Multi-Modal Network Traffic Forecasting with Large Language Models and Edge Intelligence in Smart Cities

Tobias Berry
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
berrytobias@missouri.edu

Anders Tougles
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
douglas1994@ku.edu

Abstract

The rapid proliferation of connected sensing infrastructure across urban environments has given rise to unprecedented volumes of multi-modal traffic data, encompassing vehicle counts, video feeds, weather measurements, social media signals, and more. Simultaneously, the emergence of large language models (LLMs) with strong generalization capabilities has begun to reshape spatiotemporal forecasting, moving beyond task-specific architectures toward models that leverage rich semantic representations across modalities. This paper presents an interdisciplinary systems analysis of multi-modal network traffic forecasting that integrates LLMs with edge intelligence in smart cities. We dissect the architectural trade-offs involved in placing forecasting intelligence across cloud, edge, and device tiers, examining how multi-modal data fusion, LLM fine-tuning, and edge-offloading strategies jointly determine latency, accuracy, privacy, and energy consumption. The discussion extends beyond algorithmic performance to address structural considerations of interoperability, resource governance, fairness across urban populations, robustness to distributional shifts, and sustainability in large-scale deployments. By synthesizing recent advances in graph neural forecasting, federated learning, and foundation models, the paper identifies systemic tensions between centralization and decentralization, between model expressiveness and inference cost, and between predictive accuracy and equitable service delivery. We argue that a successful deployment paradigm must treat the forecasting pipeline as a socio-technical infrastructure, embedding lifecycle management, continuous monitoring, and regulatory compliance from the outset. The analysis offers design principles and policy implications for city authorities and system architects seeking to harness LLM-driven multi-modal traffic forecasting in a responsible and resilient manner.

Keywords

Multi-modal traffic forecasting; large language models; edge intelligence; smart cities; spatiotemporal prediction; sustainability; fairness; governance.

1. Introduction

The vision of the smart city is predicated on sensing, computing, and communication fabrics that transform raw urban flows into actionable intelligence. Among the most central of these

flows, road traffic dynamics influence economic productivity, environmental emissions, and citizen well-being. Modern urban traffic management systems ingest data from a diverse array of modalities: inductive loop detectors, roadside cameras, GPS traces from vehicles and smartphones, weather stations, public transit feeds, event calendars, and even text-based incident reports or social media posts [1]. Effective traffic forecasting must therefore move beyond processing a single homogeneous time series toward fusing heterogeneous signals that capture physical movement, environmental context, and human behavior. This multi-modal challenge has historically been addressed by domain-specific models that range from classical statistical approaches such as seasonal autoregressive integrated moving average methods [2] to deep learning pipelines including long short-term memory networks [3] and, more recently, spatial-temporal graph neural networks that encode roadway topology alongside temporal dependencies [4], [5], [6], [7]. While each generation has yielded steady improvements in prediction accuracy, the growing diversity and velocity of urban data have exposed limitations in the architectural generality, transferability, and adaptability of task-specific forecasting models.

In parallel, the unprecedented scaling of natural language models has demonstrated that extremely large transformer-based architectures pre-trained on vast corpora can acquire a form of broad reasoning that generalizes across distinct downstream tasks with minimal fine-tuning. This development has triggered a wave of research that investigates whether large language models (LLMs) can absorb the structural priors of spatiotemporal data and perform zero- or few-shot traffic forecasting simultaneously across multiple modalities [10], [11]. Complementing this trend, edge intelligence is rapidly maturing as a paradigm that distributes computation toward the network periphery, thereby reducing end-to-end latency, economizing on backhaul bandwidth, and strengthening data sovereignty [8], [13]. The convergence of these two forces—LLM-based multi-modal modeling and edge-native inference—creates a design space with profound implications for how smart cities monitor and manage their traffic flows. Yet, to date, the systems community lacks a comprehensive treatment that weaves together the algorithmic, infrastructural, and governance dimensions of this convergence. The present paper provides an interdisciplinary systems analysis of multi-modal network traffic forecasting that is driven by large language models and deployed through edge intelligence architectures, paying explicit attention to structural trade-offs, sustainability, fairness, and policy.

