Yen-Chia Hsu | Research Statement

I co-design technology with stakeholders, especially citizens, to address sustainable development challenges. Citizens have solid connections to the local context and a desire to advocate for changes. **I study how technology can be used in decision-making to empower local communities with the lens of citizen science (engaging the public in research) and scientific citizenship (making science socially responsible).** I have collaborated extensively with citizens in Pittsburgh (Pennsylvania, USA) to co-design technology in addressing industrial air pollution. My work creates the link between technology and a problem of social importance through conducting practice-based research that has strong promise for social impact. My research process is dynamic and iterative. I see myself as a design engineer who applies various techniques to tackle civic concerns raised by citizens or decision-makers. The process involves a mix of problem seeking (similar to the one used by architects and designers), field deployment of working technology, documentation of the co-design process (using ethnographic methods), and evaluation of social impact (using case reports, surveys, and user behavior analysis). In this sense, I am interested in building new technology that fits into the local context and studying its impact on stakeholders. My work is often interdisciplinary, where I frequently collaborate with no-profit organizations, citizen groups, domain experts, and decision-makers.

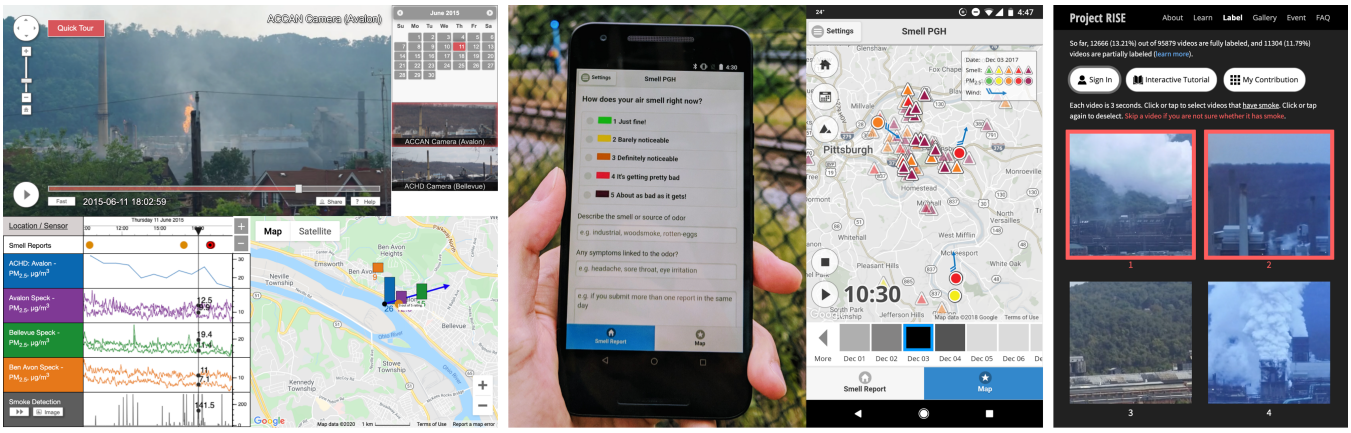

Figure 1: My works in: (a) air quality monitoring [2], (b) smell reporting [1], and (c) industrial smoke annotation [3].

I conduct field research to tackle citizen concerns while simultaneously making scientific contributions to Human-Computer Interaction (HCI) and Artificial Intelligence (AI). Regarding social impact, my air quality work has urged regulators in the United States Environmental Protection Agency to respond to industrial air pollution during a public community meeting. A mobile smell reporting system that I co-designed with citizens has attracted more than 6,500 users who contributed over 25,000 reports, used in protesting air pollution. My other air quality work has engaged over 60 citizens to label industrial smoke emissions from over 12,000 videos to train a computer vision model that supports local non-profit organizations in collecting air pollution evidence. Regarding scientific contributions, my work extends from HCI and AI with the lens of community empowerment. I focus on improving interactions between citizens and technology by integrating HCI (Research through Design, Action Research, and Co-Design) and AI (Machine Learning, Computer Vision, and Data Science). My research has made empirical, artifact, opinion, and dataset contributions to design (ACM CHI [2]), crowd computing and human computation (ACM IUI [1]), computer science (CACM [4]), and AI for social impact (AAAI [3]).

