial intelligence solutions. Since both cities and technologies continuously evolve, developing an understanding of urban artificial intelligence is not a discrete goal, but rather an ongoing effort to continue to learn and improve.

CALL FOR CONTRIBUTIONS

Urban AI is constantly evolving as technology progresses and urban centers face new challenges. For this reason, the Urban AI guide is a living document, which will be periodically updated to reflect the current state of the art. If you would like to share insights or describe a case study for the next edition, please fill out the form here: https://urbanai.fr/our-works/urban-ai-guide/

REFERENCES

Allam, Z. & Dhunny, Z. (2019). On big data, artificial intelligence and smart cities. *Cities*, 89. DOI: 10.1016/J.CITIES.2019.01.032

Anyoha, R. (2017). The History of Artificial Intelligence. Blog Science in the News. Harvard University. Retrieved October 28, 2022, from: https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Batty, M. (2018). Artificial intelligence and smart cities. *Environment and Planning* B, 15, 3–6. DOI: 10.1177/2399808317751169

Batty M. (2016). Big Data and the City. *Built Environment*, 42(3). DOI:10.2148/benv.42.3.321

Batty, M. (2009). Urban Modeling. In N. Thrift, & R. Kitchin (Eds.), International Encyclopedia of Human Geography (pp. 51-58). Oxford, UK: Elsevier. DOI: 10.1016/B978-008044910-4.01092-0

Beroche, H., Chubinidze, A., & Goelzer, L. (in press). *Geopolitics of Smart Cities: Expression of Soft Power and New Order.* Paris, France: Urban AI.

Beroche, H. (2021). Urban AI. Paris, France: Urban AI. Retrieved from: https://urbanai.fr/wp-content/uploads/2021/03/URBAN-AI-1.pdf

Beroche, H. (2022). Towards Sensible Interfaces. In Brandusescu, A. & Reia, J. (Eds.), *Artificial Intelligence in the City: Building Civic Engagement and Public Trust. Centre for Interdisciplinary Research on Montreal, McGill* University

Bostrom, N. (2017). Strategic Implications of Openness in AI Development. Global Policy, 8(2). 10.1111/1758-5899.12403

Burgess, A. (2018). *The Executive Guide to Artificial Intelligence.* Cham, Switzerland: Springer Nature. Retrieved from https://books.google.com/books/about/The_Executive_Guide_to_Artificial_Intell.html?id=avM-DwAAQBAJ&source=kp_book_description

Castillo, A., Karaiskou, A. & Gandhi., S. (June 16, 2022). Framing AI for Sustainable Cities. *Urban AI.* Retrieved from: https://medium.com/urban-ai/framing-ai-for-sustainable-cities-9723784dc97d

Cave, S., Dihal, K., & Dillon, S., (2020). *AI Narratives: A History of Imaginative Thinking about Intelligent Machines.* Oxford, UK: Oxford University Press.

City of Boston (2019). STREET BUMP. Retrieved from: https://www.boston.gov/transportation/street-bump

Consortium DATA PUBLICA - KPMG (2021). De la Smart City à la réalité des territoires connectés. French Ministry of the Economy, Finance, and Recovery. Retrieved from: https://joinup.ec.europa.eu/sites/default/files/event/attachment/2022-05/From%20Smart%20City%20to%20the%20Reality%20of%20Connected%20Territories_EN.pdf

Crawford, C. & Joler, V. (2018). Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources. *AI Now Institute and Share Lab.* Retrieved from https://anatomyof.ai

Cugurullo, F. (2020). Urban Artificial Intelligence: From Automation to Autonomy in the Smart City. *Frontiers in Sustainable Cities*, 2. DOI: 10.3389/frsc.2020.00038

DECODE (2023). What is DECODE? Retrieved from: https://decodeproject.eu/what-decode.html

Eckroth, J. (2018). *AI Blueprints: How to build and deploy AI business projects.* Birmingham, UK: Packt Publishing Ltd. Retrieved from: https://books.google.com/books/about/AI_Blueprints.html?id=sR2CDwAAQBAJ&source=kp_book_description

Faircloth, B., Connock, C., Welch, R., Elsworth, K., Escott, E. (2022). How We See Now: Traversing a Data-Mosaic. In: Chokhachian, A., Hensel, M.U., Perini, K. (eds) *Informed Urban Environments*. The Urban Book Series. Springer, Cham. DOI: 10.1007/978-3-031-03803-7_3

Fernandez-Anez, V., Fernandez-Guell, J. M., & Giffinger, R. (2018). Smart City implementation and discourses: An integrated conceptual model. The case of Vienna. *Cities*, 78. DOI: 10.1016/j.cities.2017.12.004

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., and Srikumar, M. (2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *Berkman Klein Center Research Publication No. 2020-1*. DOI: 10.2139/ssrn.3518482

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, 26. DOI: 10.1007/s11948-020-00213-5

Gordon, E. & Guarana, T. (2022). Solving for Trust: Innovations in Smart Urban Governance [white paper]. John S. and James L. Knight Foundation. Retrieved from: https://elabhome.blob.core.windows.net/elabpublication/SolvingForTrust_11-9-Web.pdf

