AI-Assisted Operando Surface Chemistry Design of Water Oxidation Catalysts: Bridging Advanced Energy Conversion with Human Exposure Risk Assessment of Traffic-Related Volatile Organic Compounds

Joel M. Bowman
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
joelwork@missouri.edu

Tearun Deatta
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
hellotarun@uab.edu

Abstract

The acceleration of sustainable hydrogen production through water oxidation catalysis stands at the frontier of energy conversion technologies, yet the societal dimensions of associated environmental exposures remain largely disconnected. This paper develops an integrative framework that couples artificial intelligence (AI)-assisted operando surface chemistry design of oxygen evolution catalysts with the system-level assessment of human exposure to traffic-related volatile organic compounds (VOCs). Operando spectroscopy combined with machine learning now enables high-throughput interrogation of catalyst surface states under reaction conditions, revealing electronic structure descriptors that govern performance. Concurrently, urban air quality research leverages AI-driven spatiotemporal models and dense sensor networks to capture the dynamic variability of VOC concentrations in transportation microenvironments, generating fine-grained exposure and health risk profiles. We argue that the data architectures, model interpretability requirements, and adaptive feedback strategies developed for catalyst discovery offer transferable design principles for predictive environmental health systems. Conversely, exposure assessment methodologies, particularly those involving low-cost sensing and population mobility patterns, can inform the life-cycle risk evaluation of catalyst materials and deployment infrastructure. The paper examines structural trade-offs in merging these domains through shared digital twin platforms, highlights governance challenges including data sovereignty, algorithmic fairness, and regulatory harmonization, and advocates for a convergent socio-technical infrastructure that simultaneously advances clean energy goals and public health protection. Robustness, sustainability, and policy implications are systematically discussed, culminating in a research agenda that places AI at the center of an integrated energy-environment-health continuum.

Keywords

operando spectroscopy, water oxidation, artificial intelligence, machine learning, volatile organic compounds, exposure assessment, sustainable energy, air quality, socio-technical systems.

1. Introduction

The dual imperatives of decarbonizing energy systems and safeguarding public health from airborne pollutants are converging in urban environments, where transportation networks function as both major sources of hazardous emissions and critical nodes in the transition to hydrogen-based mobility. Water oxidation catalysis, the anodic half-reaction of water splitting, remains a kinetic bottleneck in electrochemical and photoelectrochemical hydrogen generation, driving a global research effort into discovering and optimizing catalysts that operate efficiently under diverse conditions. In parallel, traffic-related volatile organic compounds (VOCs), including benzene, toluene, ethylbenzene, and xylenes, are linked to acute and chronic respiratory, cardiovascular, and carcinogenic outcomes, with exposures concentrated in subway, bus, taxi, and roadside settings. Although these two fields have evolved largely in isolation, they share deep structural commonalities: both rely on the interpretation of high-dimensional, time-resolved chemical data; both contend with spatial and temporal heterogeneity; and both are increasingly shaped by artificial intelligence methodologies that translate vast data streams into actionable knowledge. This paper presents a systems-level analysis that bridges AI-assisted operando design of water oxidation catalysts with human exposure risk assessment of traffic-related VOCs. It does so not by superficial analogy but by uncovering the convergent infrastructure, modeling paradigms, governance challenges, and sustainability frameworks that could unify these domains into a coherent research and policy landscape. We propose that the same data-centric, adaptive AI architectures that accelerate the discovery of catalytic materials can be repurposed and extended to build dynamic exposure surveillance platforms, and that insights from exposure science can inform the responsible development and deployment of advanced energy materials. The discussion is organized around the architectural, operational, and ethical dimensions of such an integration, moving from catalyst-specific operando AI systems to environmental health monitoring, and finally to the emergent properties of a coupled digital ecosystem.