2. The Multi-Modal Traffic Data Landscape

Understanding the systemic challenge of multi-modal forecasting requires a clear picture of the data sources that modern urban environments can potentially harness. At the physical layer, fixed sensors such as inductive loops and magnetometers provide high-frequency vehicle counts and occupancy measurements with well-characterized sparsity patterns. These are often supplemented by traffic cameras from which computer vision pipelines extract vehicle trajectories, queue lengths, and congestion levels, introducing a rich but computationally heavy visual modality. Simultaneously, floating car data and smartphone GPS probes deliver fine-grained trajectory samples albeit with varying penetration rates and privacy sensitivities. Beyond direct mobility observations, contextual modalities critically modulate traffic patterns: weather variables such as precipitation, temperature, and visibility shapes demand and safety margins; public transit schedules and real-time bus or metro loads reflect mode shifts; and event calendars or sports fixtures generate predictable demand surges that propagate through the road network in complex ways.

This heterogeneity imposes a multi-modal data fusion challenge that cannot be fully met by traditional deep learning models designed for a single input tensor structure. Early graph-based forecasting architectures such as spatial-temporal graph convolutional networks [4], diffusion convolutional recurrent networks [5], and Graph WaveNet [6] excelled at fusing sensor readings with a static road graph but struggled to incorporate language-based event descriptions or variable-resolution camera feeds. Attention-based extensions such as ASTGCN [7] offered greater flexibility in aligning spatial and temporal dimensions yet remained bound to the fixed set of modalities seen during training. A contemporary survey of deep neural approaches to traffic prediction has catalogued the accelerating yet fragmented attempts to ingest multi-source data, highlighting the recurring tension between model complexity and generalization [12]. As cities strive to unify data lakes that span structured sensor tables, visual streams, and unstructured text, the need for a forecasting substrate that can natively map all of these representations into a shared semantic space has become acute.

The rise of foundation models in the form of LLMs suggests a pathway to satisfy this need. Because LLMs are pre-trained on vast quantities of human language, they internalize a remarkably flexible representation of structure that can be redirected toward numeric time series, visual features, and symbolic encodings through appropriate tokenization and prompting strategies. Seen from a systems perspective, the multi-modal traffic data landscape is therefore transitioning from an era of bespoke fusion pipelines—where each new modality demands its own feature engineering and architectural coupling—to a potential era of universal traffic sequence modeling in which a single LLM backbone, augmented with lightweight modality-specific adapters, serves as the forecasting engine. The realization of this promise, however, hinges on addressing profound challenges of scale, latency, and placement that edge intelligence is uniquely positioned to resolve.

3. Large Language Models for Spatiotemporal Forecasting

The application of LLMs to time series and spatiotemporal data has evolved rapidly. Early works demonstrated that reprogramming a frozen large language model—by learning continuous prompts that map patches of time series into the model's embedding space—could yield competitive short-term forecasts without gradient-based fine-tuning of the base transformer [10]. Such reprogramming exploits the pre-trained attention layers' ability to capture long-range dependencies originally learned from natural language and repurposes this inductive bias for numerical sequences. Concurrently, UrbanGPT introduced a spatiotemporal large language model that explicitly conditions on geographic zones and temporal indices, thereby enabling the model to perform zero-shot forecasting across cities and sensor types after minimal language-guided instruction [11]. These advances have been complemented by architectures that generate full urban mobility scenarios through natural language conversation, opening a line of attack on interactive traffic simulation.

