I. Context and Motivation

The Global Water Crisis

The United Nations' April 2022 Summary Report demanded that water and sanitation be legally declared human rights that must be available to all,[1] including the 2.2 billion people who currently lack access to safely managed drinking water, and 4.2 billion people who currently lack safely managed sanitation.[2] The report, which was the product of a global consultation with stakeholders from 46 countries, stressed the importance of access to safe drinking water and sanitation in all aspects of life and organized the comprehensive justifications into three dimensions: social, economic, and environmental. Climate change and the coronavirus pandemic, according to the report, are exacerbating these impacts on the lives of billions of people around the world.

The Local Water Crisis

The Philippines has been suffering from a water crisis since 2000. Currently, around 7 million Filipinos (6% of the population) consume untreated surface water, and 24 million (21% of the population) lack access to basic sanitation facilities.[3] Ironically, more Filipinos own mobile phones than sanitary toilets, resulting in widespread open defecation.[4] This practice has been identified as the primary source of pollution and water contamination in Metro Manila's two largest bodies of water: the Pasig River and Manila Bay. In addition to domestic waste brought about by urbanization, rapid industrialization has been identified as another major reason why Manila Bay[5] and Pasig River[6] were deemed 'biologically dead', with industrial wastes from tanneries, textile mills, food processing plants, distilleries, chemical, and metal plants accounting for one-third of total pollution.[7]

II. Statement of the Problem

Manila Bay and Pasig River could have been sources of clean water for domestic and industrial use, means of transport, places of recreation and shelter, marine life habitats, and thriving ecosystems.[8] These bodies of water could have helped in alleviating the country's water crisis and improving the lives of the 13.8 million residents of the National Capital Region and its neighboring most populous region, Calabarzon.[9]

It cannot be assumed that the government lacked initiatives and interventions. The Manila Bay Rehabilitation Program was established three years ago,[10] whereas the Pasig River Rehabilitation Commission has been in operation for more than two decades.[11] The latter was decommissioned in 2019 and absorbed into the Manila Bay Task Force (MBTF), which is now in charge of restoring water quality in both bodies of water.[12] The task force is led by the Department of Environment and Natural Resources (DENR) and collaborates with 12 government agencies and quasi-governmental organizations, including water privatization, supply, treatment, and distribution utility companies like Maynilad Water Services, Inc. (Maynilad), Metropolitan Waterworks and Sewerage System (MWSS), and Local Water Utilities Administration (LWUA).[13] Apart from water quality improvement, rehabilitation, and resettlement, the MBTF was also tasked with ensuring sustained law enforcement and monitoring.[14] Despite this, no effective policy measures to address this environmental-social-health issue have been implemented, and the conditions of the two target water bodies continue to deteriorate.[15] [16]

Although water pollution costs the country about 1.3 billion dollars per year,[17] numerous researches have identified insufficient government funding as the main impediment to addressing the water quality problem in Manila Bay and Pasig River. This deficit in finances and focus resulted in poor implementation of projects;[18] thus, the goal of this data science project is to propose economically viable and easily implementable solutions to assist the Manila Bay Task Force in fulfilling its mandate to rehabilitate Manila Bay and the Pasig River.

III. Use Case of Data Science in Addressing the Problem

In developing countries, rapid industrialization, combined with limited resources, has contributed to an increased difficulty for monitoring agencies to identify sources of pollution. Studies have advocated the efficiency of developing a proper water quality monitoring system that can organize, aggregate, and share information across related agencies, leading to better quality enforcement. One such study by Kumpel et al.[19] was conducted among water quality agencies and institutions across six sub-Saharan African countries. The systems these agencies employed to collect and report information were assessed, and key hindrances to the efficient flow of information (and corresponding action) were found. These barriers were the lack of synthesis of the data collected, poor access to sharing of information among related agencies, and the lack of digital infrastructure to store, analyze and monitor the data collected.

Technological advancements such as artificial intelligence, machine learning, and robotics have already aided businesses across a wide range of industries. Water quality monitoring agencies are no exception, and a number of use cases involving data science have been explored in research throughout the globe. In Taiwan[20], miniature sensors with the Internet of Things (IoT) technology were stationed in 100 sites to monitor water quality in real-time, with the objective of preserving the Nankan River. The river was situated in Taoyuan, the largest industrial city in Taiwan. The project aimed to trace sources of pollution using geographical data of river paths and databases of factories and their characteristics. It also provided early warning systems for pollutant detection and determined hotspots of water pollution using clustering analysis and water quality prediction models. The system used in this research successfully reduced the scale of areas investigated as sources of pollution, improved government response, and optimized limited human manpower.

