Using Computer Vision to Detect Contaminated Medical Waste in Hospital Operating Rooms
(Technical)

The Social and Ethical Implications of Climate-Oriented AI Applications: A Comparative Study of the U.S. and India
(STS)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Sonali Luthar
Fall, 2020
Department of Computer Science

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature  Date_____

Approved_____ Date_____
Madhur Behl, Department of Computer Science

Approved_____ Date_____

Sharon Tsai-Hsuan Ku, Department of Engineering and Society

Introduction

Microsoft's AI for Earth is an initiative which fosters artificial intelligence and cloud solutions for global environmental challenges. Climate Change AI is a volunteer organization comprised of academics and industry leaders who believe in machine learning's impactful role in tackling climate change. AI for Social Good is a program by Google which focuses on utilizing their AI expertise to solve humanitarian and environmental challenges.

The message is clear. AI is increasingly connected to sustainable issues, and yet research in this area of intersection is still both innovative and novel. Deeply passionate about both of these subjects, I believe it is crucial to examine how they may best interact with each other. Hence, I explore this topic further in both my technical and STS research.

My technical project focuses on material waste, termed the Internet of Wasted Things. First, we have created the system SCRAP which uses computer vision to detect recyclable objects at the point of disposal, then guides the user to the correct waste bin with LED lights. Now, we seek to tackle medical waste in hospital operating rooms by again using computer vision to differentiate between used and unused medical tools and products.

The STS portion examines AI applications overall in global sustainability by analyzing the state of artificial intelligence in climate change, its limits and opportunities, and any potential humanitarian costs and benefits. This is done through a comparative study in the U.S. and India where the countries' starkly different approaches will provide insight into mobilizing AI for combating climate change. The ethical and social implications for each country will then be evaluated through biases existing in technological solutions and government regulations, and also by who is at risk for experiencing any harms. I will conclude by offering a glimpse into the probable future and highlight avenues requiring the greatest need of work or change.

Technical Project

The proposed technical project is a continuation of many semesters of work on an idea termed the Internet of Wasted Things (IoWT). Currently, recyclable waste management in the U.S. typically occurs downstream after waste has already been collected and commingled. Sorting high volumes of waste at this stage is expensive, slow, and manual. Similar to how energy and water are now often IoT-monitored in building spaces, the IoWT seeks to bring waste management to the 21st century through the integration of AI and IoT components for improved detection, classification, auditing, and reduction of material waste.

Thus far, we have created SCRAP, or the System for Classifying Recyclables through Automated Process. This closed-loop automated system improves categorization of recyclable waste items disposed by an individual by using computer vision to detect the waste object in hand and then lighting up the correct recycling bin with LED lights. We use low-cost cameras to scale our project for real-time detection and guidance. Our models were trained using custom data on the YOLOv3 (Redmon & Farhadi, 2018) architecture, with our best mean average precision being for Model 2 at 93.77%. Figure 1 shows the output detection of the baseline YOLO model and our trained Model 1 and Model 2 on images from the dataset across our three categories of interest: cup, bottle, and paper. As seen in the figure, the trained models can confidently detect each object type, whereas the baseline model cannot. We decided to make Model 2 our final model for the SCRAP system.

For my capstone technical project in Spring, 2020, I will build upon my work in creating SCRAP and expand the breadth of the IoWT umbrella. We plan to collaborate with colleagues from the Medical Center at UVA with the goal of adopting computer vision detection to classify used and unused surgical tools in hospital operating rooms (ORs). The motivation for this project

