Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 4, 2025

DEVELOPING EDGE COMPUTING SOLUTIONS FOR IOT DEVICES TO REDUCE LATENCY AND ENHANCE REAL-TIME DECISION-MAKING

Rizwan Ul Haq[*1], Fateh Aman[2], Muhammad Ahmed Majeed[3], Sohaib Raza[4], Amjid Khan[5], Rahmat Hussain[6], Muhammad Kashif Majeed[7]

[*1, 2]Department of Computer Science & IT, Superior University 10 Km Lahore Rd Sargodha (40100) Punjab, Pakistan.
[3]Üsküdar Üniversity İstanbul Altunizade, Üniversite Sok. No:14, 34662 Üsküdar/ İstanbul, Türkiye.
[4]BS (Telecommunication) Department of Communication and Cyber Security BZU Multan.
[5]CEO Friends Associates Islamabad Pakistan
[6]Institute of Computer Science & Information Technology, University of Science and Technology Bannu.
[7]Faculty of Engineering Science and Technology, Iqra University, Karachi 75500, Pakistan. School of Electronic Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China.

[*1]rizwanfast@gmail.com, [2]amanmalik967@gmail.com, [3]iammuhammadahmedmajeed@gmail.com, [4]raza454sohaib@gmail.com, [5]ceo@friendsassociates.net, [6]imrahmatwazir@gmail.com, [7]mkashif@iqra.edu.pk

DOI: https://doi.org/10.5281/zenodo.15302206

Keywords
Edge Computing, Multimodal AI, Real-Time Decision-making, AI Optimization, Model Pruning, Quantization, AI Accelerators.

Article History
Received on 11 March 2025
Accepted on 11 April 2025
Published on 29 April 2025

Copyright @Author

Corresponding Author: *
Rizwan Ul Haq

Abstract

Multimodal AI integrated with Edge Computing provides better real-time decisions through their synergy by efficiently processing data nearby its origin points along with analyzing multiple input data such as images and sensors. Both technologies create essential functionality when combined for speeding up autonomous car operations and healthcare patient monitoring systems. Various features of edge devices limit their processing ability as well as spending energy and securing data. The study evaluates three optimization approaches that reduce model size through pruning and use precision quantization for accuracy reduction along with customized AI processors to boost processing speed which resolves these limitations. The purpose behind these solutions generates minimal performance loss for sophisticated AI models which execute on constrained edge devices. Various domains stand to benefit from data-driven real-time decision-making applications because of multimodal AI's power when merged with edge computing processes. Advanced hardware and software systems develop continuously which enlarges existing boundaries to produce increasingly intelligent and quick systems.

INTRODUCTION

The deployment of cloud computing supports data processing together with storage and energy management while edge computing unites the cloud computing features by placing emphasis on local processing of data. The execution speed together with necessary bandwidth requirements decrease significantly in real-time applications. The edge operates local processes through IoT gateways or routers while the cloud system handles complex processing needs [1]. The decentralized approach allows for faster operation speeds that can also function in offline mode. Edge computing implements real-time analysis of IoT sensors at manufacturing edges to process industry data as one of its main application points. Multimodal artificial intelligence refers to a system which enables artificial intelligence to merge different types of data including text, images, audio and video information.

Multimodal Artificial Intelligence system gathers different inputs to create a unified understanding which resembles how humans perceive information. Through contextual analysis this approach produces better outputs which are also more accurate [3]. An unethical and numerous model chain formation occurs through connecting different models with aligned components. Two diverse sensorial inputs focus on trees - visual pictures of tree branches with sounds of rustling leaves - to present a multisensory experience. The combination yields improved contextual understanding about the matter [4].

