YAHOO! JAPAN ニュース

Japan-born AI Chip Startup EdgeCortix Raises Funds from Saudi Arabia as well

By Kenji Tsuda

A Japanese AI chip startup, EdgeCortix, was selected by the Ministry of Investment of the Kingdom of Saudi Arabia in February and has been chosen to participate in the country's 'National Semiconductor Hub'. EdgeCortix is a distinguished Japanese startup based in Ginza, Tokyo. However, it has quickly taken steps that could serve as a valuable reference for current Japanese semiconductor and system companies, such as securing funding from overseas and establishing international branches. Let me briefly introduce the company.

EdgeCortix is a fabless semiconductor company founded in 2019 by Dr. Sakyasingha Dasgupta, who has experience at the RIKEN Institute and the Max Planck Institute in Germany (Figure 1). Commonly known as Sakya, the CEO of EdgeCortix, is widely recognized for his commitment to fostering a diverse and inclusive environment, valuing talent above all, irrespective of nationality.

Although its headquarters is located in Ginza, Tokyo, the actual chip development and engineering work is carried out in their lab, in Musashi-Kosugi, Kawasaki City.

Figure 1: EdgeCortix Founder & CEO, Dr. Sakyasingha Dasgupta, commonly known as Sakya. (Photo taken by the author)

The semiconductor chip being developed is a high-performance AI chip designed for edge applications. Initially, to verify the operation of the AI chip with the algorithm they had envisioned, a prototype was made using FPGAs (an IC that can be programmed to create custom circuits). In fact, the prototype edge AI chip named SAKURA-I passed a radiation exposure reliability testing conducted by NASA in the United States without any malfunctions. In conventional semiconductor chips (especially memory and logic), exposing them to radiation could cause a bit to flip from state 1 to state 0 (or vice versa). This phenomenon is called a soft error, as it can be corrected by resetting the system when the power is turned off. SAKURA-I has also undergone proton (ion) and heavy ion irradiation tests at Massachusetts General Hospital and Lawrence Berkeley National Laboratory to verify its reliability. The results from these tests revealed that SAKURA-I is highly resistant to failure, even in environments with radiation, such as outer space.

Until now, AI chip development has been divided into chips for data centers and chips for devices at the edge, such as computers and smartphones. Companies designing chips for data centers include NVIDIA, Intel, and AMD, as well as startups like Cerebras, SambaNova, and Esperanto. For edge applications, there are companies such as Tenstorrent and EdgeCortix. In general, data center chips consume a large amount of power, making them difficult to use in outer space. In contrast, edge AI chips are particularly effective in space.

EdgeCortix's chip, SAKURA-I, does not use error correction circuits like ECC, nor does it employ a redundant configuration. However, it has a self-healing function, using features such as self-refresh to address data inversion caused by radiation. While the details of the technology have not been disclosed, conceptually, the system monitors data via software and, if an error occurs, traces the signal path to reproduce the correct data in normal mode.

Furthermore, this function is also included in the dedicated ASIC chip, SAKURA-II, which is also expected to have strong radiation resistance. SAKURA-II has a power consumption of 8W and achieves 60 TOPS (Trillion Operations Per Second) - they plan to develop even higher AI performance with the upcoming SAKURA-X series. By the way, Microsoft defines the AI PC feature, Copilot+, as having a performance of over 40 TOPS.

In November 2024, EdgeCortix secured a subsidy of 4 billion yen through NEDO (New Energy and Industrial Technology Development Organization) to develop the SAKURA-X series. Moving forward, the company aims to apply this technology in areas such as AI-RAN for edge base stations in 5G Advanced and 6G, as well as enhancing the capabilities of robots. At SEMICON Japan, they also demonstrated several examples. These include an AI model that generates captions to describe in real-time images from videos (Figure 2), an AI model that recognizes people and measures their temperature (body temperature) and distance in the depth direction (Figure 3), as well as a text-based example of a chatbot similar to ChatGPT.