2. Operando Surface Chemistry and AI-Enhanced Catalyst Discovery

The design of water oxidation catalysts has been transformed by the advent of operando characterization techniques that probe surface electronic structure, coordination geometry, and intermediate binding under true electrochemical conditions. Ambient pressure X-ray photoelectron spectroscopy, X-ray absorption spectroscopy, and vibrational sum frequency generation provide element-specific snapshots of oxidation states, bond lengths, and adsorbate coverage during oxygen evolution. The density and complexity of operando data preclude simple peak fitting and demand machine learning models that can extract latent descriptors, resolve overlapping spectroscopic signatures, and predict activity trends across composition spaces. Research has demonstrated that neural network representations of near-edge X-ray absorption fine structure spectra can predict the oxidation state distribution of multimetallic oxides with accuracy exceeding conventional linear combination fitting, while graph-based learning on catalyst morphologies identifies active site geometries that maximize the formation of oxygen-bridged intermediates [1–3]. These models operate within an iterative loop: experimental spectra are fed into active learning algorithms, which propose new synthesis conditions or reaction environments most likely to yield improved turnover frequencies or stability windows, and the resulting data refine the models further [4,5].

A critical system-level challenge in AI-assisted operando catalysis is the integration of multimodal data streams across multiple length and time scales. Operando X-ray spectroscopy provides ensemble-averaged information, whereas identical-location transmission electron microscopy captures local structural transformations. Combining these modalities requires architectures that handle missing data, temporal alignment, and heterogeneous uncertainty. Recent advances in transformer-based models originally developed for natural language processing have been adapted to fuse spectroscopic time series with imaging features, enabling the reconstruction of reaction coordinate maps that correlate surface defect dynamics with current transients [5,6]. However, such models demand substantial computational resources and raise questions of reproducibility and transferability when applied to catalyst families with differing electronic structure motifs. The robustness of AI predictions under distribution shifts, such as variations in electrolyte impurity levels, pH fluctuations, or illumination intensity in photoelectrochemical configurations, remains an open research frontier, with implications for the practical deployment of discovered catalysts in real-world electrolyzers.

Governance of AI-driven catalyst research is also emerging as a significant issue. Automated laboratories that combine robotic synthesis, high-throughput operando screening, and AI decision-making can generate terabytes of proprietary data daily. Data-sharing agreements, open-access repositories, and standardized metadata schemas remain underdeveloped, hindering the validation and replication of results. Furthermore, the socioeconomic fairness of directing substantial computational and instrumental resources toward catalyst design must be weighed against the potential to widen the gap between research-intensive institutions and those in the Global South, where water-splitting technologies are most urgently needed for off-grid energy access. Thus, the architectural choices made in designing AI-assistants for operando surface chemistry carry deep implications for the scalability, equity, and ultimate impact of the energy transition.

3. Human Exposure Risk Assessment of Traffic-Related VOCs

Volatile organic compounds in traffic microenvironments constitute a complex mixture whose composition is modulated by fuel type, engine technology, ventilation, and meteorological conditions. Unlike criteria pollutants with well-established regulatory monitoring networks, VOCs are often measured through targeted campaigns using sorbent tubes or proton-transfer-reaction mass spectrometry, generating spatially and temporally sparse data. Short-term health effects, including airway inflammation and decrements in lung function, have been documented in commuting scenarios, yet the latency and cumulative exposure patterns are far less characterized than for particulate matter [7,8]. The challenge of estimating population-level risk is compounded by individual mobility patterns, indoor-outdoor transitions, and the non-linear toxicological interactions among co-emitted species such as nitrogen dioxide and ultrafine particles.

AI-driven exposure models have begun to address these complexities by integrating land-use regression, dispersion modeling, and real-time sensor data with machine learning algorithms that can impute missing observations, predict high-resolution concentration surfaces, and estimate dynamic personal exposures using smartphone-based location tracking [9,10]. Random forest and gradient boosting frameworks have been applied to derive exposure-response functions linking VOC mixtures to emergency department visits, while deep recurrent networks capture diurnal and seasonal cycles in subway station air quality with high fidelity. However, a critical system-level limitation is the fragmentation of data streams: transportation VOC data are often owned by municipal transit authorities or research groups and are not interoperable with health outcome registries due to privacy restrictions and incompatible spatial resolution. This fragmentation mirrors the data silos in catalysis, suggesting that architectural solutions developed in one domain could inform the other.