More recent work has pushed the frontier of LLM-based traffic intelligence by encoding the topological structure of road networks directly into the token stream processed by a large language decoder. The TIDES framework [14] leverages a powerful language model backbone to ingest spatial sensor graphs, temporal sequences, and auxiliary textual context within a unified sequence-to-sequence paradigm, achieving notable gains in long-horizon forecasting accuracy. The framework incorporates positional and spatial relation embeddings that allow the decoder's self-attention to recover non-Euclidean dependencies without an explicit graph convolution stage. Similarly, the ST-LLM model treats both spatial adjacency matrices and traffic feature series as input tokens that can be processed within a single

transformer stack, using a carefully designed spatial tokenizer and a temporal patching mechanism to preserve fine-grained dynamics while respecting the token budget [15]. Collectively, these models suggest that the era of the self-contained, multi-modal traffic predictor is approaching; yet the computational appetite of such models—often requiring tens of billions of parameters and thousands of token lookups per time step—raises serious concerns about their suitability for real-time, resource-constrained urban deployments.

Equally important is the question of generalization across urban contexts. A single LLM-based forecaster pre-trained on large corpora of city traffic data, weather archives, and map tiles may be able to adapt to a new neighborhood or a previously unseen city with only a handful of local examples, thereby amortizing the enormous training cost over many deployments. This transfer capability could be a decisive advantage for municipalities with limited data science resources, provided that the inference footprint can be managed. The tension between model expressiveness and deployment pragmatism therefore sets the stage for edge intelligence to play a mediating role, distributing the computational load across a hierarchy of compute nodes.

4. Edge Intelligence Architecture and Deployment

Edge intelligence architectures organize computing, storage, and networking resources along a cloud-to-thing continuum, pushing decision logic closer to data sources. In a smart city traffic forecasting context, this continuum can be segmented into cloud data centers, regional edge servers co-located with cellular base stations or traffic management centers, and on-device processors embedded in roadside units, connected vehicles, or embedded cameras. Each tier offers a different combination of compute capacity, memory, storage, energy budget, and communication latency. Deep forecasting models have historically been hosted in the cloud due to their computational demands, but this model creates a single point of failure, incurs high round-trip latency, and forces massive raw data exfiltration from the edge, undermining privacy and bandwidth efficiency [8]. The introduction of LLM-based multi-modal forecasters, whose inference costs can be orders of magnitude higher than those of a compact graph convolution network, amplifies these tensions.

Federated learning offers a partial remedy by enabling model training across geographically distributed edge nodes without transferring raw data to a central server [9], [13]. In a federated multi-modal forecasting setting, each roadside unit or traffic management zone could fine-tune a shared LLM backbone using locally collected sensor streams, camera feeds, and event metadata, then transmit only model gradients or parameter deltas to a coordinating server. This approach preserves data locality and complies with increasingly stringent data protection regulations while still harnessing diverse urban patterns to improve model robustness. However, the communication overhead of aggregating large language model updates across potentially hundreds of edge nodes, combined with the statistical heterogeneity of local traffic distributions, can lead to convergence instability and straggler effects that demand sophisticated aggregation and compression protocols.

An alternative deployment strategy involves splitting the model pipeline between tiers. A lightweight feature extractor or tokenizer can preprocess multi-modal data at the far edge—converting camera frames into object counts and trajectory vectors, or embedding event text into dense representations—while the heavy LLM decoder runs on a regional edge server that has greater GPU capacity and more predictable power supply. This split-inference paradigm slashes the volume of data that must traverse the wide-area network and allows rapid local reconfiguration if the cloud link is disrupted. The cost, however, includes the need for tight

synchronization between the encoder and decoder portions and the difficulty of optimizing the split point for heterogeneous hardware. Moreover, decisions about which modalities to process locally and which to offload become deeply consequential for system-wide metrics such as freshness of the forecast, energy consumption at battery-powered edge nodes, and the latency budget available for downstream control actions such as adaptive traffic signal timing.