UNDERSTANDING MULTIMODAL AI

The integration of Edge Computing and Multimodal AI pertains to the implementation and operation of multimodal AI on edge devices. Through this convergence model edge computing benefits from its quick processing along with Multimodal AI's ability to obtain worldwide contextual data. The localized processing enables networking infrastructure to operate without transferring large datasets into the cloud because domain-specific privacy restrictions together with bandwidth constraints exist [5]. The combination allows users to receive immediate insights while helping drive quick responses that prove essential for autonomous systems running in intelligent healthcare and augmented/virtual reality environments. The integration proves suitable when multiple data varieties need real-time processing for various applications [6]. Autonomous systems perform immediate responses through analyzing multiple sensor inputs which occur in scenarios like autonomous vehicle operations with sensor data or medical device monitoring of patient vital signs and examination results. Augmented reality applications which integrate vocal commands and manual control make up the interactive features of this system. Advanced AI models encounter deployment hurdles when implemented on edge devices primarily because these devices have restricted computing power together with energy restrictions. Optimal solutions combined with advanced hardware systems represent the main way to overcome these technical obstacles [7].

Benefits of Multimodal AI

The benefits of multimodal AI stem from combining and decoding data across modalities, which gives it the following advantages against unimodal:

AI:Improved Precision and Insight:
As Multimodal AI integrates complementary information from diverse modalities, it can provide a deeper representation, resulting in better predictions and insights [8].

Enhanced Contextual Understanding:
Multimodal AI has superior contextual understanding due to the interaction of various modality of data.

Stability and dependability:
Multimodal systems are sometimes more robust with respect to noiseand missing data [9].

Natural and Intuitive Interaction:
Multimodal AI allows for more natural and intuitive human computer interaction.

Love Hate Relationship with Multimodal AI

Even though Multimodal AI has great potential, there are several challenges that stop us from implementing it fully:

Data Fusion:
It is challenging to fuse and interpret data of different types.

Solution to Challenge:
The proper link between different types of information maintains crucial importance for interpretation but strong alignment capabilities should reflect data complexity when solving this challenge [10].

Computational Complexity:
The resource constraints of devices make this technology difficult to use because of its high computation needs [11].

Lack of Data at Scale and Bias:
The collection of such datasets is usually difficult even when meeting the needs of specialty fields [12].

The existing biases in the collected datasets have the potential to become intensified through model usage and create inaccurate or unjustified results.

Interpretability and Explain ability:
Healthcare instruments specifically need total transparency due to their dependence on developed trust which makes opaque systems less likely to gain adoption. Scientists remain dedicated to researching methods that identify the reasoning basis for model prediction outputs [13].

Generalization and Transfer Learning:
Transfer learning techniques serve as possible solutions for this trend since they enable the transfer of learned information between different problems which are closely related [14].

EDGE COMPUTING THE NECESSARY PLAYER IN THE REAL-TIME DECISIONS

Proximity to Data Source:
The system achieves its main benefit from being capable of gathering data without disruption. The data aggregation methods of Edge computing brings information closer to its original sources which leads to substantial reduction in response delays. Real-Time applications require low latency since they benefit from this technology which finds its most critical use in driverless cars along with industrial automation systems and remote surgery operations [15]. Edge devices operate at the location of data resources to execute calculations before sending them thus cutting down network transmission delays. The proximity of this processing function reduces the required bandwidth consumption. Network resources remain preserved as well as communication costs decrease because only critical information or synthesized conclusions should be transferred to the cloud [16]. Such functionality proves ideal regardless of whether bandwidth levels are restricted or processing large volumes of data needs to be performed. The local handling of data at the edge improves the security measures and safeguards personal information. The act of processing sensitive information on location systems reduces the possibility of data vulnerabilities that may occur during transport or storage on cloud platforms. Natural language data requires exceptional attention to security because it contains many types of sensitive information including healthcare and financial records [17]. The computer system at the edge point stores data within its local environment which lets operators keep complete control over data access security activities.

Applications of Edge Computing in Critical Environment:
The leading benefit which edge computing provides relates to its positioning near data origin points. The processing of location-based data increases performance because it occurs near the data source [18]. Real-Time applications alongside individual automobiles and industrial automation and remote surgery require this functionality because they need immediate responses. Local processing capacity of edge devices helps organizations to handle decisions and response tasks at the site where data originates instead of sending raw data to cloud servers. The system reduces bandwidth needs because data distribution through transmission acts as an operational limitation especially in areas with minimal connection capabilities [19]. Data privacy remains secure while sensitive data transmission decreases through minimization of insecure network vulnerabilities. Edge computing technology gives programs the ability to work steadily as data connectivity becomes irregular or discontinuous which makes them operate better in complex situations [20].