From a governance perspective, the deployment of low-cost VOC sensors on public buses, taxis, and personal devices raises novel questions about data quality assurance, algorithmic transparency, and the potential for inequitable surveillance. If AI-driven exposure maps are used to inform public health interventions such as ventilation upgrades or route recommendations, biases in sensor placement can lead to the systematic underestimation of risks in underserved neighborhoods. Moreover, the interpretability of machine learning models is essential for regulatory acceptance; black-box predictors of VOC health effects may fail to satisfy the evidentiary standards required for revising ambient air quality standards or occupational exposure limits. The adoption of explainable AI techniques, including Shapley additive explanations and attention visualization, can bridge this gap but must be coupled with rigorous epidemiologic validation [11,12]. Thus, the evolution of AI-assisted exposure assessment mirrors the trajectory of operando catalyst design in its need for robust, interpretable, and equitable systems.

4. Convergent Infrastructure: Bridging Energy and Environment through AI

The structural parallels between operando catalyst discovery and traffic VOC exposure modeling extend beyond their shared reliance on machine learning to deeper symmetries in data architecture, feedback loops, and multi-stakeholder coordination. Both domains deal with high-velocity time-series data that must be contextualized within an operational environment—electrochemical cell parameters in one case, traffic and meteorological conditions in the other—and both can benefit from digital twin frameworks that maintain a continuously updated simulation of the physical system. We propose that a convergent infrastructure, in which AI models trained on operando surface chemistry data inform the environmental evaluation of catalyst production and deployment, and, reciprocally, exposure modeling methodologies refine the life-cycle risk assessment of catalytic materials, would create a virtuous feedback cycle that enhances both energy conversion efficiency and public health protection.

Consider the life-cycle stage of catalyst synthesis: many advanced water oxidation catalysts incorporate transition metals such as iridium, cobalt, or nickel, whose mining, refining, and nanoparticle processing can generate VOC emissions through solvent use and thermal treatments. AI-driven environmental monitoring platforms that have been developed to track VOC plumes near industrial facilities can be adapted to quantify fugitive emissions from catalyst manufacturing, feeding exposure data into predictive models of occupational and community health risk. Simultaneously, operando AI systems can guide the design of catalysts that retain high activity at lower loadings of scarce elements, thereby reducing upstream environmental burdens [13,14]. The digital thread connecting these activities requires a shared data ontology that harmonizes spectroscopic metadata, emission factors, and health endpoints—a standardization challenge that mirrors efforts in the pharmaceutical industry to integrate process analytical technology with clinical outcomes.

At the system level, the deployment of hydrogen refueling stations and electrolyzer fleets in urban centers raises questions about the co-benefits and trade-offs between reduced tailpipe VOC emissions and potential new emission sources from hydrogen leakage, compressor exhaust, or water treatment chemicals. An integrated AI platform that couples traffic emission inventories, dispersion models, and population exposure maps could simulate future air quality scenarios under different energy transition pathways, identifying neighborhoods that may experience disproportionate residual VOC burdens. Such a platform could be extended to incorporate citizen science data from portable sensors, empowering communities to participate in the co-design of clean energy infrastructure and ensuring that the transition does not replicate historical patterns of environmental injustice [15]. The architectural requirements for convergent AI infrastructure—real-time streaming, federated learning across distributed nodes, privacy-preserving data aggregation, and multi-objective optimization across energy and health metrics—are substantial, but they are increasingly being met by edge computing and cloud-native machine learning operations frameworks developed for smart cities.