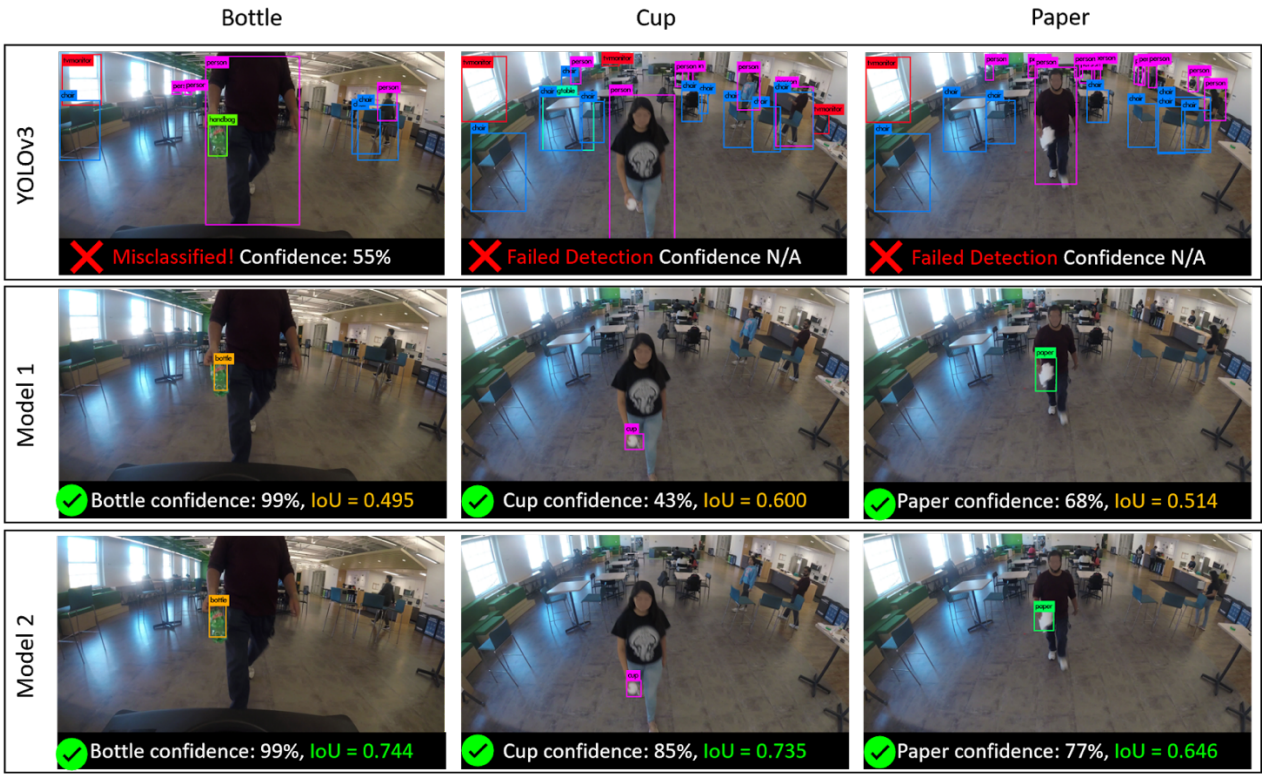

Figure 1. Examples of detection with output label, confidence, and IoU for each category.

is similar to that of SCRAP; that is, much material is wasted during operations through opening tools or other medical material and not using that material during surgery. We hope to audit the waste being generated from ORs using AI. With this data, we have the potential to (1) provide recommendations to hospitals on how best to educate surgeons and attending nurses to reduce their medical waste generated in surgery by targeting the items most wasted, (2) introduce cost-savings for a medical center by reducing the amount of often expensive material that is used, and (3) begin sustainable waste management practices within hospital spaces by reducing the amount of material waste generated.

Our expected deliverable at the end of the semester is still in the works as we navigate what is possible in a clinical space given the COVID-19 situation. At minimum we hope to implement the same visual detection as seen in SCRAP but for used and unused tools in an OR.

STS Prospectus

Introduction

Uncovered social, ethical, and technical potentials lie at the intersection of the booming artificial intelligence industry and the dire need for solutions to the climate crisis. A look at both the U.S. and India will provide a comprehensive view of those potentials. The U.S. has long been technologically powerful and India is on the rise. Each country's unique political, financial, and social factors have bred differing perspectives on climate change, technology, and ethical concerns. A comparison of their states will prove useful in assessing AI and climate overall.