Edge computing proves highly suitable for different industries since it supports numerous devices that need time-sensitive processing and reduced latency. Confidential manufacturing operation tasks like equipment fault prediction along with safety protection enhancements are managed by local real-time analysis of sensor data conducted by edge devices. The system reduces operational interruptions while allowing predictive maintenance procedures to boost system performance [21]. The process of data analysis in real-time by self-driving vehicles uses edge computing for navigation guidance and obstacle recognition and decision-making functions [22]. The necessary low-latency functionality of edge computing makes it possible to achieve secure autonomous operations without failures. Through intelligent healthcare edge computing medical staff conduct continuous patient surveillance to provide prompt personalized treatments to their patients. Various variants of intelligent security surveillance incorporate integrated AI algorithms into edge devices to instantly process video feeds for abnormality detection which enhances security operations [23]. Verifying the information at locations where data storage takes place results in a reduction of transmitted video data while simultaneously improving reaction times and decreasing bandwidth requirements [24]. An edge computing system along with integrated model operated a multimodal AI system at the same time.

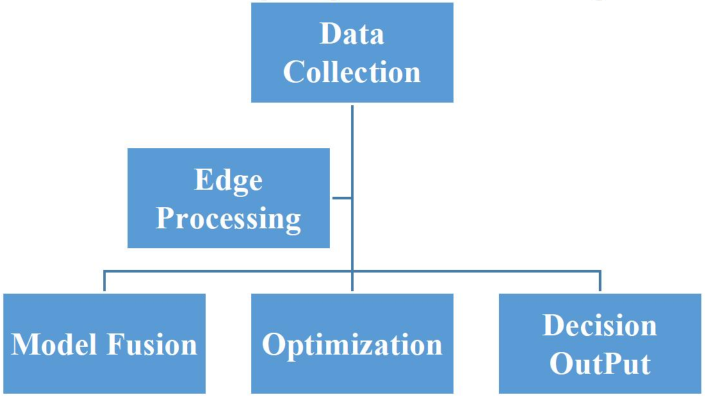
THE UTILITY OF MULTIMODAL AI IN EDGE DEPLOYMENTS

The potential of edge computing remains high yet various obstacles appear. Due to their technical constraints Edge devices contain lower processing power and reduced memory together with decreased energy capabilities compared to cloud servers. Models of artificial intelligence require several optimizations and resource-management strategies to successfully deploy them onto constrained devices [25]. Features of edge computing struggle with maintaining device security and protecting edge data because edge devices exist in dispersed locations and accessible through physical attacks. Exclusive security platforms and standardized access control methods need implementation. Many dispersed edge devices will pose management complexities when device numbers increase which warrants solutions capable of scaling edge resource deployments and management tasks. Edge computing impacts every part related to IT infrastructure regardless of its geographic location. Standardization issues with edge devices or platforms create barriers when attaining interoperability between different equipment. Devices and systems with various traits need patented methods and interfaces to exchange data while enabling communication. The reduction of dependence on cloud connection exists in Edge computing but inconsistent bandwidth continues to present technical difficulties [28]. The implementation of X2 handover within IMR requires synchronized communication systems to successfully migrate across different Cooperating Regions with varying network conditions. The concept of edge computing remains uncertain even though it commonly gets promoted throughout its growing popularity [29]. All critical applications require necessary assurance about the reliability and long-lasting performance of edge devices and applications. Operations need fault tolerance alongside redundancy measures to handle errors as preventative measures against interruptions [30].