Gordon, E. & Mugar, G. (2020). *Meaningful Inefficiencies: Designing for Public Value in an Age of Digital Expediency.* Oxford, UK: Oxford University Press. Retrieved from: https://books.google.com/books/about/Meaningful_Inefficiencies.html?id=-O_GDwAAQBAJ&source=kp_book_description

Jyoti, R., et al. (2021). IDC FutureScape: Worldwide Artificial Intelligence and Automation 2022 Predictions. *IDC FutureScape.* Revtrieved from https://www.idc.com/getdoc.jsp?containerId=US48298421

King, B. A., Hammond, T., Harrington, J. (2017). Disruptive technology: Economic consequences of artificial intelligence and the robotics revolution. *Journal of Strategic Innovation and Sustainability, 12(2)*, 53–67. DOI: 10.33423/jsis.v12i2.801

Koseki, S. et al. (2022). AI and Cities: Risks, Applications, and Governance. *UN Habitat and Mila.* Retrieved from: https://unhabitat.org/sites/default/files/2022/10/artificial_intelligence_and_cities_risks_applications_and_governance.pdf

Macrorie, R., Marvin, S. & While, A. (2019). Robotics and automation in the city: a research agenda. *Urban Geography*, 42(2). DOI: 10.1080/02723638.2019.1698868

Muller, M. et al. (2019). How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* DOI: 10.1145/3290605.3300356

Murphy, O. & Villaespesa, E. (2020). AI: A Museum Planning Toolkit. *The Museums + AI Network.* DOI: 10.13140/RG.2.2.33549.90085

Popelka, S. (July 18, 2022). AI for Urban Gamification. *Urban AI.* Retrieved from: https://medium.com/urban-ai/ai-for-urban-gamfication-debb3cf04a0a

Saltz, J. & Heckman, R. (2015). Big Data Science Education: A Case Study of a Project-Focused Introductory Course. *Themes in Science and Technology Education, 8(2).* Retrieved from: https://eric.ed.gov/?id=EJ1131005

Saran, N. (September 29, 2021). Towards Better Governance of Urban Data: Concrete Examples of Success. *Urban AI.* Retrieved from: https://medium.com/urban-ai/towards-better-governance-of-urban-data-concrete-examples-of-success-ade6472d427d

Scammell, R. (September 24, 2020). Half of UK data leaders say their company data risks ethnic bias. *Verdict.* Retrieved from: https://www.verdict.co.uk/ethnic-data-bias/

Schmoeker, A. & Lu, J. (April 14, 2021). Realizing our democratic values: unveiling the invisible in our shared spaces. *Urban AI.* Retrieved from: https://urbanai.fr/realizing-our-democratic-values-unveiling-the-invisible-in-our-shared-spaces/

Silva, E. & Wu, N. (2010). Surveying Models in Urban Land Studies. J*ournal of Planning Literature*, 27(2). DOI: 10.1177/0885412211430477

Tegmark, M. (2017). Life 3.0: *Being Human in the Age of Artificial Intelligence.* New York, USA: Knopf.

Tomitsch, M. & Haeusler, M. H. (2015). Infostructures: Towards a Complementary Approach for Solving Urban Challenges through Digital Technologies. *Journal of Urban Technology*, 22(3). DOI: 10.1080/10630732.2015.1040296

UN DESA. (2021). *The Sustainable Development Goals Report 2021.* New York, USA: UN DESA. Retrieved from https://unstats.un.org/sdgs/report/2021/

Verhulst, S., Young, A., & Sloane, M. (2021). The AI Localism Canvas. *Informationen zur Raumentwicklung*, 48(3). Retrieved from: https://ssrn.com/abstract=3946503

Warwick, K., & Shah, H. (2016). Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence, 28(6),* 989–1007. DOI: 10.1080/0952813X.2015.1055826

Woetzel, J., Remes, J., Boland, B., Lv, K., Sinha, S., Strube, G., Means, J., Law, J., Cadena, A., von de Tann, V. (2018) Smart Cities: Digital solutions for a more livable future. McKinsey Global Institute. Retrieved February 10, 2023, from: https://www.mckinsey.com/capabilities/operations/our-insights/smart-cities-digital-solutions-for-a-more-livable-future

Yigitcanlar, T. et al. (2021). Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(1). DOI: 10.3390/joitmc7010071

Yigitcanlar, T., Desouza, K., Butler, L., & Roozkhosh, F. (2020). Contributions and Risks of Artificial Intelligence (AI) in Building Smarter Cities: Insights from a Systematic Review of the Literature. *Energies*, 13(6). DOI: 10.3390/en13061473

Yigitcanlar, T. & Inkinen, T. (2019). *Geographies of Disruption: Place Making for Innovation in the Age of Knowledge Economy.* Cham, Switzerland: Springer Nature. Retrieved from https://link.springer.com/book/10.1007/978-3-030-03207-4

Yigitcanlar, T. & Kamruzzaman, M. D. (2018). Does smart city policy lead to sustainability of cities? *Land Use Policy*, 73. 49-58. DOI: 10.1016/j.landusepol.2018.01.034