

Edge AI

In today's world, where we rely heavily on data, it is truly clear that technology needs to become smarter and more capable. Undoubtedly, edge computing and artificial intelligence stand out as powerful and advanced technologies in our current era. Edge computing refers to the computing that takes place at the edge of a network. Here edge is the point from where the data will be delivered or processed. And artificial intelligence (AI) means creating machines capable of performing tasks that need human intelligence. Both play a vital role on the Internet of Things (IOT) ecosystem. However, with technological advancements, there is a growing preference for applied AI i.e., Edge AI that enables real-time decision making. Edge AI is simply the combination of Edge Computing and Artificial Intelligence. Edge AI refers to the practice of processing and analyzing data at the edge of a network closer to the source by intelligent edge devices with built-in processing power and AI capabilities. It means that AI algorithms and capabilities are directly deployed on edge devices such as sensors, smart phones and other IOT devices. And the data storage and data computation are done locally at the device where the data is being produced without having sent it to cloud. By processing data at edge, IOT devices can make real-time decisions with high responsiveness and reduce the dependency on cloud services. The major benefit of Edge AI over other computing applications is that Edge AI enables real-time data processing and decision-making locally on the device itself.

Edge AI Architecture Overview

Edge AI, also known as Edge Intelligence, brings intelligence to IOT sensors and devices so they can check and analyze data in real time. It combines AI algorithms, machine learning frameworks and specific hardware to support AI interference on edge devices. The fundamental components of edge AI are datasets, AI models, AI applications and hardware components. The architecture includes the following processes:

a. Data Acquisition

It is the process of gathering information from different devices and sensors. In context of Edge AI, data acquisition is done by edge devices using cameras, sensors and other sources at the edge of network. Sensors are embedded on edge devices to collect real-time data from the physical world. There are diverse types of sensors to collect various data. Cameras are used to collect visual information. Accelerometers and gyroscopes are used for motion detection. Microphones are used to capture audio data. Temperature Sensors are used to check climate conditions. Similarly, there are different other sensors sensitive to environmental changes used to capture data. Light sensors, biometrics sensors, gas sensors are other few examples of sensors used in data collection in Edge AI architecture.

b. Data Transmission

It is the process of sending collected data from sensors to other components in the network. In edge AI, data transmission process consists of hardware components and communication protocols. Various hardware components like Arduino and Raspberry Pi are popularly used for data transmission. These hardware components are connected with various communication protocols such as Wi-Fi module, Bluetooth, GSM module,

LoRa for effective transmission of data. They have board processing capabilities and connectivity options like HDMI, Ethernet, I2C and GPIO pins. Here the function of hardware component is sensor interfacing and data transmission through communication protocols. Wi-Fi module enables wireless communication while GSM module is used to communicate to mobile devices.

c. **Data Processing & Management**

In edge AI, data processing is directly done on devices itself. The process involves deployment of machine learning algorithms on edge devices. This is done by using chips like Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). Here TPUs and GPUs are special processors used to process AI and machine learning tasks locally. These chips perform computations very quickly. They can even perform complex and multiple computations at a time. This allows quick processing and quick results. The chips also respond according to the AI trained model. The processing unit also involves noise reduction, data augmentation and normalization here. The processed data are stored on the devices itself locally. After the processing, the trained AI model generate predictions and produce outcomes according to their training. Edge devices can use applications or databases to store data.

d. **Data Analytics & Visualization**

Edge devices are equipped with AI interference which helps to analyze meaningful data and extract meaningful insights from it. AI algorithm running on devices analyzes pre-processed data in real time to make decisions promptly. Then the device produces useful insights after interpreting the results produced from AI's analysis. Tools like Apache Spark, Edge ML and TensorFlow lite are used for edge analytics. Data analytics helps to organize and interpret data to find patterns and trends. And those findings are represented in charts, graphs or dashboards to understand complex information. Tools like Tableau and Power Bi are used for data visualization.

Edge AI Real World Applications

Edge AI has potential to transform many industries such as Healthcare, Transportation, manufacturing and a lot. There are diverse range of applications for Edge AI across different industries. The Edge AI solutions are very capable and being widely used in our daily lives. Various smart devices and wearables that can fit in our palm are equipped with amazing capabilities due to Edge AI. In the domain of healthcare, remote surgery is one of the examples of Edge AI applications which could be beyond imagination if there was no advancement in Edge AI. Hospitals are using different Edge AI devices for Remote Patient Monitoring System (RPMS). Wearable devices equipped with edge AI capabilities such as are used for real-time analysis of health data of patients. Hospitals use AI equipped medical imaging devices to quickly analyze X-rays, MRI and other diagnostic images. Thermal imaging cameras are used to quickly measure body temperatures. Smart watches & fitness bands provide real time health and fitness insights without connectivity. Smart home devices like voice assistants, thermostats, cameras are used widely. Smart surveillance cameras are used for traffic management and public safety. Edge AI systems are used in the retail industry for inventory management and real-time customer analytics. In agriculture, edge AI applications include analyzing soil condition and monitoring crop health

for precision farming. Farmers can employ Edge AI on devices like drones to analyze data in real-time. This helps them make prompt decisions on when to irrigate, apply pest control, or take other measures to enhance crop yields. Quick decision-making with this technology contributes to improved agricultural outcomes. AI-enabled technologies have been very crucial in poultry health and welfare management (Ojo et al., 2022). Edge AI helps farmers to verify early diseases of animals and make farm management efficient. Edge AI even has a significant role in the education sector. Smart classroom, real-time feedback on the assessments & facial recognition for attendance of students are some popular edge AI applications used by colleges and institutions in this field. Edge AI has also been very crucial in the banking sector. The applications of edge AI in banking sector include ATM security, live financial markets, biometric authentication. It is also used for real-time analysis of transactions that helps to prevent frauds. One of the most notable and impactful applications of edge AI can be seen in the field of transportation. Chatbots are being used by many organizations to enhance customer support. Undoubtedly, the most ground-breaking technology of our era is autonomous vehicles. Self-driving cars use object detection and recognition to detect road conditions and produce real-time decision-making itself without depending on cloud processing. This doesn't only show a lack of dependence on cloud servers but also ensures instant response in their decision-making following safety and responsiveness.