Examples of Integration

Edge computing integration with Multimodal AI technology enables the transformation of advanced traffic management systems for smart cities. The implementation of smart traffic signal systems involves deploying edge devices across intersections to process data from visual cameras together with sensor data from traffic sensors and location information from GPS devices and textual social media feeds according to [31]. AI multimodal algorithms evaluate these data elements to optimize traffic flow before using the information to predict congestion while readjusting traffic signal timing schedules. Real-Time processing at the edge point leads to speed-efficient traffic while reducing congestion which improves transportation quality within smart cities [32]. An intelligent traffic system converts real-time video stream data from cameras into accident detection alerts which combines it with traffic monitoring sensor inputs and GPS-enabled zone identification for traffic congestion detection. These modalities enable the system to identify traffic light activation times along with directing traffic flow modifications and immediate emergency service notifications. The collaborative network creates more efficient urban traffic that reduces congestion and strengthens safety measures throughout metropolitan areas [33].

Figure 1: Workflow for real-time decision-making using multimodal AI on edge devices



Benefits of Integration

Implementation of multimodal Artificial Intelligence with edge computing provides organizations with multiple advantages including advanced real-time insights which benefit from combined capability [34]. Organisations can achieve better insight through Multimodal AI that processes diverse data types when using edge computing to deliver immediate processing for time-sensitive applications. Systems gain better and faster Real-Time decision-making capabilities through edge-based multisensory processing and analysis of data. The choice made during a critical final millisecond becomes crucial to determining survival outcomes in life-threatening safety scenarios [35].

The strategy to process sensitive data near its origin at the edge where it is created reduces visibility of the information which minimizes exposure to data breach threats that could endanger cloud-stored information. The protection of confidential data becomes crucial for applications holding sensitive material such as healthcare information as evidence indicates [36]. Edge computing processes data by its source through localized processing which reduces the transmission of essential data to the cloud that leads to decreased bandwidth expenses. Advantageous conditions exist when bandwidth limits exist or data sets grow large. Edge computing allows devices to maintain autonomous operations as they can operate independently with variable cloud connectivity. The reliability and technical stability of applications becomes stronger thanks to edge computing technology in complex or isolated areas [37]. A significant enhancement of user experience has occurred because Multimodal AI on Edge created an intuitive technological interface which allows effortless human-machine interaction. The adaptable platform allows computers to process joint voice commands and hand motions and detect visual cues which results in better user experiences through enhanced convenience [38].

Weighting AI Deployment in Edge Computing

Different challenges hinder the implementation of AI models on edge devices: Resource Constraint produces high deployment expenses for powerful AI models until operators optimize these systems thoroughly. The combination of model compression techniques with pruning and quantization methods minimizes the model size requirements as well as computing needs to maintain equivalent performance results [39]. The edge computing environment shows clear divergence because it encompasses numerous devices running different hardware systems together with unique operational capabilities. The AI models function best when having access to specialized development tools and frameworks which enable operation across different hardware systems [40]. The synchronization of decentralized edge device data structures can become difficult to manage because their coordination requires complex solutions. The required data for training and inference needs effective data pipelines joined with synchronization techniques to work properly [41]. When AI models travel to end users for deployment the security of both model data and applications must be the top priority. System security measures should be deployed to protect the system from unauthorized access and model theft while preventing data breaches. AI models deployed in the edge environment should automatically adjust their operations through monitoring active changes within their environments and data patterns. The model needs features for constant learning that should adapt its architecture to maintain performance stability throughout periods [43].

The decentralized training capabilities of federated learning address machine learning privacy issues because it enables model development without dataset transmission between edge and central server locations which creates optimal conditions for edge learning scalability [44]. Model pruning serves as an extensive compression technique used for deep learning models thus enabling large model deployment to edge devices which have limited computational capacity. Model parameters shift from their original 32-bit floating point version to lower precision formats known as quantization which affects their precision level. Model performance speeds up dramatically after parameter reduction since it requires decreased energy usage while operating on edge devices [45]. The accuracy-related decline becomes tolerable to various time-sensitive operational systems despite quantization methods. The purpose of AI accelerators lies in their capability as specialized hardware which boosts AI computation performance. The Tensor Processing Unit (TPU) along with Graphics Processing Unit (GPU) functions as computing accelerators designed specifically for executing complex calculations needed to operate artificial intelligence functions. These processors outperform generic CPU designs in their performance execution. Edge devices have better performance and better energy efficiency in executing complex AI tasks through the use of these accelerators [46]. Other innovative techniques exist along with pruning and quantization which help compress models for edge deployment.