On a smaller scale, even leaner technologies such as machine learning models on their own can be useful for water quality monitoring in developing countries. In cases where there is no access to IoT technology, the water quality testing process can still be improved—an important step that is oftentimes a bottleneck. In a 2020 study[21], machine learning algorithms were used to classify the water quality index of water samples based on parameters such as pH, dissolved oxygen, biological oxygen demand (BOD), conductivity, fecal coliform, total coliform, and nitrate. The researchers used the Support Vector Machine (SVM), Naive Bayes, and K-nearest neighbors (KNN) models. Using the model with the highest performance, SVM, they were able to predict water quality classification with 97.01% accuracy. Prediction of future water quality using such methods would surely be useful for informing policy in developing countries.

IV. Proposed Data Products

We propose that the Manila Bay Task Force adopt and own the following data products:

Automated Pollutant Detection Model

To address the water pollution problem observed in Manila Bay and Pasig River, our team proposes to utilize data collected from IoT devices such as drones, and then apply different machine learning algorithms to the images collected by these devices to identify specific hotspots of polluted water with minimal human intervention. Current practices to evaluate compliance of different establishments to water quality standards would involve the cumbersome process of conducting an ocular inspection, which is followed by the laboratory testing of water samples collected from the drainages near these areas. However, conducting confirmatory laboratory tests could be very costly especially if there are a lot of water quality parameters that have to be evaluated. With machine learning, it is possible to achieve cost savings through the reduced dependency on laboratory tests to correctly determine which sites are hotspots of water pollution. Likewise, this tool would also help in accelerating the execution of different policy interventions needed to resolve the water quality problems observed in areas that flow water into the Manila Bay and Pasig River. This tool would help the MBTF strategically deploy their response teams that are in charge of clean-up drive efforts.

Water Quality Risk Mapping

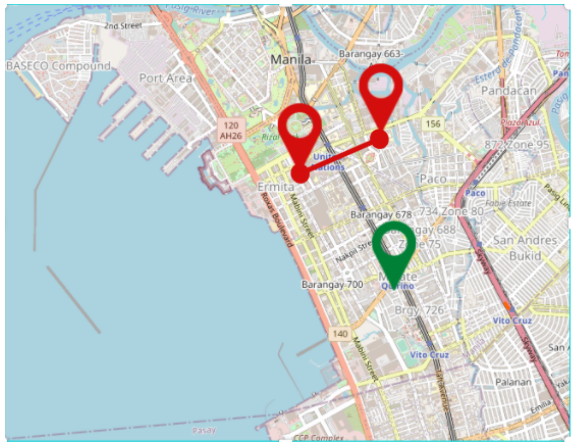

Figure 1. Water Quality Risk Map Prototype

Although the use of machine learning on image data helps in identifying whether water bodies found close to establishments are still compliant with water quality standards or not, this data product does not make predictions towards areas that have not yet been inspected but deem to have similar characteristics of establishments that were already established to be non-compliant to water quality regulations. With this, the team would also like to implement a water quality risk-mapping tool that would help in identifying which particular areas in Metro Manila (that were not inspected yet) should be prioritized for an ocular inspection and water sample testing. To be able to implement this, the team would need access to geospatial data that outlines the number of establishments that are nearby a particular area. Likewise, a sewer network map would be helpful for the policymakers to identify how water flows from the different drainage pipes found near establishments up to the water bodies. The sewer map would help identify potential hubs where most discharge water from establishments tends to flow. By being able to detect these central hubs in the sewer network, MBTF would be able to identify new possible areas that they must constantly monitor in order to preserve the quality of water that flows into Manila Bay and Pasig River.

Centralized Decision Support Tool

To be able to fully utilize the data and insights generated, our data products will be combined into a centralized decision support tool. This tool will display a dashboard of insights generated from the deployed models, key performance indicators the agency may need to monitor and report to external entities, and a user interface to record collected data more efficiently. This tool will also be linked to the additional data collected through drones and site visits. With this tool, MBTF will be able to optimize their limited resources and manpower by spending fewer labor hours on extensively manual work. Through the integration of the results generated by the prior data products proposed, policymakers would be able to gain a bigger picture of how close or far they are in achieving their water body rehabilitation targets. Likewise, the centralization of water quality monitoring efforts would be helpful in the enforcement of policies targeted to preserve the cleanliness of the Manila Bay and Pasig River.