Designing an edge intelligence fabric that supports LLM-driven multi-modal forecasting thus requires a careful negotiation between centralization and decentralization. A fully centralized approach simplifies model management and version control but is fragile under network partition and fails to exploit local data patterns. A fully decentralized approach maximizes resilience and privacy but fragments the training signal and complicates policy compliance auditing. The optimum typically lies in a hierarchical federation where a global LLM is periodically updated through secure aggregation of local adaptations, while latency-critical inference operations are executed on regional edge servers or even on vehicular on-board units for safety-related predictions. Such a system must also incorporate a lightweight orchestration layer that monitors resource availability, data drift, and quality-of-service contracts, dynamically routing inference requests to the most suitable tier. These architectural decisions transcend purely technical trade-offs and intersect with governance frameworks that determine how data stewardship, access rights, and accountability are distributed among city agencies, technology vendors, and citizens.

5. System-Level Trade-offs and Infrastructure Considerations

The shift toward multi-modal LLM-based forecasting deployed across the edge introduces a series of system-level trade-offs that span performance, cost, reliability, and maintainability. Perhaps the most immediate tension exists between model accuracy and inference latency. Larger LLM backbones with billions of parameters exhibit a superior ability to assimilate heterogeneous contextual signals and produce well-calibrated multi-step forecasts, yet their token-by-token generation process can easily exceed the sub-second latency envelope demanded by real-time traffic control applications. Model compression via quantization, knowledge distillation into smaller student networks, or speculative decoding can narrow this gap, but each technique introduces new failure modes such as reduced numerical precision that degrades rare-event prediction or increased model staleness if distillation cycles cannot keep pace with evolving traffic patterns.

A second critical trade-off concerns data freshness versus bandwidth consumption. Edge-deployed forecasters that can operate on local multi-modal streams avoid the delays and packet loss associated with backhaul networks, but they depend on high-quality local sensors. If a roadside camera fails or a loop detector becomes miscalibrated, the edge model must either fall back to a degraded unimodal forecast or request cloud assistance, thereby spiking wide-area traffic at precisely the moment of greatest operational stress. Designing a multi-tier data plane that gracefully degrades under sensor failure requires explicit redundancy management and real-time data quality monitoring, turning the forecasting system into a self-aware infrastructure rather than a static pipeline. Such capabilities intersect with supervisory control and data acquisition architectures already deployed in traffic management centers, suggesting that the forecasting module must be integrated into broader urban operating systems rather than deployed as a standalone application.

Energy and carbon footprints constitute a third systemic consideration. Training an LLM from scratch on multi-modal traffic data can consume tens to hundreds of megawatt-hours of electricity, leading to substantial embodied carbon emissions that must be weighed against the

operational emissions reductions that more accurate routing and signal control might enable [18]. Edge inference, while low-power per device, aggregates across tens of thousands of nodes and can exceed the energy budget of solar-powered roadside units if not carefully managed. Hardware-software co-design strategies such as dedicated neural processing units and model cascades—where a compact model runs continuously and a larger LLM is invoked only when uncertainty exceeds a threshold—offer a promising direction for minimizing the product of energy and latency while preserving high-value accuracy gains.

Infrastructure interoperability is a further, often underestimated challenge. Smart city deployments typically evolve through procurement cycles that yield fragmented sensor networks managed by different agencies, each with proprietary data formats and access control policies. A multi-modal LLM forecaster that requires unified and timestamp-aligned streams must therefore be accompanied by a robust data fabric layer that handles schema mapping, temporal synchronization, quality imputation, and access control enforcement. This data fabric must itself be resilient to administrative boundaries and capable of enforcing geo-specific policies—for instance, facial recognition features may need to be aggregated within a camera's jurisdiction and only anonymized count features shared with the regional forecasting model. The consequence is that the forecasting architecture cannot be designed in isolation; it must co-evolve with the city's data governance middleware, demanding joint investment by transportation departments, IT agencies, and privacy regulators.