As excitement around AI grows, its power is tested against pervasive societal issues like healthcare, education, transportation, energy, and more. Growth is inevitable as engineers uncover the powerful applications of AI; however, it is then the duty of developers to proceed with caution. Climate change is a complex, multifaceted issue, and innovative solutions including AI are absolutely necessary. However, treading new waters necessitates a thorough and sincere evaluation of the creation and ethics of the new AI to yield the most productive and benevolent results. The literature review presented demonstrates the novelty of the study of AI's application to climate initiatives and resulting ethical implications.

Research Questions

1. What are the technical roles, limits, and potentials of AI in climate solutions?
2. What is the state of AI and climate separately in both India and the U.S. and what is the state of their intersection?
    a. What government regulations, proposals, or goals are in place?
    b. What national initiatives are currently underway?

3. What creates the discrepancies between these two countries' systems?
    a. What is the current and near-future expected infrastructure?
    b. Who are the system builders and stakeholders?
    c. What are the reverse salient?
    d. How great is the momentum?
4. What are the ethical and social implications of AI providing solutions to climate change in each country?
    a. Who is helped and harmed and by what means?
    b. Are AI applications designed "ethically" or could/should improvements be made?
5. What is the expected future for each country and what social and ethical implications does that carry?
    a. What factors does this future depend on?

Literature Review

Much research is written individually on ethical AI, climate change and AI, and sustainability ethics; however, little to none is written on these areas intertwined, especially in comparing the U.S. and India. My research will therefore be a novel inquiry in this field.

The AI and climate state of the U.S.

Ghaffary (2019) in Vox reflects on the executive order "American AI Initiative" released under President Trump in February, 2019. The article reviews the order's goals and weighs strengths against criticisms. For example, the author claims the order "acknowledges the need for the government to ensure that civil liberties are protected in the use of AI," but doesn't address these concerns directly. This piece and others give insight into America's state of AI. A common

theme is fear our lead is slipping to countries with comprehensive plans like China (Bloomberg Opinion Editorial Board 2019).

However, these pieces fail to critically examine how political states of AI affect the technology created. My research will examine how the infrastructure and politics of AI in the U.S. either lends or does not lend itself to generating creative climate solutions.

The State of Climate Tech 2020 by PwC claimed investment in climate-related technology is a "new frontier of venture" by considering the markets and challenges of various climate sectors. It succeeded in displaying monetary impacts and opportunities, but did not draw any hard connection between climate and AI; it even compared climate tech to the AI boom without suggesting climate tech may be a subset of the AI revolution.

The AI and climate state of India

Artificial intelligence is a growing field in India. Analytics AI Magazine (Thomas, 2020) reported their study with Jigsaw Academy on the AI market in India and provided:

1. The AI market value in India at $6.4 billion.
2. A breakdown for AI market size based on company type.
3. A breakdown for AI market size based on industry, including retail, telecom, banking, energy, etc.

Although this report effectively defined financial stakes of AI in India, specific sustainable endeavors are not highlighted.

Vempati (2016), a digital strategist and political commentator, argued AI-ready infrastructure is critical for national security and preparedness. Her final recommendations included the necessity to adopt a deliberate national policy to drive AI innovation and the need to

identify public sector applications where AI could make significant impact. Again, this paper failed to identify trends between AI and climate change and focused instead on national security.

Connecting AI to climate initiatives is crucial due to India's dire state. A commissioned research report by the U.S. National Intelligence Council adequately summarizes: "India is both a major greenhouse gas emitter and one of the most vulnerable countries in the world to projected climate change. The country is already experiencing changes in climate and the impacts of climate change, including water stress, heat waves and drought, severe storms and flooding, and associated negative consequences on health and livelihoods" (National Intelligence Council, 2009).

There exists a lack of discussion into the full extent of connection between AI and climate in India. A national strategy report (NITI Aayog, 2018) details AI's potential intervention into areas such as smart cities and infrastructure and smart transportation. Before this vision is realized, AI and climate must be thoroughly examined, including potential impacts to ensure solutions are truly beneficial.

