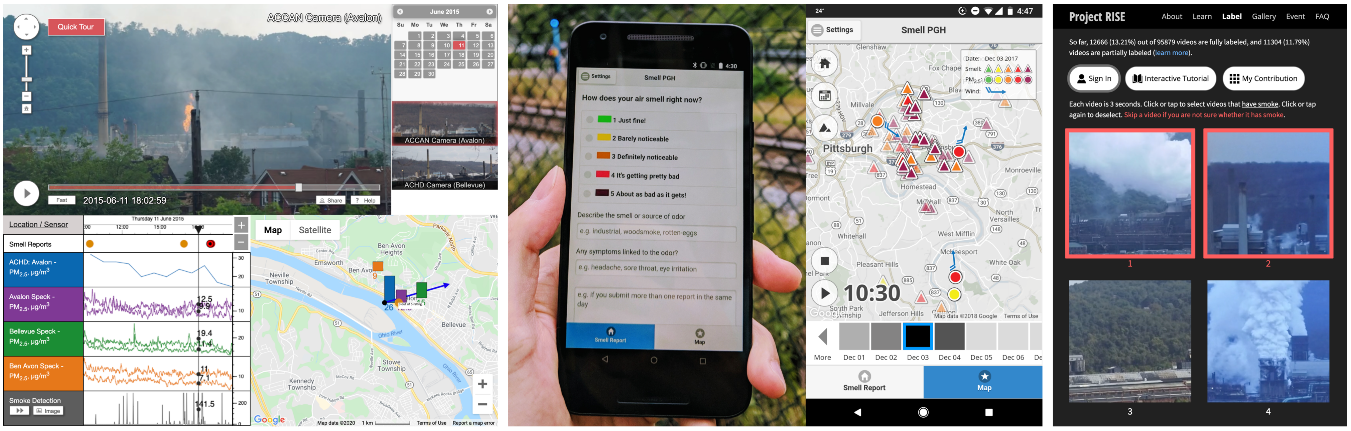
Yen-Chia Hsu | Research Statement



(a) (b) (c)

Figure 1: (a) Air quality monitoring system [4], (b) Smell Pittsburgh [3], and (c) Project RISE [5].

My research sits between AI (Artificial Intelligence) and HCI (Human-Computer Interaction), where I apply multi-disciplinary knowledge to address critical social and environmental concerns raised by local citizens. **I frame my research field as "Community-Empowered AI", where I co-design, implement, deploy, and evaluate interactive AI systems to empower communities.** Conventionally, when engaging residents in addressing local concerns, scientists lead research activities and tend to have more power over citizens. However, this top-down approaches usually pursue technology interventions that treat local participants as experimental subjects, which may cause adverse impacts to communities. I have proposed an alternative, **Community Citizen Science (CCS)**, to closely connect research and social issues by empowering everyday citizens to produce scientific knowledge, represent their needs, address their concerns, and advocate for impact [6]. Through CCS, I integrate sustainable HCI and AI to advance the current science-oriented method by emphasizing continued community engagement after technology interventions.

My works emphasize going from research to practice to impact and have been featured in more than 60 media articles. I have collaborated in developing an air quality monitoring system that urged US EPA (United States Environmental Protection Agency) regulators to respond to the air pollution caused by a petroleum coke refinery during a public meeting [2, 4]. Also, one month after the meeting, the parent company of the refinery announced its closure, which was the ultimate goal that the community had tried to achieve for decades. In certain ways, this air quality monitoring system indirectly caused a social impact that **improved the living quality of over 70,000 residents**. Based on the experiences learned from the air quality monitoring system, I also collaborated in developing a mobile smell reporting system, which **has attracted more than 6,500 users who contributed over 25,000 reports** in 32 months since its release [3]. The system enabled community activists to debate air quality issues with scientific evidence, which urged the local health department regulators to enact rigorous rules for coke plants [1]. Recently, I co-designed a system that **has engaged over 60 citizens to label industrial smoke emissions from over 12,000 videos**, which were used to train a computer vision model to recognize smoke [5]. In the future, I plan to establish Community Citizen Science as a formal research field with the goal of empowering communities and scientists to represent their voices, reveal local concerns, and shape more equitable power relationships.

What is Community Citizen Science?