V. Socio-Technological Systems Framework

Technical Characteristics

Monitoring and detection can also be performed using drone shots or through CCTV cameras above exposed water bodies. This solution replicates the visual inspection approach without the need to travel physically to the location. Alternatively, water quality detection can be performed using IoT devices. Microsoft AI Lab's Clean Water AI has developed a single prototype that costs below $500.[22,23] The device can identify if the water sample is contaminated or not as illustrated in Figure 2 below. The prototype includes an IoT sensing application and an edge device that can analyze a microscopic image of the water. Additionally, it can host a machine learning model that can detect water quality in real-time. The information can then be transmitted back and used for appropriate immediate action and policy implementation.

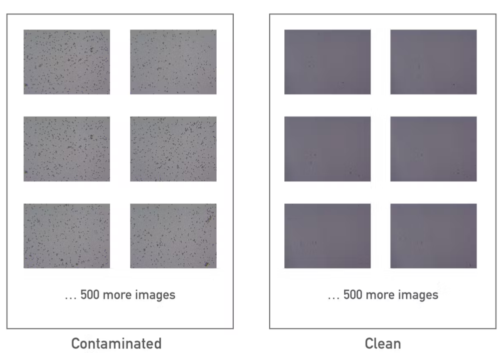

Figure 2. Sampled contaminated and clean water images from Clean Water AI

Social Dimension

The drone and CCTV images captured can be misconstrued as encroaching on the privacy of individuals that pass by the field of vision of the camera. Reactive policies addressed to non-compliant locations would have a Big Brother feel, particularly if sanctions are imposed on specific individuals or establishments identified through the video feed. The alternative IoT solution avoids privacy concerns as it only captures microscopic images and would be installed near the source of the water. However, there is still a remote possibility that it could be a way to gather private information from nearby establishments.

Institutional Context

The IoT device solution lacks post-detection action beyond the financial investment required for the prototype. The $500 seed investment (or roughly equivalent to PHP 25,000) for a single location is too steep of a price. It remains to be seen if a single-use testing service using chemicals that costs around $10[24] in developed countries would be more cost-efficient. For urban Metro Manila, allocating the budget for labor that would perform cleaning activities and education programs for efficient waste management instead can be the clearer optimal choice.

Adoption and Use

The solution can be more impactful if both the financial and technical barriers to using the data products are reduced as these are not financially feasible for both residential and small business use. Other than government adoption, another option is to have private companies include the use of these devices as part of their respective CAPEX investments (amounts to PHP 13.7 Billion[25] for MWSS) with the goal of reducing the long-term overhead in monitoring the water quality in their treatment plants and respective zone jurisdictions. Should it yield beneficial results, then the data product would be promoted organically.

VI. Resources Needed

For this solution to make a positive impact on the management of the water quality, sufficient resources must be allocated to support the implementation of this tool. The Philippines is not exempt from access issues such as the digital divide, and a number of the key parties involved in the project may not have equal access to necessary technologies, made even more pronounced due to the COVID-19 pandemic.

Infrastructure

To make full use of the data products, especially those designed to be used during fieldwork, agency employees must be provided with computers and mobile phones that meet the required specifications to run the computations and access the mobile dashboard. They should also be provided with WiFi devices that can enable access to the geospatial visualization while conducting site visits. Apart from user-specific devices, investments in IoT technology in the form of CCTV or drones will also be required to supply data for the machine learning algorithms. Access to geospatial data across regions of responsibility including sewage networks will also be necessary for this type of analysis.

Generic Services

Due to the difficulty of financing in-house databases in developing countries, we also recommend renting a cloud database storage such as Amazon Web Services to house the needed data. Techniques such as parallel processing can also be used to make up for the lack of computational power of individual machines. Similarly, if in-house data scientists and data engineers are not available, the agency championing this product can outsource the necessary tasks to consulting agencies. These agencies will be responsible for the proper design of the models, integration of all data products, and structuring of the database. Making use of services in this way can allow infrastructure demand to scale together with the steady increase of operations.

Capacities and Knowledge Skills

Lastly, even if the major components are outsourced, the agency involved must also update the capacities and skills of their employees in order for them to understand how to make the most of this product. Data literacy training can familiarize employees with the mechanism behind the model's generated results, and when combined with their innate domain expertise, may unlock opportunities for more in-depth insights.

VII. Limitations

Financial Aspect

To be able to implement this product, the agency must first and foremost address the financial feasibility of the digital transformation program. Although this proposed solution is designed to be more economically efficient, it is not without cost. The new total budget allocated to the rehabilitation of Manila Bay is PHP 1.56 billion,[26] and proper allocation of this funding must be done to ensure there are enough finances allocated to undertake a digital transformation. There must be a sufficient budget to support capital expenditures such as the cost of database storage, processing and computing capabilities, the use of drones to capture images, outsourcing of data science and data engineering tasks, and employee skills training. Ultimately, proceeding with this solution is determined by whether or not the payback period is expected to be realized as soon as possible. If financial projections from the agency result in a positive net present value over their desired time frame, they may choose to proceed with the endeavor.