"Ethical" AI

Literature on ethical AI includes no shortage of definitions, propositions, and contentions. For example, the High-Level Expert Group on Artificial Intelligence set up by the European Commission created the *Ethics Guidelines for Trustworthy AI* (Hleg, 2019) which laid a framework of trustworthy AI consisting of three components: lawful AI, ethical AI, and robust AI. They define four ethical principles of AI, (1) respect for human autonomy, (2) prevention of harm, (3) fairness, and (4) explicability, and also detail seven key requirements needed for the realization of trustworthy AI.

In contention, Ryan (2020) asserted AI cannot be "trustworthy" because it does not have capacity for emotions or self-responsibility; instead, the term “reliable” AI shifts ethical responsibility onto developers.

A concept similar to ethical AI is AI for social good, gaining popularity with the most obvious example being Google's AI for Social Good initiative. Floridi et al. (2020) outlined seven essential factors for a successful AI and succeeded in highlighting specific areas prone to unethical use, but the advice is unspecific. For example, the paper advises "designers should adopt safeguards which ensure that non-causal indicators [in data] do not inappropriately skew interventions" but does not explain how that may be accomplished.

Ethical AI is discussed broadly in literature; in my research, I seek to apply the overarching principles of these frameworks to the field of climate action. I hope to uncover unique cautions in climate and outline specific measures for avoiding them.

AI and Climate Change

Rolnick, et. al. (2019) give an extremely comprehensive analysis of AI potentials for climate action. Their audience expands to researchers and engineers, entrepreneurs and investors, corporate leaders, and local and national governments. They have provided a stunning assessment across thirteen climate solution domains, along with subdomains, relevant to specific machine learning techniques, such as computer vision, NLP, causal inference, and more.

A thorough examination of this paper would provide substantial data for my first research question. The only way this paper falls short is in not continuing the discussion, i.e. evaluating context of political environments and assessing social benefit and cost. They also do not discuss ongoing initiatives in these fields. I hope to accomplish both in my own research paper.

Climate Change and Humanitarian Concerns

Some populations are much more vulnerable to the most brutal effects of climate change. Marina and Ribot (2012) explore how climate interventions too may prove harmful to vulnerable populations. The conclusion states "climate adaptation and mitigation often fall short of promoting environmental justice – in the form of rights, recourse and representation – that might make these interventions more locally relevant, equitable and therefore sustainable." They make valid points on the precarious nature of creating solutions to a difficult issue like climate change, but do not focus on any specific solution type. They caution, "researchers must understand the conditions under which climate change interventions are both created and implemented – to ensure that the goals of burden sharing, fairness in adaptation, representation in decision making, and the promotion of social justice and wellbeing remain priorities at all scales of climate change negotiations." In my own research I hope to do a similar analysis but on the potential effects of AI-climate applications specifically to vulnerable populations.

STS Framework

I will conduct my research using Thomas Hughes' Large Technical Systems (LTS) framework. This framework views technological revolution as an evolution resultant of existing momentum. I would like to evaluate momentum differences in both India and the U.S. through factors that created them, i.e. the interactive components, system builders, and reverse salient. I suspect relevant components will include the state of the environment and how soon solutions must be realized; the technological abilities of each community; the political leanings of each government; the voice of marginalized communities susceptible to harm; and corporations who create technological solutions and/or stand to gain monetarily in some fashion. System builders

are those in charge of each of these components. Reverse salient may include the limits of AI technology and the laws governing both tech and climate.

Methods for Data Collection

I will collect data through study of relevant, reputable documents and literature, namely:

1. Research and publications from Partnership on AI, including *The Role of Demographic Data in Addressing Algorithmic Bias* and *Human-AI Collaboration Trust Literature Review: Key Insights and Bibliography*.
2. National and international climate reports and goals, such as IPCC reports, UN Climate Reports, The U.S. National Climate Assessments, etc.
3. National policies on AI for the United States and India.
4. Resources on AI and social good, such as Google's AI for Social Good and the paper *Applying artificial intelligence for social good* (Chui, et. al., 2019).
5. Resources on AI and climate change, such as Climate Change AI (Rolnick, et. al., 2019).