My research has three special characteristics that bring different challenges. First, I conduct research in the wild from a social impact domain to understand unknown unknowns (those that we do not even know to exist). This is different from known unknowns, where the research questions can be clearly defined, experimented with hypothesis testing, and tested using well-structured data. In my research, scientists often have limited control, and the outcome can be largely different from the expectations. Therefore, new research methodologies are needed to tackle such uncertainty. Second, my research is often under time pressure from society, which means it is urgent to frame the problem with stakeholders and make various tradeoffs timely in the decision-making process. Also, citizens usually

are concerned about being treated as experimental subjects. Thus, we need new ways of collaborating with citizens in research to address such time pressure and ethical concerns. Third, my research requires time-consuming and large-scale field studies, which means that collecting field data can be very difficult, and the citizen-contributed data can be biased and noisy. Also, designing experiments can be unethical or infeasible, and the evaluation requires multiple aspects from the scientific and social perspectives. Hence, new pipelines are required in collecting, processing, and analyzing such large-scale hybrid data. Below are my three major works that reflect these characteristics.

Community-Empowered Air Quality Monitoring System [2] (`http://shenangochannel.org/`). In Pittsburgh (Pennsylvania, USA), 70,000 residents suffered from hazardous smoke emissions with fine particulates from a petroleum coke refinery. Although regulators and scientists in the local health department conducted air pollution studies, the result was rarely disseminated in an understandable form to the public. So, residents organized a community to gather evidence about how air pollution affected the local region. But the data that they collected manually was often held in low regard due to the lack of scientific validity. Over time, this non-transparency exacerbated a lack of trust between citizens and regulators. As technological supporters, we aided the community in setting up outdoor air quality sensors and monitoring cameras. We also distributed Google Forms to capture odor complaints. To visualize the collected hybrid data, we developed a web-based interactive system. Residents could use the system to search for and create videos with smoke emissions manually. Because this manual approach required considerable human efforts, we implemented a computer vision tool to automatically detect smoke and produce videos, which could then be curated and shared. With the system, residents could tell stories with scientific evidence about what happened and how these events affected the local neighborhood. Our analysis of user behaviors showed that both manual and automatic approaches are valuable when gathering evidence. The manual approach encouraged communities to generate videos with smoke emissions. The automatic approach with computer vision reduced the workload of finding smoke videos, which sustained long-term participation. Our survey study showed that the system promoted critical discussions with policy-makers and empowered citizens to take community actions. Also, the smoke videos provided by the system, combined with affected residents' stories, enabled effective communication of air pollution between citizens and regulators at a public meeting[1].

Smell Pittsburgh: Community-Empowered Mobile Smell Reporting System [1] (`https://smellpgh.org`). Cities like Pittsburgh deployed expensive monitoring stations sparsely to measure air pollution. However, in this way, it is difficult to understand how residents experience air pollution at the neighborhood level, such as poor smell. Currently, Pittsburgh citizens can report industrial odors to the municipality via its phone line or website. But citizens may not remember the exact time and location that pollution odors occurred, which results in missing data. Also, there is no real-time feedback of sharing experiences, making it hard to know if air pollution is at a neighborhood or city-wide scale. We developed Smell Pittsburgh, a mobile application for citizens to report pollution odors with time and location data. The system visualizes odor reports, which enables residents to view if others also share similar experiences. Also, by forwarding the reports to regulators, the system adds more weight to these reports, and regulators can access the data to track pollution sources. We used the reports and air quality measurements to train a machine learning model that predicts upcoming pollution odors and sends push notifications to users. We also statistically explained relationships between smell reports and air quality measurements. Our survey study found that motivations for citizens to use Smell Pittsburgh came mainly from internal factors, including the desire to contribute evidence, altruism, and the ability to validate personal experiences. Although stakeholders typically view odor reports as subjective and noisy, this work shows that smell data can help identify air pollution patterns and empower communities to pursue better air quality. The data were presented by advocates at the Board of Health meeting, which shows that the system encouraged citizens to communicate local issues with regulators[2]. The regulators later announced that they would enact rigorous rules for petroleum coke plants, which shows that this

[1]Link to the social impact of the air quality monitoring system – `https://www.post-gazette.com/news/environment/2015/11/19/Regulators-reviewing-Shenango-Coke-Works-compliance-with-2012-consent-decree/stories/201511190230`

[2]Link to the social impact of Smell Pittsburgh – `https://triblive.com/local/allegheny/13878930-74/allegheny-county-health-department-defends-air-quality-efforts-plans-stricter-coke-plant`

community empowerment approach can benefit agencies in effective decision-making. Such impacts can hardly be achieved by applying traditional methods that only involve researchers and regulators.