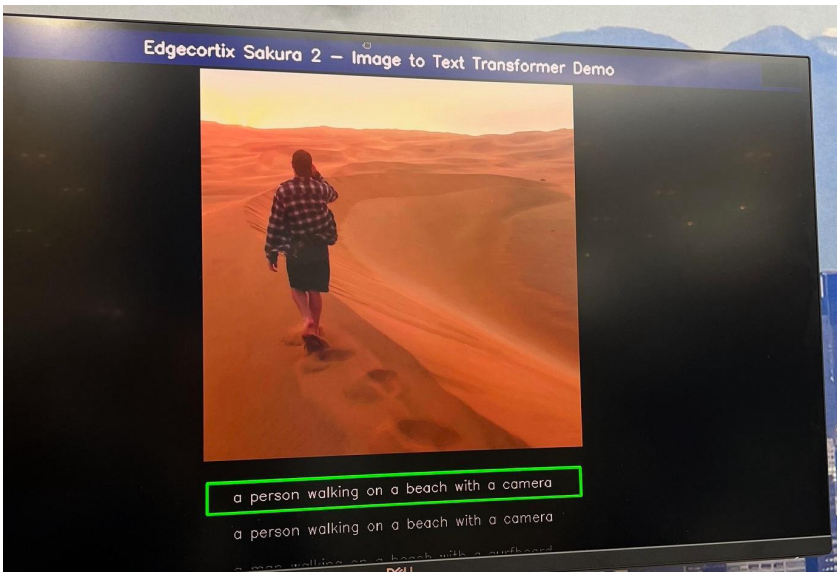

Figure 2: Displaying what the video depicts as text. (Photo taken by the author)

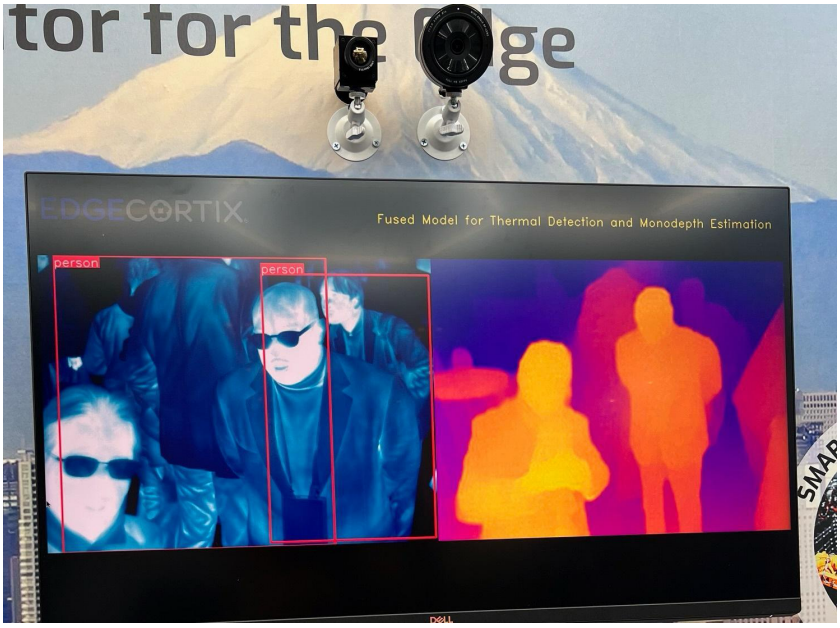

Figure 3: Recognizing people in the video and thermal detection and measuring their depth only using only an RGB camera. (Photo taken by the author)

Startups need to secure funding while developing their products. In this case, Saudi Arabia's Ministry of Investment has selected EdgeCortix's AI chip, and the company will participate in the 'National Semiconductor Hub (NSH)' that began in 2024. The Saudi government is investing 1 billion SAR (Saudi Riyals), approximately 40 billion yen, in this hub. There are plans to train 5,000 IC circuit design engineers by 2030. As part of Saudi Arabia's Vision 2030, the goal is to establish 50 fabless semiconductor companies, including through attraction, by 2030.

In this way, the Japanese startup has not only secured funding from initial investors such as SBI Investment and Renesas Electronics but has also raised funds from NEDO and Saudi Arabia. Moreover, it is actively exploring partnerships with potential users, including NASA. Of course, they are also focusing on technology development and have presented a roadmap for the development of new edge AI chips. They are said to be adopting a new data flow architecture as well. Their strategy of balancing both technology and marketing seems to follow a similar path to that of NVIDIA. We will be watching EdgeCortix's progress closely.

Translation prepared by EdgeCortix

- Yahoo Japan (2025/02/03). Full original Japanese article: https://news.yahoo.co.jp/expert/articles/5373795c222952d42595945bdf35e76b63c57d44