A small student model learns to copy the behavior of a large complex teacher model through this technique. A less costly student model becomes feasible through this method because it maintains most of the teacher model's correctness [47]. Big matrices within the model can be decomposed into smaller matrices to decrease parameter costs and computational expenses through Low Rank Factorization. The privacy protection benefits of Federated Learning become critical because it enables distributed AI model training throughout different local devices called edge devices. Through this method devices conduct their own training operations before sending model parameters to their central server rather than raw data [48]. Devices maintain data privacy by using this method to draw knowledge from other devices' network-experienced information. Several applications which execute on energy-limited edge devices need optimized AI algorithms for decreasing their resource usage. Algorithms need to be designed to save power consumption or an adaptive method must be created that adjusts power consumption based on the processing requirements of tasks such as dynamic voltage and frequency scaling [49]. The goal remains to decrease power drain while keeping AI capacities operational. AI deployment efficiency becomes more possible when designers create edge-specific architecture systems for these environments. Networks of lightweight size together with hybrid inclusive models that incorporate multiple AI techniques should be used to achieve accurate and efficient solutions.

MULTIMODAL AI ON EDGE DEVICES: TECHNICAL CONSIDERATIONS AND CHALLENGES

Multi-modal artificial intelligence needs higher computational resources than single-modal artificial intelligence since it processes numerous sources simultaneously. The high computational requirements of these models create an essential barrier when trying to deploy complex models onto hardware with limited resources [51].

One major limitation of implementing AI at the edge involves power consumption because the combination of high energy requirements and frequent communication practices can drain the restricted battery power of different edge devices. Proper performance needs energy-efficient algorithms together with hardware acceleration to increase device operational life span.

Warmer processing of sensitive healthcare information at the edge reduces network-based data risks yet requires sufficient protection mechanisms on edge tools to stop breaches. The correct interpretation of multimodal analysis demands synchronized data from different modalities according to both time and contextual relationship [53]. Time discrepancies between audio-video data combinations create errors that subsequently produce ambiguous results because timing functions as an essential data property. Complexities rise as algorithms need better regulation of their operating times. The following difficulties require innovative solutions which include quantization and pruning techniques for model compression to lower multimodal AI model requirements before edge deployment.

Future Trends and Research Directions

Efficient algorithms for multimodal fusion will be developed to improve performance on edge devices while minimizing resource usage according to research [54]. The deployment of planetary AI in edge computing environments becomes feasible and enables precise real-time decisions through complex models because 5G networks and subsequent networks will provide better bandwidth along with faster latency speeds. Edge computing with multimodal AI signals the advent of multiple operational fields including self-driving vehicles and personalized medical solutions and virtual reality enhancement. Edge device applications continue to extend because of diminishing technical limitations [56]. The existing research prerequisites need additional attention despite recent progress. Enhanced algorithm development needs to improve multimodal fusion efficiency while creating deployment energy saving techniques along with protection measures for edge AI systems. Edge device capacity will grow and new applications will emerge because of these developed improvements.

Conclusion

The coupling of multimodal AI systems with edge computing technology enables immediate decision processes for diverse range of applications. Resource needs of multimodal AI models present two significant problems because these models do not work effectively on edge devices that have limited processing capabilities. For edge deployment suitability model compression relies on methods which include pruning along with quantization and knowledge distillation and low-rank factorization techniques. The processing rate can be boosted through hardware systems that include GPUs and TPUs. Edge devices have restricted processing power which gets exhausted quickly through the execution of detailed AI models. The improvement of battery longevity depends on electricity-efficient algorithms and hardware in combination with dynamic voltage and frequency scaling along with other techniques. The protection of confidential information has become essential when companies process sensitive data through edge devices since security remains a primary concern for every business. The model training process under federated learning takes place within individual devices before transmitting only the learned parameters to a central server for distribution. Well-synchronized multimodal datasets from several sensors. Timing variances create challenges that need complex algorithms to resolve them before successful data fusion can take place. The solution to these challenges will be simple through new AI algorithm and hardware systems and security standards development. The combination of Multimodal AI and edge computing technology will create revolutionary changes in healthcare systems as well as autonomous systems and smart city applications to power real-time decision-making processes of the future.