Project RISE: Recognizing Industrial Smoke Emissions [3] (`https://smoke.createlab.org`**).** Currently, air quality advocacy groups in Pittsburgh rely on a manual approach (US EPA Visual Opacity Reading) to determine if smoke emissions violate the permit issued to the facility, which is laborious and time-consuming. Computer vision (CV) techniques are promising in identifying smoke emission automatically. However, existing datasets are not of sufficient quality nor quantity to train the robust CV models needed to support air quality advocacy. We co-designed and deployed a system with air quality advocates to invite citizens to annotate whether industrial smoke emissions exist in video clips (obtained from our camera monitoring network). The design and use of this system were iteratively refined based on community feedback. We recruited volunteers to help label smoke through two workshops with air quality advocates, three presentations during community events, and two guest lectures at universities. Our dataset contains 12,567 video clips from 30 days spanning four seasons over two years, all taken in daylight. The labeled clips have 19 views cropped from three panoramas taken by cameras at three locations. The dataset covers various characteristics of smoke, including opacity and color, under diverse weather and lighting conditions. We ran experiments using I3D ConvNet architecture with Inception-v1 layers (a representative deep neural network for action recognition) to establish a robust performance baseline. Project RISE demonstrated a way to empower citizens through AI research and make practical design challenges explicit to others. Our survey study showed that volunteers' motivations involved the importance of supporting air quality advocacy, the wish to help scientific research, and the desire to push regulators in law enforcement. We have deployed the AI model to recognize smoke. Community activists and health department officers have used the system to curate a list of severe pollution events as evidence to conduct air pollution studies. We envision that this work can encourage others to keep communities and citizens at the center of every AI system design stage.

Future Research Plans

I plan to promote the "scientists as citizens" concept to make scientific research socially responsive and responsible, especially for the Sustainable Development Goals. My vision is to reinvent community-university partnerships, going from top-down academy leadership to a future of shared prosperity by establishing sustainable relationships among citizens, scientists, and other stakeholders. I have started this vision by piloting air quality advocacy projects and creating the Community Citizen Science [4] (CCS) framework accordingly. CCS embraces participatory democracy in scientific research, adopts community co-design in developing systems, and rebalances power relationships by disseminating data-driven evidence. I plan to use CCS to explore: *(i)* how technology can strengthen the link between science and society, *(ii)* how researchers can take on the role of supporters to facilitate using and disseminating technology, and *(iii)* how technology can empower citizens to produce scientific evidence. I have collaborated with researchers and stakeholders (citizens and non-profit organizations) with multidisciplinary skills, including software engineering, design, data science, AI, citizen engagement, community outreach, environmental health, and public policy. I will continue the interdisciplinary work with the vision of democratizing science when I pursue grants to fund my research and collaborations. Below is an outline of my 5-year research plan.

AI for Citizen Science and Scientific Citizenship. Modern AI techniques, such as machine learning and computer vision, can help communities make sense of complex data and identify critical evidence. However, the current process of engineering AI systems focuses mainly on model performance (e.g., f-score), making AI a black box for citizens and stakeholders. I am interested in exploring sustainable citizen-scientist collaborations to democratize AI research. How can citizens participate in AI research, including data collection, data labeling, model training, model validation, system deployment, and impact evaluation? How can scientists collaborate with citizens in applying AI to address local concerns and take community actions, such as environmental health and public policy? How can citizens' feedback be taken into account when fine-tuning and refining AI systems? I am also curious

about the co-design methodology that can create meaningful social impact and raise citizens' confidence in pursuing sustainability. How can we adopt the existing research methods (or propose new ones) to document the community co-design practice in building AI systems? What are the common design patterns that can be applied in co-creating AI systems? What do AI equity, fairness, accountability, and transparency mean to citizens and other stakeholders, and how can we communicate the associated findings through community outreach? Besides, citizen science projects can suffer from biases, which means a small number of volunteers contribute a large amount of data. If the biased information is used to inform policy-making, it could lead to negative social consequences, such as racial discrimination in facial recognition and the discriminatory practice of denying services (redlining). How can we identify and mitigate biases in citizen-contributed data? How can we co-design AI systems to reach a broader audience and achieve deep participatory democracy? How can we take the biases into account when training and refining AI models? How can we diversify human interventions and add diverse human values to the data collection process?