Employee Acceptance

Another limitation that must be considered is the willingness of agency employees to adopt this thrust at digital transformation. Employees may be aware of advances in technology causing certain jobs to become obsolete, and this may lead to resistance stemming from a fear of unemployment. Apart from this, resistance to new technology may also be ingrained in the culture of government agencies, possibly caused by a lack of exposure to AI technologies, which are primarily seated in the private sector. It is, therefore, necessary to conduct employee reskilling and upskilling programs to keep their roles significant, and continuously promote an agency-wide culture change towards digitization and innovation.

Resident Participation

The residents surrounding the water bodies to be monitored must also be taken into account in the deployment of this solution. For instance, around 233,000 informal settlers reside around the Manila Bay area,[27] leading to the discharge of waste into the waterbody. Households remain a source of pollution, and awareness must be raised in communities about the initiatives being introduced and how they stand to benefit. Communication with the residents must remain open and collaborative to ensure all key stakeholders are taken into account, securing continued support for this project.

VIII. Ethical Considerations

The analysis that follows is based on the five core principles for ethics in artificial intelligence derived from bioethics: Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability.

It is safe to claim that the entire project complied with the Beneficence principle as the overarching goal is to advance SDG6 - ensure availability and sustainable management of water and sanitation for all.[28] The proposed data products are deemed to benefit a large population, while also contributing to environmental sustainability. When the Non-Maleficence and Autonomy principles are considered, however, a point of contention is the potential infringement of privacy in terms of data gathering. The use of IoT devices to detect pollutants in real-time may jeopardize the privacy of 4.4 million residents of the Pasig River catchment area[29] and 2.8 million informal settlers living along Manila Bay's waterways.[30] To avoid this, the MBTF must obtain clear and informed consent from residents and provide them with the option of opting out of the data collection process. The scope of data use and handling should also be clearly communicated. Furthermore, the data science team promotes transparency and also encourages the task force to be truthful about the methodologies and algorithms employed in the proposed solutions. The *hows* and *whys* of the project should be specified and explained to all stakeholders.

The proposed data products are unlikely to exacerbate the already existing inequalities in the informal settlements. However, the concern lies in how the MBTF decision-makers will utilize and integrate the generated insights in carrying out their mandates and in crafting policies, particularly the sanctions on non-compliant residents and industrial companies. This ethical and legal analysis is limited to the proposed outputs and does not extend to policy-making.

IX. References

1 United Nations. (2022). Global Online Stakeholder Consultation: Themes for Interactive. Dialogues, Summary Report. *United Nations 2023 Water Conference*. https://sdgs.un.org/sites/default/files/2022-04/

2 United Nations. (2021). Water and Sanitation. https://sdgs.un.org/topics/water-and-sanitation.

3 Samion, Nur Khairiyah Binte Mohd. (2020). The Water Crisis: The Philippines. *Wateroam.* https://www.wateroam.com/social-awareness/the-water-crisis-philippines.

4 Habito, Cielito F. (2020). More basic than cell phones. *Inquirer.net*. https://opinion.inquirer.net/127989/more-basic-than-cell-phones.

5 Chaigne, Thai. (2021). Plastic waste chokes last mangrove forests in Philippines' Manila Bay. *The Observers.* https://observers.france24.com/en/asia-pacific/20211011-plastic-pollution-mangrove-forest-philippines-manila-bay.

6 Gorme, J.B., Maniquiz, M.C., Song, P., Kim, L. (2010). The Water Quality of the Pasig River in the City of Manila, Philippines: Current Status, Management and Future Recovery. *Korean Society of Environmental Engineering*. http://www.kpubs.org/article/articleMain.kpubs?articleANo=.

7 Baclig, C. (2021). Pasig River makes international waves despite being dead. *Inquirer.net*.

https://newsinfo.inquirer.net/1446482/pasig-river-makes-international-waves-despite-being-dead.

8 Gorme, J.B., Maniquiz, M.C., Song, P., Kim, L. (2010). The Water Quality of the Pasig River in the City of Manila, Philippines: Current Status, Management and Future Recovery. *Korean Society of Environmental Engineering*. http://www.kpubs.org/article/articleMain.kpubs?articleANo=.

9 Philippine Statistics Authority. (2021). Highlights of the Philippine Population 2020 Census of Population and Housing (2020 CPH). https://psa.gov.ph/content/highlights-philippine-.