REFERENCES

1. [2301.00774v3] SparseGPT: Massive Language Models Can Be Accurately Pruned in OneShot. (n.d). https://arxiv.org/pdf/2301.00774v3.pdf
2. 2019], [O 4 N. (2019, November 4). Smart City Intelligent System Traffic Congestion Optimization using Internet Of Things. https://arxiv.org/abs/1911.01286
3. 2020, M R O D 1. (2020, December 14). MultiModal OnDevice AI: Heterogeneous Computing Once More?
4. 2020, M R O D 1. (2020, December 14). MultiModal OnDevice AI: Heterogeneous Computing Once More?. https://www.sigarch.org/multimodalondeviceaiheterogeneouscomputingoncemore/
5. Ahmed, S N P W. (2021, July 1). A Review on Edge Analytics: Issues, Challenges, Opportunities, Promises, Future Directions, and Applications. https://arxiv.org/pdf/2107.06835v1.pdf

6. B, A L O O P R S A S S G. (2023, January 1). Edge AI: A survey. https://www.sciencedirect.com/science/article/pii/S2667345223000196

7. Bagchi, S., Siddiqui, M., Wood, P., & Zhang, H. (2019, December 20). Dependability in edge computing. Association for Computing Machinery, 63(1), 5866. https://doi.org/10.1145/3362068

8. Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2021, June 10). A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. Springer Science+Business Media, 38(8), 29392970. https://doi.org/10.1007/s00371021021667

9. Bhardwaj, K. (2019, October 23). EdgeAI: A Vision for Deep Learning in IoT Era. https://deepai.org/publication/edgeaiavisionfordeeplearninginiotera

10. Bilal, K., Khalid, O., Erbad, A., & Khan, S U. (2017, October 18). Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers. Elsevier BV, 130, 94120. https://doi.org/10.1016/j.comnet.2017.10.002

11. Cao, K., Liu, Y., Meng, G., & Sun, Q. (2020, January 1). An Overview on Edge Computing Research. Institute of Electrical and Electronics Engineers, 8, 8571485728. https://doi.org/10.1109/access.2020.2991734

12. Das, A., Dash, P K., & Mishra, B K. (2017, December 30). An Innovation Model for Smart Traffic Management System Using Internet of Things (IoT). Springer International Publishing, 355370. https://doi.org/10.1007/9783319706887_15

13. Frantar, E., & Alistarh, D. (2023, January 1). SparseGPT: Massive Language Models Can Be Accurately Pruned in OneShot. Cornell University. https://doi.org/10.48550/arxiv.2301.00774

14. Gossett, S. (2022, May 1). What Is Edge Computing?. https://builtin.com/cloudcomputing/whatisedgecomputing

15. Gupta, K., & Shukla, S. (2016, February 1). Internet of Things: Security challenges for next generation networks. https://doi.org/10.1109/iciccs.2016.7542301

16. Hasan, M T., Hossain, M A E., Mukta, M S H., Akter, A., Ahmed, M., & Islam, S. (2023, May 11). A Review on DeepLearningBased Cyberbullying Detection. Multidisciplinary Digital Publishing Institute, 15(5), 179179. https://doi.org/10.3390/fi15050179

17. Hilmani, A., Maizate, A., & Hassouni, L. (2020, September 11). Automated Real-Time Intelligent Traffic Control System for Smart Cities Using Wireless Sensor Networks. Wiley, 2020, 128. https://doi.org/10.1155/2020/8841893

India, S J A K M P S U C. (2019, November 5). Smart City: Traffic Management System Using Smart Sensor Network.