6. Governance, Fairness, and Policy Implications

Insofar as traffic forecasts directly influence resource allocation—signal timings, congestion pricing, public transit dispatch, emergency vehicle routing—the outputs of an LLM-based multi-modal forecasting system carry distributive consequences that raise profound questions of public interest. Machine learning models are known to embed biases present in training data, and urban traffic data are no exception. Sensor density is often correlated with neighborhood income, leading to systematically lower forecast accuracy in historically under-invested areas. When such forecasts guide adaptive signal control or dynamic lane assignments, the resulting decisions can entrench mobility inequities, privileging well-instrumented arterials over underserved neighborhoods. Fairness in algorithmic systems is not merely a technical metric to be optimized post-hoc but a structural property that must be negotiated through inclusive design processes [17]. For a city-scale traffic forecasting deployment, this implies that the spatial distribution of sensor investments, the choice of accuracy objectives, and the thresholds for model updates must be subjected to public scrutiny and ongoing audit.

The multi-modal nature of the data intensifies these governance concerns because modalities differ in their privacy implications. GPS traces and camera-derived re-identification features are far more sensitive than anonymized loop detector counts, yet both may improve forecast accuracy. Edge intelligence architectures can partially reconcile accuracy and privacy by processing sensitive modalities locally and sharing only privacy-preserving embeddings or aggregated statistics with higher tiers, aligning with the principle of data minimization enshrined in regulations such as the GDPR [20]. However, such architectural choices must be codified in enforceable data governance agreements that bind all parties in the supply chain, from sensor manufacturers to cloud service providers. Mechanisms for algorithmic transparency—such as model cards, data sheets, and impact assessments—need to be integrated into the lifecycle management of the forecasting infrastructure to allow independent auditors and citizen watchdogs to verify compliance.

Accountability for forecast-driven decisions presents a further policy challenge. If an LLM-based multi-modal predictor systematically underestimates travel demand on a critical evacuation route, leading to delayed emergency response during a natural disaster, attributing responsibility across model developers, system integrators, edge hardware providers, and municipal operators is legally ambiguous. This ambiguity can be reduced by establishing clear service-level objectives that specify not only average accuracy but also worst-case tail performance on equity-sensitive corridors, and by mandating fallback to simpler, more interpretable models when the LLM's uncertainty estimates exceed pre-defined bounds. In effect, the forecasting system must be embedded in a governance framework that treats it as a component of critical public infrastructure, subject to reliability standards comparable to those applied to power grids and water systems. Developing such standards in the rapidly evolving landscape of LLM research requires proactive collaboration between academic researchers, city planners, standards bodies, and civil society.

7. Sustainability and Robustness

Sustainability in this context encompasses both environmental and operational longevity dimensions. The trend toward ever-larger foundation models must be reconciled with the imperative to reduce the carbon intensity of urban computing. On the model supply side, techniques such as sparse mixture-of-experts architectures and retrieval-augmented generation can dramatically reduce the number of parameters activated per inference, making it feasible to run diverse traffic forecasting tasks on shared edge infrastructure. On the demand side, a continuous training paradigm that incrementally updates an LLM with streaming data rather than periodically retraining from scratch can amortize the training energy across a model's usable lifetime, aligning with circular economy principles. Nevertheless, rigorous lifecycle assessments are needed to ensure that the operational gains in traffic efficiency—such as reduced idling and fewer stops—outweigh the global warming potential of manufacturing and powering the distributed computing fabric over multi-year horizons [18].

Robustness against distributional shifts and adversarial perturbations must be engineered into the system from the ground up. Traffic dynamics can change abruptly due to construction, special events, or extreme weather, causing the joint distribution over multi-modal input channels to shift in ways that violate the independent and identically distributed assumptions underpinning standard LLM fine-tuning. Benchmarks designed to measure in-the-wild distribution shifts, such as those introduced for computer vision and natural language processing, have yet to be fully adapted to spatiotemporal forecasting, but the underlying principle of evaluating models under realistic covariate and label shifts is directly applicable [19]. A robust edge LLM architecture would need to integrate continual test-time adaptation—using lightweight monitoring modules that detect statistical drift in incoming sensor streams and trigger local model recalibration or fallback to a safe, interpretable baseline. Additionally, the system must be hardened against adversarial inputs, including intentionally manipulated camera frames or falsified GPS breadcrumbs that aim to degrade forecast quality and, through it, influence congestion pricing or traffic enforcement actions. Designing certified robustness guarantees for multi-modal LLM forecasts remains an open research frontier with significant implications for public trust in AI-augmented urban management.