Crowdsourcing, Sensing, and Analytic for Sustainable Cities. Cities are facing sustainability challenges regarding their rapid development, such as air quality, inclusive public spaces, food safety, disaster resilience, affordable housing, public health, and renewable energy. These challenges often involve many stakeholders (e.g., citizens, scientists, policy-makers) who can have tensions and conflicts of interest. To understand the challenges in the local context, it is essential to co-design and deploy technology with stakeholders in crowdsourcing citizen feedback, sensing the urban environment, and analyzing the large-scale hybrid data. We need to move beyond the mindset of testing technology in urban contexts to treating technology as urban infrastructure that sustains long-term participation. In the short term, I plan to continue my collaboration with the Citizen of the Earth Foundation (a non-profit organization that advocates for environmental justice) and the g0v open government community in Taiwan on engaging citizens to annotate illegal factories on farmlands using satellite images. I also plan to continue my work of labeling and sensing industrial smoke emissions with Pittsburgh communities and the CREATE Lab at Carnegie Mellon University. In the long term, besides tackling environmental concerns, I am interested in exploring ways of co-creating tools and methods to engage stakeholders in addressing other types of sustainable development challenges that can have controversial perspectives. For example, how can we build a platform to discuss if the municipalities should deploy more security cameras in public spaces? How can we use crowdsourcing and urban analytics to understand citizens' concerns about returning to the shared workspace from COVID-19 lockdown? I am also curious about the methodologies of co-creating a sustainable ecosystem. How can we create such an ecosystem and a solid community of stakeholders? How can we encourage constructive conversations among stakeholders using crowdsourcing? How can citizens, scientists, policy-makers, and industry co-create knowledge in advocating for sustainable cities?

Co-Design and Co-Research for Diverse, Equal, and Inclusive Society. The current design practice and scientific research to address diversity, equality, and inclusion (DEI) are primarily centered on designers and scientists, based on the assumption that they can empathize with citizens who are impacted by the problems. Beyond doing so to be "like" someone, I believe it is more important to be "with" someone to co-create shared experiences and historicity. By being with the citizens affected by DEI concerns, designers and scientists acknowledge that it is challenging to understand citizens' experiences fully. Only by admitting this weakness can designers and scientists sincerely involve citizens in the center of the design process when creating artifacts. I am interested in studying the methodology for establishing such an DEI ecosystem of shared prosperity among local stakeholders (communities, universities, and municipalities), especially in developing information technologies, including AI, robotics, data analytics, and visualization. How can we create and deploy such information technologies in reducing inequality and pursuing environmental injustice? How can we equip citizens with tools to overcome the structural barriers that are disproportionately affecting underserved communities? How can we apply design and scientific research of information technologies to advocate for diversity, equality, and inclusion? I am also interested in co-creating technology to forge a common ground of collaboration in DEI challenges. How can we engineer visualization and urban analytics systems to reveal the challenges and identify patterns of DEI concerns in our city? How can we empower local citizens to collect, analyze, and disseminate data related to DEI challenges?

References

[1] **Yen-Chia Hsu**, Jennifer Cross, Paul Dille, Michael Tasota, Beatrice Dias, Randy Sargent, Ting-Hao (Kenneth) Huang, and Illah Nourbakhsh. 2020. Smell Pittsburgh: Engaging Community Citizen Science for Air Quality. *ACM Transactions on Interactive Intelligent Systems* 10, 4, Article 32 (Nov. 2020), 49 pages. DOI:http://dx.doi.org/10.1145/3369397

[2] **Yen-Chia Hsu**, Paul Dille, Jennifer Cross, Beatrice Dias, Randy Sargent, and Illah Nourbakhsh. 2017. Community-Empowered Air Quality Monitoring System. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1607–1619. DOI:http://dx.doi.org/10.1145/3025453.3025853

[3] **Yen-Chia Hsu**, Ting-Hao (Kenneth) Huang, Ting-Yao Hu, Paul Dille, Sean Prendi, Ryan Hoffman, Anastasia Tsuhlares, Jessica Pachuta, Randy Sargent, and Illah Nourbakhsh. 2021. Project RISE: Recognizing Industrial Smoke Emissions. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 17 (May 2021), 14813–14821. https://ojs.aaai.org/index.php/AAAI/article/view/17739

[4] **Yen-Chia Hsu** and Illah Nourbakhsh. 2020. When Human-Computer Interaction Meets Community Citizen Science. *Communications of the ACM* 63, 2 (Jan. 2020), 31–34. DOI:http://dx.doi.org/10.1145/3376892