19. Klas, G. (2017, January 1). Edge Computing and the Role of Cellular Networks. IEEE Computer Society, 50(10), 4049. https://doi.org/10.1109/mc.2017.3641649

20. Luo, R C., & Kuo, C. (2016, March 24). Intelligent SevenDoF Robot With Dynamic Obstacle Avoidance and 3D Object Recognition for Industrial Cyber–Physical Systems in Manufacturing Automation. Institute of Electrical and Electronics Engineers, 104(5), 11021113. https://doi.org/10.1109/jproc.2015.2508598

21. Miz, V., & Hahanov, V. (2014, September 1). Smart traffic light in terms of the cognitive road traffic management system (CTMS) based on the Internet of Things. https://doi.org/10.1109/ewdts.2014.7027102

22. Mohsen, F., Ali, H., Hajj, N E., & Shah, Z. (2022, October 26). Artificial intelligencebased methods for fusion of electronic health records and imaging data. Nature Portfolio, 12(1)

23. Mohsen, F., Ali, H., Hajj, N E., & Shah, Z. (2022, October 26). Artificial intelligencebased methods for fusion of electronic health records and imaging data. Nature Portfolio, 12(1). https://doi.org/10.1038/s41598022225144

24. Narejo, S., Pandey, B., Esenarro, D., Rodríguez, C., & Anjum, M R. (2021, May 11). Weapon Detection Using YOLO V3 for Smart Surveillance System. Hindawi Publishing Corporation, 2021, 19. https://doi.org/10.1155/2021/9975700

25. Nikolopoulos, B V W B K S. (2016, September 7). Challenges and Opportunities in Edge Computing. https://arxiv.org/pdf/1609.01967v1.pdf

26. Nyle, D R S N S. (2021, November 22). The Benefits of Edge Computing in Healthcare, Smart Cities, and IoT. https://arxiv.org/abs/2112.01250

27. Palla, T. (2021, May 29). INTELLIGENT TRAFFIC MANAGEMENT USING BIG DATA ANALYTICS AND IOT. https://tharunpalla43.medium.com/intelligenttrafficmanagementusingbigdataanalyticsandiota5e841f2cf7d?source=post_internal_links2

28. Pearson, S. (2012, June 27). Privacy, Security and Trust in Cloud Computing. , 342. https://doi.org/10.1007/9781447141891_1

29. Projects, C T W. (2006, May 24). Edge computing. https://en.wikipedia.org/wiki/Edge_computing

30. Sethi, P S. (2022, April 11). Opportunities and Challenges in Edge Computing Parminder Singh Sethi Medium. https://medium.com/@singhsethi.parminder/opportunitiesandchallengesinedgecomputing76ff4643edcb?source=read_next_recirc03994cd49_e943_4d72_a119_26b554fd4e95

31. Shi, D., Tao, C., Jin, Y., Yang, Z., Yuan, C., & Wang, J. (2023, January 1). UPop: Unified and Progressive Pruning for Compressing Vision Language Transformers. Cornell University. https://doi.org/10.48550/arxiv.2301.13741

32. Sujata, B E B. (2020, December 10). Artificial Intelligence at the Edge. https://arxiv.org/abs/2012.05410

33. Summaira, J., Li, X., Shoib, A M., Li, S., & Abdul, J. (2021, January 1). Recent Advances and Trends in Multimodal Deep Learning: A Review. Cornell University. https://doi.org/10.48550/arxiv.2105.11087

34. Traffic Management Case Studies | SMATS. (2020, November 4). https://www.smatstraffic.com/casestudies

35. Trihinas, D., Thamsen, L., Beilharz, J., & Symeonides, M. (2022, September 1). Towards Energy Consumption and Carbon Footprint Testing for AIdriven IoT Services. https://doi.org/10.1109/ic2e55432.2022.00011