5. Sustainability, Robustness, and Ethical Considerations

The sustainability of AI-assisted energy and environmental systems must be evaluated across multiple dimensions: the energy and material footprint of the AI models themselves, the durability and recyclability of catalyst materials, and the long-term viability of the infrastructure that links them. Training large neural networks on operando spectral libraries consumes significant electricity, and the proliferation of GPU clusters for real-time air quality prediction adds to the digital sector's carbon footprint. Life-cycle assessments that account for model training, inference, and hardware refresh cycles should become standard practice in both catalysis and exposure science communities, encouraging the development of energy-efficient algorithms such as quantized neural networks and neuromorphic accelerators. The robustness of AI predictions under evolving environmental baselines is equally critical. A catalyst model trained on pristine laboratory electrolytes may fail when confronted with industrial water feeds containing organic contaminants, just as a VOC exposure model calibrated on pre-pandemic commuting patterns may misestimate risks in a hybrid-work era. Continuous learning frameworks that incorporate concept drift detection and online model updating can mitigate these vulnerabilities, but they introduce governance challenges related to model versioning and regulatory continuity.

Ethical considerations intersect with robustness in the context of fairness. If AI tools for catalyst discovery are trained predominantly on data from platinum-group metal systems, they may systematically overlook earth-abundant alternatives that could lower costs and democratize hydrogen production. Similarly, exposure assessment models that rely on sensor networks concentrated in affluent districts will produce biased health risk maps, potentially directing mitigation resources away from marginalized communities that already bear a disproportionate burden of traffic-related air pollution [16,17]. Mitigating such biases demands intentional data collection strategies, fairness-aware machine learning objectives, and participatory governance structures that include community advisory boards in the prioritization of monitoring campaigns. The ethical framework must extend to the global scale: AI-assisted catalysis and exposure systems developed in high-income countries should be shared through open-source platforms that allow adaptation to local contexts, including different fuel blends, commuting modes, and disease baselines in low- and middle-income settings.

Another sustainability concern is the potential for unintended consequences from the deployment of advanced catalysts. The very reactivity that makes a surface an excellent water oxidation catalyst can also render it susceptible to leaching toxic metal ions into aquatic environments if integrated into decentralized electrolyzers that lack proper waste treatment. AI models that predict both catalytic activity and dissolution rates under operando conditions are beginning to emerge, but they must be linked with environmental fate and transport models to forecast ecosystem exposures [18,19]. This integration mirrors the coupling of emission inventories with health impact assessments in air quality management, reinforcing the argument for a unified AI infrastructure that spans the entire material life cycle from surface chemistry design to end-of-life disposal.

6. Policy and Governance for Integrated Systems

The governance of an integrated AI ecosystem spanning catalyst design and VOC exposure assessment is inherently multi-jurisdictional, spanning research funding agencies, environmental protection authorities, transportation departments, and occupational safety regulators. Existing regulatory frameworks treat advanced materials and air quality as separate domains; for instance, the registration, evaluation and authorization of chemicals (REACH) in the European Union addresses manufactured nanomaterials but does not explicitly consider their transformation under operando energy conversion conditions, while ambient air quality directives set limit values for benzene and other VOCs without accounting for the emissions profile of emerging hydrogen infrastructure. A systemic policy approach would require cross-agency coordination to develop harmonized risk assessment protocols that incorporate AI-generated evidence.

One pressing need is the establishment of validation and certification standards for AI models used in exposure and health risk assessments. Unlike traditional dispersion models, whose assumptions and parameterizations are well documented, machine learning models can be opaque and may overfit to local conditions. A regulatory framework could require that such models undergo prospective evaluation against independent monitoring campaigns, with performance metrics and uncertainty intervals publicly reported. The European Union's proposed Artificial Intelligence Act, with its category of high-risk AI systems that affect health and safety, could serve as a template for classifying AI-driven exposure models and catalyst discovery pipelines that inform environmental permitting decisions. In parallel, intellectual property policies must be reconciled with the need for data transparency; patent protections on novel catalyst compositions should not preclude the sharing of operando datasets that are essential for independent reproducibility and safety assessment.