8. Conclusion

Multi-modal network traffic forecasting powered by large language models and deployed through edge intelligence represents a transformative architectural paradigm for smart cities,

yet it also surfaces a dense web of system-level interdependencies that demand interdisciplinary scrutiny. This paper has argued that the migration from task-specific, sensor-bound forecasters toward general-purpose LLM backbones capable of ingesting diverse urban data modalities promises substantial gains in predictive coverage, transferability, and operational flexibility. At the same time, realizing these gains in a manner that is simultaneously performant, private, fair, robust, and sustainable requires a deliberate shift from point-solution design toward integrated socio-technical infrastructure planning. Edge intelligence provides a critical enabling layer by distributing the high computational costs of LLM inference across a hierarchy of nodes, thereby bounding latency, preserving data sovereignty, and enabling localized adaptation. Yet the resulting architectural degrees of freedom—where to split the model, how frequently to aggregate updates, which data to process locally, and how to allocate energy—cannot be resolved through algorithmic optimization alone; they are fundamentally governance questions that intersect with public values, regulatory frameworks, and long-term infrastructure investment strategies. Urban computing research has long emphasized that the technical and the social are co-produced [21], and the deployment of LLM-based forecasting in smart cities offers a vivid case study of this co-production. Future work should pursue open benchmarks that capture the full systems stack, multi-stakeholder design methodologies, and life-cycle evaluation frameworks that can guide cities toward equitable and resilient traffic intelligence futures.

References

1. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of Things for smart cities. IEEE Internet of Things Journal, 1(1), 22–32.

2. Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of Transportation Engineering, 129(6), 664–672.

3. Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, 54, 187–197.

4. Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (pp. 3634–3640). AAAI Press.

5. Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In International Conference on Learning Representations.

6. Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for deep spatial-temporal graph modeling. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (pp. 1907–1913). AAAI Press.

7. Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, 33, 922–929.

8. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.

9. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282). PMLR.

10. Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., Chen, P., Liang, Y., Li, Y., Pan, S., & Wen, Q. (2024). Time-LLM: Time series forecasting by reprogramming large language models. In International Conference on Learning Representations.

11. Li, Z., Xia, L., Xu, Y., Wu, L., & Zheng, Y. (2024). UrbanGPT: Spatio-temporal large language models. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM.

12. Tedjopurnomo, D. A., Bao, Z., Zheng, B., Choudhury, F. M., & Qin, A. K. (2022). A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. IEEE Transactions on Knowledge and Data Engineering, 34(4), 1544–1561.

13. Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y. C., Yang, Q., Niyato, D., & Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 22(3), 2031–2063.

14. Zhang, C., Zhang, H., Qiao, J., Li, Z., & Alouini, M. S. (2025). TIDES: Traffic Intelligence with DeepSeek Enhanced Spatial Temporal Prediction. IEEE Journal on Selected Areas in Communications.

15. Liu, W., Zheng, Z., Zhong, J., & Yang, Y. (2024). ST-LLM: Spatial-temporal large language model for traffic prediction. arXiv preprint arXiv:2401.10134.

16. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE, 107(8), 1738–1762.

17. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 149–159). PMLR.

18. Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., So, D., Texier, M., & Dean, J. (2022). The carbon footprint of machine learning training will plateau, then shrink. Computer, 55(7), 18–28.

19. Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R. L., Gao, I., Lee, T., Chi, E. A., Bui, T., Bahl, A., Sen, P., Beam, A., Stojanov, P., Maharaj, T., Yurochkin, M., … Liang, P. (2021). WILDS: A benchmark of in-the-wild distribution shifts. In Proceedings of the 38th International Conference on Machine Learning (pp. 5637–5664). PMLR.

20. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. AI Magazine, 38(3), 50–57.

21. Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology, 5(3), 1–55.