36. Varghese, B., Lara, E D., Ding, A Y., Hong, C., Bonomi, F., Dustdar, S., Harvey, P., Hewkin, P., Shi, W., Thiele, M., & Willis, P. (2021, July 1). Revisiting the Arguments for Edge Computing Research. IEEE Computer Society, 25(5), 3642. https://doi.org/10.1109/mic.2021.3093924

37. Wang, Y J. (2011, September 1). A City Intelligent Transportation Management Command System. Trans Tech Publications, 9798, 863866. https://doi.org/10.4028/www.scientific.net/amm.9798.863

38. Wasserblat, O Z L B S. (2021, November 10). Prune Once for All: Sparse PreTrained Language Models. https://arxiv.org/pdf/2111.05754v1.pdf

39. What is edge computing?. (2023, February 6). https://www.redhat.com/en/topics/edgecomputing/whatisedgecomputing

40. What Is Multimodal AI? (2023, July 10). https://app.twelvelabs.io/blog/whatismultimodalai

41. Wu, J., Gan, W., Chen, Z., Wan, S., & Yu, P S. (2023, December 15). Multimodal Large Language Models: A Survey. https://doi.org/10.1109/bigdata59044.2023.10386743

42. Xu, D., Li, T., Li, Y., Su, X., Tarkoma, S., & Hui, P. (2020, March 26). A Survey on Edge Intelligence

43. Xu, D., Li, T., Li, Y., Su, X., Tarkoma, S., & Hui, P. (2020, March 26). A Survey on Edge Intelligence. http://dblp.unitrier.de/db/journals/corr/corr2003.html#abs200312172

44. Xu, U R H G A H E S Z. (n.d). A Case for Elevating the Edge to be a Peer of the Cloud. https://dl.acm.org/doi/10.1145/3447853.3447859

45. Yan, G., & Qin, Q. (2020, January 1). The Application of Edge Computing Technology in the Collaborative Optimization of Intelligent Transportation System Based on Information Physical Fusion. Institute of Electrical and Electronics Engineers, 8, 153264153272

46. Yan, G., & Qin, Q. (2020, January 1). The Application of Edge Computing Technology in the Collaborative Optimization of Intelligent Transportation System Based on Information Physical Fusion. Institute of Electrical and Electronics Engineers, 8, 153264153272. https://doi.org/10.1109/access.2020.3008780

47. Yao, Z., Ma, L., Shen, S., Keutzer, K., & Mahoney, M W. (2021, May 30). MLPruning: A Multilevel Structured Pruning Framework for Transformerbased Models. Cornell University. http://export.arxiv.org/pdf/2105.14636

48. Zhang, C., Yang, Z., He, X., & Deng, L. (2020, March 1). Multimodal Intelligence: Representation Learning, Information Fusion, and Applications. Institute of Electrical and Electronics Engineers, 14(3), 478493

49. Zhang, C., Yang, Z., He, X., & Deng, L. (2020, March 1). Multimodal Intelligence: Representation Learning, Information Fusion, and Applications. Institute of Electrical and Electronics Engineers, 14(3), 478493. https://doi.org/10.1109/jstsp.2020.2987728

50. Zhang, M. (2020, October 17). Deep Learning in the Era of Edge Computing: Challenges and Opportunities

51. Zhang, M. (2020, October 17). Deep Learning in the Era of Edge Computing: Challenges and Opportunities. https://deepai.org/publication/deeplearningintheeraofedgecomputingchallengesandopportunities

52. Zhang, Z Z X C E L L Z K L J. (2023, November 10). Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing

53. Zhang, Z Z X C E L L Z K L J. (2023, November 10). Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing. https://ieeexplore.ieee.org/document/8736011/

54. Zhu, M., & Gupta, S. (2017, January 1). To prune, or not to prune: exploring the efficacy of pruning for model compression. Cornell University. https://doi.org/10.48550/arxiv.1710.01878

55. Zomaya, S D H Z W F J Y S D A Y. (2023, December 4). Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence

56. Zomaya, S D H Z W F J Y S D A Y. (2023, December 4). Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. https://ieeexplore.ieee.org/document/9052677/