Benefits of Edge AI

Edge AI is beneficial for its outstanding features like reduced latency, real-time processing, increased efficiency and improved privacy. Edge AI solutions used in different scenarios help in different areas like security, safety and easiness in our life. Since there would be no requirement to communicate this data to the internet in Edge AI architecture, we don't have to worry about latency or wider internet connectivity. This enables offline functionality on devices. Edge AI frameworks are compatible with a variety of devices such as sensors, cameras, smartphones, and IoT devices. This feature allows them to be used in various kinds of hardware configurations. Deployment of encryption, authentication and authorization on edge devices prevent data breaches. There is no need for transmission of enormous amounts of data to centralized servers on Edge AI which allows efficient use of bandwidth. By processing data on devices, it is more energy efficient compared to other computing architectures. Sensitive information can be stored in the device itself securely. Edge AI can prevent people from different accidents and diseases by making it valuable in the field of healthcare and safety. Different autonomous devices can be used for the prevention of accidents and dangers by alerting people if devices detect dangers. Alarms and alerting systems are extremely useful that uses edge AI. Suppose sensors in valves at a petroleum pipeline detect dangerously high pressure and shutoff needs to be triggered as soon as possible. Sensor with AI capabilities analyzes the pressure data taking and shuts off the valve automatically if it needs to be. So, if it was on cloud, it would need to communicate to the centralized server first. The shutoff instructions might have come too late. Until then the accident could have already taken place. Hence instant decision-making and response makes Edge AI more applicable than cloud computing AI and other architectures.

Challenges and Future Directions

Aside from advantages, there are also challenges faced in Edge AI. Most of the edge devices have battery as their power source but running AI algorithms on those devices might drain batteries too fast. It has limited computational power. The main problem with Edge AI is that the smart algorithms and their valuable information are associated in an environment we can't control or trust. Edge AI applications might contain personal data that could be very vulnerable to attackers. Also, AI models require time to time maintenance and updates which might be a challenge concern. Edge AI hardware security is also one of the major security concerns here. People could damage the hardware to exploit and gain unauthorized access. The edge AI solutions might not be applicable for every place. Like autonomous vehicles could not be suitable in our country Nepal due to its difficult road conditions having uphill and downhill roads.

These challenges can be overcome with advancement in hardware and network components. Edge AI can be powerful when edge cloud operation is implemented. We could see many powerful IOT devices in the coming days in future. Devices like our phone and watch can become smarter than now. Robotic Process Automation (RPA) is predicted to have more edge applications in the future. Federative Learning techniques is also considered to be more in practice in coming days. Edge-to-Edge communication is at initial stage where edge devices communicate with each other directly with improved security and enhanced user experience. The future of Edge AI looks better in healthcare as robotic surgeries and different other advanced technologies are upgrading day by day. The world will adopt Edge AI in future everywhere for different things. Rise of super applications and edge devices integrated with 5G technology are being breakthrough technologies. In future, edge AI will create numbers of research possibilities and find applications in different sectors like smart IOTs, advanced manufacturing and efficient urban planning for smart city (Elmoulat et al., 2020).

Conclusion

Combining edge computing and AI together is like a big step forward in technological revolution. We could already see so many powerful and intelligent devices at the earliest stage of its revolution. From this we can expect more possibilities of AI-driven solutions. Edge AI is making our world more connected and intelligent, where devices are highly intelligent and perform tasks beyond our imaginations. Secure embedded systems, data encryption, hardware-level security, and anti-tampering measures must be implemented in edge AI ecosystems to make it more effective and secured.

References

- Debauche, O., Mahmoudi, S., Mahmoudi, S., Manneback, P., Bindelle, J. and Lebeau, F. (2020) Edge Computing and Artificial Intelligence For Real-time Poultry Monitoring. *Procedia Computer Science* [online]. 175, pp. 534-541. [Accessed 02 January 2024].
- Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S. and Zomaya, A. (2020) Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. *IEEE internet of things journal* [online]. 7 (8), pp. 7457–7469. [Accessed 02 January 2024].
- Elmoulat, M., Debauche, O., Mahmoudi, S., Mahmoudi, S., Manneback, P. and Libeau, F. (2020) Edge Computing and Artificial Intelligence For Landslides Monitoring. *Procedia Computer Science* [online]. 177, pp. 480-487. [Accessed 02 January 2024].
- Wang, X., Han, Y., Leung, V., Niyato, D., Yan, X. and Chen, X. (2020) *Edge AI: Convergence of Edge Computing and Artificial Intelligence* [online]. 1st ed. Singapore: Springer Singapore Pte. Limited. [Accessed 02 January 2024].
- Zhu, Y. (2021) Network Public Opinion Prediction and Control Based on Edge Computing and Artificial Intelligence New Paradigm. *Wireless Communications and Mobile Computing* [online]. 2021, pp. 1-11. [Accessed 02 January 2024].