Data collection would lend quantitative measurements of states of AI and climate in the U.S. and India, such as the dollar value of AI initiatives, climate risk assessments, existing infrastructure for technological climate solutions, etc. Then, each country's national goals will provide quantitative and qualitative data. Furthermore, I expect to research qualitative examples of AI applications to climate and resulting quantitative measurements of AI's role in climate change. Lastly, I will find both qualitative and quantitative means for evaluating ethical AI.

Assessing the bias present in my research is a core component of reaching a conclusion because I seek to ultimately focus on social and ethical implications. Government documents will contain political bias, and I am interested in its influence. As I am researching, I will also

continuously ask who is present in the discussion of this research. For example, are communities who will be affected most included when creating the technology?

I will also tackle bias by conducting interviews as necessary at UVA to supplement my research, such as with staff members from the Office for Sustainability or the Recycling Department, with professors working on sustainable AI initiatives, or with professors researching the humanitarian elements of climate change.

Timeline

Feb, 2021: Complete all literature research.

March, 2021: Complete any necessary interviews.

April, 2021: Coalesce research into a comprehensive thesis.

May, 2021: Apply finishing touches.

Conclusion

This STS research is expected to be a novel examination using the LTS framework into the U.S.'s and India's current state of AI, climate, and ethics. It is a necessary inquiry with which we gain a clear picture of what the near future holds. Literature is plentiful and comprehensive in each individual area of concern, but the goal of my research is to combine insights from all of these areas of study into a complete picture of what lies ahead. It will highlight areas of both concern and promise and ultimately provide greater insight into how we may traverse the path to a technologically advanced, climate-safe, and equitable future.

References:

AI for Earth. (n.d.). Retrieved December 04, 2020, from https://www.microsoft.com/en-us/ai/ai-for-earth

AI for Social Good. (n.d.). Retrieved November 02, 2020, from https://ai.google/social-good/

Bloomberg Opinion Editorial Board. (2019, February 15). Trump's AI Plan Is a Good Start. Retrieved December 04, 2020, from https://www.bloomberg.com/opinion/articles/2019-02-15/artificial-intelligence-trump-s-smart-american-ai-initiative

Chui, M., Harrysson, M., Manyika, J., Roberts, R., Chung, R., Nel, P., & Heteren, A. (2019, November 20). Applying artificial intelligence for social good. Retrieved December 05, 2020, from https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good

Floridi, L., Cowls, J., King, T.C. *et al.* (2020). How to Design AI for Social Good: Seven Essential Factors. *Sci Eng Ethics* **26,** 1771–1796. https://doi-org.proxy01.its.virginia.edu/10.1007/s11948-020-00213-5

Ghaffary, S. (2019, February 13). Trump's executive order on AI, explained. Retrieved December 04, 2020, from https://www.vox.com/2019/2/13/18222433/trump-executive-order-ai-explained

Hleg, A. I. (2019). Ethics guidelines for trustworthy AI. *B-1049 Brussels*.

Joint Global Change Research Institute, Battelle Memorial Institute, Pacific Northwest Division. (2009). India: The Impact of Climate Change to 2030. *National Intelligence Council*.

Marino, E., Ribot, J. (2012). Special Issue Introduction: Adding insult to injury: Climate change and the inequities of climate intervention. *Global Environmental Change*. https://doi.org/10.1016/j.gloenvcha.2012.03.001.

NITI Aayog. (2018). National Strategy for Artificial Intelligence. *Author*.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv*.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., ... & Luccioni, A. (2019). Tackling Climate Change with Machine Learning. *arXiv preprint arXiv:1906.05433*.

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Sci Eng Ethics* **26,** 2749–2767 https://doi-org.proxy01.its.virginia.edu/10.1007/s11948-020-00228-y

The State of Climate Tech 2020. (2020). PwC. Retrieved December 04, 2020, from https://www.pwc.com/gx/en/services/sustainability/publications/state-of-climate-tech-2020.html

Thomas, S. (2020, November 08). Report: State of Artificial Intelligence in India - 2020. Retrieved November 22, 2020, from https://analyticsindiamag.com/report-state-of-artificial-intelligence-in-india-2020/

Vempati, S. S. (2016). India and the artificial intelligence revolution. *Carnegie Endowment for International Peace*.