Citizen science refers to the framework that empowers amateurs and professionals to form partnerships and produce scientific knowledge jointly. A major science-oriented strand aims to facilitate scientific research and address large-scale problems that are infeasible for experts to tackle alone. One example is Galaxy Zoo[1], which invites volunteers to classify a large number of galaxies online. On the other hand, Community Citizen Science, the main framework

[1]Link to Galaxy Zoo – `https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/`

of my research, aims to democratize science by equipping citizens with technology to directly target concerns that are raised by communities [6]. CCS is a particular case of citizen science that:

- **Embraces participatory democracy** through scientific research to influence policy-making and address local concerns that community members wish to advocate for themselves
- **Adopts community co-design** to develop interactive systems with advocacy groups, who are deeply grounded in local cultures and can bring diverse expertise to inform the design and use of computational tools
- **Rebalances power relationships** by disseminating critical data-driven evidence from the bottom-up to inform and convince decision-makers about the perceptions of community concerns

CCS is related to sustainable HCI, which studies the intervention of technology for increasing the awareness of sustainability, changing user behaviors, and influencing attitudes of stakeholders. CCS seeks to generate scientific knowledge from community data to support exploration, understanding, and dissemination of local environmental, social, and ethical concerns. CCS further explores: *(i)* how technology can strengthen the link between science and society, *(ii)* how researchers can take on the role of supporters to facilitate utilizing and disseminating technology, and *(iii)* how technology can empower citizens to produce scientific evidence and rebalance power relationships.

Community-Empowered Air Quality Monitoring System[2]

In several local neighborhoods 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. Over time, this non-transparent approach exacerbated a lack of trust between citizens and regulators. Unlike this traditional agency-led approach that leaves citizens out of the loop during decision making, we adopted Community Citizen Science to empower residents in advocating for better air quality.

Although residents organized a community to gather evidence about how air pollution affected the local region, the data that they collected manually was often held in low regard due to the lack of scientific validity. 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 (sensor readings, odor complaints, and monitoring videos), we developed a web-based interactive system (Figure 1, left) [4]. 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 corresponding videos, which could then be curated and shared. With the system, residents could tell stories with concrete scientific evidence about what happened and how these events affected the local neighborhood.

Co-designing and deploying this system has offered insights for CCS and led to social impacts. From the analysis of server logs, we recommend combining manual and automatic approaches when gathering evidence. The manual approach encouraged communities to generate videos with smoke emissions. The automatic approach with computer vision reduced the workload by expediting the process of finding smoke videos, which sustained long-term participation. From a survey study, we found statistically significant increases in self-efficacy and a sense of belonging among users. Open responses in the survey showed that the system promoted critical discussions with policy-makers and empowered citizens to take community actions. Moreover, the smoke videos provided by the system, combined with affected residents' stories, have enabled effective communication of air pollution between citizens and regulators at a public meeting [2]. The acting director of the EPA from the Region III Air Protection Division in Philadelphia pointed at the videos projected on a screen and said: "But what I see in the video, is totally unacceptable." This impact shows the value of evidence-based visualization. In December 2015, the refinery was closed by its parent company, which was the ultimate goal that the community had tried to achieve for decades.

[2]Link to the air quality monitoring system – `http://shenangochannel.org/`

Smell Pittsburgh: Community-Empowered Mobile Smell Reporting System[3]

Cities like Pittsburgh deployed expensive monitoring stations sparsely to measure the impact of air pollution. However, in this way, it is difficult to understand how residents experience air pollution at the neighborhood level. Based on lessons learned from developing the air quality monitoring system, residents linked odor experiences with living quality. Although Pittsburgh citizens can report industrial odors to the local health department via its phone line or website, the quality of the data is doubtful. Citizens may not remember the exact time and location that pollution odors occurred, which results in missing data that can affect the statistical analysis of identifying pollution sources. Furthermore, the reporting process is not transparent. There is no real-time feedback or ways of sharing experiences, which makes it hard to know if air pollution is at a neighborhood or city-wide scale.

Smell Pittsburgh is an interactive system for citizens to report pollution odors with time and location data via mobile devices (Figure 1, middle) [3]. The system visualizes odor reports in real-time, which enables residents to view if others also share similar experiences. Moreover, by forwarding the reports to regulators, the system adds more weight to these reports, and regulators can access the data to address potential pollution sources. Also, we used the reports and air quality measurements to train a machine learning model that predicts upcoming pollution odors and send 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 [1]. The regulators later announced that they would enact rigorous rules for petroleum coke plants, which shows that this CCS approach can also benefit agencies in effective decision-making [1]. Such impacts can hardly be achieved by applying traditional methods that only involve researchers and regulators.