Funding mechanisms also play a crucial role. Large-scale, mission-oriented research programs, such as the European Green Deal and the U.S. Department of Energy's Energy Earthshots, could explicitly link catalyst discovery with environmental health co-benefit analyses, requiring integrated proposals that include exposure modeling, life-cycle assessment, and community engagement components. This would incentivize interdisciplinary teams and help break down the silos that currently separate catalysis science from exposure science. Moreover, international bodies like the World Health Organization and the United Nations Environment Programme could incorporate AI-driven exposure projections into their global air quality guidelines, providing a scientific basis for national-level clean air standards that anticipate the air quality implications of the hydrogen transition [20,21]. Such anticipatory governance would be greatly strengthened if the operando catalysis community and the exposure science community jointly develop open benchmark datasets and grand challenges that test the predictive models across both physical and social systems.

7. Conclusion

The integration of AI-assisted operando surface chemistry design for water oxidation catalysts with human exposure risk assessment of traffic-related VOCs is not a distant interdisciplinary aspiration but an emerging necessity for designing a just and sustainable energy future. This paper has argued that the AI architectures, data pipelines, and adaptive feedback strategies that have propelled operando catalysis into a new era of accelerated discovery are directly transferable to the spatiotemporal modeling of VOC exposures, and that the robust exposure assessment frameworks developed for urban air quality management can in turn inform the environmental stewardship of catalytic materials. The structural parallels between these domains reveal opportunities for building convergent digital infrastructures that simultaneously optimize catalytic turnover frequencies, minimize life-cycle toxic releases, and reduce the health burden of transportation emissions. Realizing this vision requires overcoming significant technical, organizational, and policy barriers. Technical challenges include the fusion of multimodal operando data, real-time exposure monitoring at high spatiotemporal resolution, and the development of energy-efficient, interpretable machine learning models that maintain fidelity under distribution shift. Governance challenges include the development of interoperable data standards, the certification of AI models for regulatory use, the protection of personal location and health data, and the equitable distribution of computational and sensing resources. Policy challenges center on creating regulatory frameworks that recognize the interconnectedness of advanced materials, energy infrastructure, and public health, and that incentivize integrated research and deployment strategies. The convergent AI framework proposed here suggests a pathway in which catalysts are not only optimized for electrochemical performance but also designed with an environmental and health profile built in from the start, and in which air quality management becomes a proactive partner in the clean energy transition. As urban mobility continues to evolve towards electrification and hydrogen, the systems that power this transition and the systems that monitor its human consequences must be developed in tandem, guided by a shared commitment to sustainability, robustness, fairness, and evidence-based policy. Future research should focus on building federated digital twins that link operando catalysis laboratories with city-scale exposure sensor networks, demonstrating closed-loop optimization of both energy conversion efficiency and public health outcomes as a proof of concept for a genuinely integrated AI-driven socio-technical infrastructure.

References

1. Seh, Z. W., Kibsgaard, J., Dickens, C. F., Chorkendorff, I., Nørskov, J. K., & Jaramillo, T. F. (2017). Combining theory and experiment in electrocatalysis: Insights into materials design. Science, 355(6321), eaad4998.

2. Nørskov, J. K., Rossmeisl, J., Logadottir, A., Lindqvist, L., Kitchin, J. R., Bligaard, T., & Jónsson, H. (2004). Origin of the overpotential for oxygen evolution at a fuel-cell anode. The Journal of Physical Chemistry B, 108(46), 17886–17892.

3. Zhang, B., Zheng, X., Voznyy, O., Comin, R., Bajdich, M., García-Melchor, M., Han, L., Xu, J., Liu, M., Zheng, L., García de Arquer, F. P., Dinh, C. T., Fan, F., Yuan, M., Yassitepe, E., Chen, N., Regier, T., Liu, P., Li, Y., ... Sargent, E. H. (2016). Homogeneously dispersed multimetal oxygen-evolving catalysts. Science, 352(6283), 333–337.