10 Department of Environment and Natural Resources. (2016). Fact Sheet on Manila Bay Rehabilitation. https://denr.gov.ph/index.php/2-uncategorised/853-compendium-of-enr-statistics-2016.

11 Gorme, J.B., Maniquiz, M.C., Song, P., Kim, L. (2010). The Water Quality of the Pasig River in the City of Manila, Philippines: Current Status, Management and Future Recovery. *Korean Society of Environmental Engineering*. http://www.kpubs.org/article/articleMain.kpubs?articleANo=.

12 Cabico, G. (2021). DENR: More action needed but state of Pasig River 'a lot better now'. *Philstar Global*. https://www.philstar.com/headlines/2021/06/28/2108692/denr-more-action-needed-.

13 Office of the President. (2019). Administrative Order No. 16. *Official Gazette*. https://www.officialgazette.gov.ph/downloads/2019/02feb/20190219-AO-16-RRD.pdf.

14 Department of Environment and Natural Resources. (2016). Fact Sheet on Manila Bay Rehabilitation. https://denr.gov.ph/index.php/2-uncategorised/853-compendium-of-enr-statistics-2016.

15 Cabico, G. (2021). DENR: More action needed but state of Pasig River 'a lot better now'. *Philstar Global*. https://www.philstar.com/headlines/2021/06/28/2108692/denr-more-action-needed-.

16 Baclig, C. (2021). Pasig River makes international waves despite being dead. *Inquirer.net*. https://newsinfo.inquirer.net/1446482/pasig-river-makes-international-waves-despite-being-dead.

17 Marrone, Rhonda. (2016). Water Pollution in the Philippines: Causes and Solutions. *Borgen Magazine*. https://www.borgenmagazine.com/water-pollution-in-the-philippines/.

18 Gorme, J.B., Maniquiz, M.C., Song, P., Kim, L. (2010). The Water Quality of the Pasig River in the City of Manila, Philippines: Current Status, Management and Future Recovery. *Korean Society of Environmental Engineering*. http://www.kpubs.org/article/articleMain.kpubs?articleANo=

19 Kumpel, E., MacLeod, C., Stuart, K., Cock-Esteb, A., Khush, R., & Peletz, R. (2020).

From data to decisions: understanding information flows within regulatory water quality monitoring programs. *NPJ Clean Water*, *3*(1). https://www.nature.com/articles/s41545-020.

20 Chen, F. L., Yang, B. C., Peng, S. Y., & Lin, T. C. (2020). Applying a deployment strategy and data analysis model for water quality continuous monitoring and management. *International Journal of Distributed Sensor Networks*, *16*(6), 155014772092982. https://doi.org/10.1177/1550147720929825.

21 Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water Quality Prediction Using Artificial Intelligence Algorithms. *Applied Bionics and Biomechanics*, 2020, 1–12. https://doi.org/10.1155/2020/6659314.

22 Clean Water AI. (Accessed April 2022). *Clean Water AI Official Website*. https://cleanwaterai.com/.

23 Ma, P., Shenk, J., Hong S.A., Han, S. (2018). Clean Water AI. *Hackster.io*. https://www.hackster.io/clean-water-ai/clean-water-ai-e40806.

24 Daphtary, N. (2021). Cost of Well Water Testing in 2021. *Chemtech International*. https://chemtech-us.com/articles/cost-of-well-water-testing-in-2021/.

25 Manila Water Company, Inc. (2022). Manila Water Hits Record 13.7B Capex for 2021. *Manila Water Official Website.* https://www.manilawater.com/customer/agos/2022-03-16/manila-water-hits-record-13-7b-capex-for-2021.

26 Quismoro, Ellson. (2021). Manila Bay task force ready to strike vs. polluters on March 22, World Water Day. *Manila Bulletin*. https://mb.com.ph/2021/03/16/manila-bay-task-force-ready-to-.

27 Department of Environment and Natural Resources. (2016). Fact Sheet on Manila Bay Rehabilitation. https://denr.gov.ph/index.php/2-uncategorised/853-compendium-of-enr-statistics-2016.

28 United Nations. (2015). Goal 6. https://sdgs.un.org/goals/goal6.

29 Baclig, C. (2021). Pasig River makes international waves despite being dead. *Inquirer.net*. https://newsinfo.inquirer.net/1446482/pasig-river-makes-international-waves-despite-being-dead.

30 Luna, Franco. (2019). Do informal settlers really contribute most of the waste in Manila Bay? *Philstar Global*. https://www.philstar.com/headlines/2019/11/11/1967913/Do-informal-settlers-really-contribute-most-waste-manila-bay.