Project RISE: Recognizing Industrial Smoke Emissions[4]

Currently, citizens who wish to advocate for better air quality rely on a manual approach (US EPA Visual Opacity Reading[5]) to determine if smoke emissions violate the permit issued to the facility, which is laborious and time-consuming. Our prior work has shown that using computer vision (CV) techniques to identify smoke can help residents collect visual evidence of pollution, empower citizens to pursue environmental justice, and influence the attitude of regulators [4]. However, the CV model that we developed in the prior work was based on heuristic approaches that were highly tuned for one specific camera. Also, existing datasets are not of sufficient quality nor quantity to train the robust CV models needed to support air quality advocacy.

To overcome the data quality and quantity problem, we co-designed and deployed a video labeling system with air quality advocates [5]. The system allowed us 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 strong performance baseline. By fine-tuning the I3D ConvNet with various temporal

[3]Link to Smell Pittsburgh – `https://smellpgh.org`
[4]Link to Project RISE – `https://smoke.createlab.org`
[5]Link to US EPA Visual Opacity Reading – `https://www.epa.gov/emc/method-9-visual-opacity`

processing techniques, we revealed the challenge of using the temporal information in our dataset.

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 are working with our system to curate a list of severe pollution events as evidence to conduct air pollution studies. We envision that our work can encourage others to keep communities and citizens in the center of every AI system design stage.

Research Agenda

My prior works in Community Citizen Science have demonstrated that pursuing sustainability relies heavily on the understanding of local concerns. The implications of designing interactive AI systems to support CCS may depend on local context and can be difficult to generalize or replicate. In the future, I plan to establish CCS as a field that **emphasizes going from research to practice to impact**. Specifically, I am excited to explore the application of extending sustainable HCI with CCS. The formal training that I have received in both computer science and architectural design makes me very well suited for integrating computational and design thinking in developing interactive systems for CCS. Several future research topics are outlined below.

- **Community-Empowered AI:** Modern AI techniques, such as machine learning and computer vision, have the potential to help communities make sense of complex data and identify critical evidence. I am interested in exploring methodologies for citizen-scientist collaboration in AI research. I believe that such collaboration can create meaningful social impact and raise communities' confidence in addressing sustainability issues.
- **Smart City:** I am interested in using cyber-physical systems to address local community concerns, especially those related to sustainability (e.g., air quality, housing inequality, food safety, energy conservation). I believe that it is essential to co-design and deploy technology with communities. In this way, we treat technology as urban infrastructure, just like how architects and urban planners co-design buildings and open spaces.
- **Evidence-Based Visualization:** I believe visualizing sustainability issues with data-driven evidence is critical in influencing decision-makers and bringing about common discourse. Visualizations can raise public awareness and pave the road of building solutions around community concerns. I am interested in exploring both tangible and intangible interaction techniques to visualize multiple types of quantitative and qualitative data.

References

[1] Theresa Clift. 2018. Allegheny County Health Department defends air quality efforts, plans stricter coke plant rules. (2018). `https://triblive.com/local/allegheny/13878930-74/allegheny-county-health-department-defends-air-quality-efforts-plans-stricter-coke-plant`.

[2] Don Hopey. 2015. Regulators reviewing Shenango Coke Works' compliance with 2012 consent decree. (2015). `https://www.post-gazette.com/news/environment/2015/11/19/Regulators-reviewing-Shenango-Coke-Works-compliance-with-2012-consent-decree/stories/201511190230`.

[3] **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). `DOI:http://dx.doi.org/10.1145/3369397`

[4] **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)*. ACM, 1607–1619. `DOI:http://dx.doi.org/10.1145/3025453.3025853`

[5] **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. 2020. Project RISE: Recognizing Industrial Smoke Emissions. *arXiv preprint arXiv:2005.06111* (2020). `https://arxiv.org/abs/2005.06111` (Accepted and to be appeared in AAAI 2021).

[6] **Yen-Chia Hsu** and Illah Nourbakhsh. 2020. When Human-Computer Interaction Meets Community Citizen Science. *Communications of the ACM* 63, 2 (Jan. 2020), 3134. `DOI:http://dx.doi.org/10.1145/3376892`