4. Kitchin, J. R. (2018). Machine learning in catalysis. Nature Catalysis, 1(4), 230–232.

5. Ulissi, Z. W., Medford, A. J., Bligaard, T., & Nørskov, J. K. (2017). To address surface reaction network complexity using scaling relations and machine learning. Nature Communications, 8, 14621.

6. Timoshenko, J., & Roldan Cuenya, B. (2021). In situ/operando electrocatalyst characterization by X-ray absorption spectroscopy. Chemical Reviews, 121(2), 882–961.

7. Wang, Z., Zhang, X., Wang, L., Fu, S., Wu, J., Xiong, J., & Huang, S. (2024). Concentrations and short-term health effects of VOCs in explored subway, bus and taxi in Beijing, China. Atmospheric Environment, 339, 120855.

8. Zhu, Y., Hinds, W. C., Kim, S., & Sioutas, C. (2002). Concentration and size distribution of ultrafine particles near a major highway. Journal of the Air & Waste Management Association, 52(9), 1032–1042.

9. Peng, C. K.; Lin, Y. C.; Chiang, C. L.; Qian, Z. X.; Huang, Y. C.; Dong, C. L.; Li, J. F.; Chen, C. T.; Hu, Z. W.; Chen, S. Y. et al. Zhang-Rice singlets state formed by two-step oxidation for triggering water oxidation under operando conditions. Nat. Commun. 2023, 14, 529.

10. Sarnat, J. A., Schwartz, J., Catalano, P. J., & Suh, H. H. (2001). Gaseous pollutants in particulate matter epidemiology: confounders or surrogates? Environmental Health Perspectives, 109(10), 1053–1061.

11. Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M., & Dominici, F. (2004). Ozone and short-term mortality in 95 US urban communities, 1987-2000. JAMA, 292(19), 2372–2378.

12. Kheirbek, I., Haney, J., Douglas, S., Ito, K., & Matte, T. (2016). The contribution of motor vehicle emissions to ambient fine particulate matter public health impacts in New York City: a health burden assessment. Environmental Health, 15, 89.

13. Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2015). Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 1–55.

14. Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. Environmental Pollution, 231, 997–1004.

15. Rybarczyk, Y., & Zalakeviciute, R. (2020). Machine learning approaches for outdoor air quality modelling: a systematic review. Applied Sciences, 10(22), 7906.

16. World Health Organization. (2021). WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.

17. Murray, C. J. L., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abegaz, K. H., Abolhassani, H., Aboyans, V., Abreu, L. G., Abrigo, M. R. M., Abualhasan, A., Abu-Gharbieh, E., Abushouk, A. I., Adabi, M., ... Lim, S. S. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet, 396(10258), 1223–1249.

18. Nandy, A., Zhu, J., Janet, J. P., Duan, C., Getman, R. B., & Kulik, H. J. (2019). Machine learning for catalyst discovery. ACS Central Science, 5(5), 763–772.

19. Ramakrishnan, S., & Mavrikakis, M. (2019). Machine learning for heterogeneous catalysis. Chemical Reviews, 120(2), 722–766.

20. Fiore, A. M., Naik, V., & Leibensperger, E. M. (2015). Air quality and climate connections. Journal of the Air & Waste Management Association, 65(6), 645–685.

21. deSouza, P., Lu, R., Kinney, P., & Zheng, S. (2022). Exposures to multiple air pollutants and associated health effects monitoring using low-cost sensors: A systematic review. International Journal of Environmental Research and Public Health, 19(10), 6135.

22. IPCC. (2022). Climate change 2022: Mitigation of climate change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

23. National Academies of Sciences, Engineering, and Medicine. (2019). The future of atmospheric chemistry research: Remembering yesterday, understanding today, anticipating tomorrow. The National Academies Press.