Presentation Title: Lightweight AI and Sustainable Generative AI

Speaker: Prof. Young-Im Cho (Department of Computer Engineering, Gachon University, Republic of Korea)

Abstract:

The rapid advancement of generative artificial intelligence (Generative AI) has led to significant technological innovations while simultaneously raising critical challenges. These include increased computational requirements, substantial energy consumption, heightened environmental impacts, and growing ethical concerns. This presentation introduces Lightweight AI as a core technical approach to mitigating these challenges and provides an overview of the current progress in relevant international standardization activities.

Specifically, the presentation will focus on the initiatives undertaken by ISO/IEC JTC 1/SC 42, with particular emphasis on ISO/IEC TS 42111 (Guidance on Lightweight AI Systems) and ISO/IEC TS 25258 (Hybrid AI Inference Framework). These standards offer structured methodologies for the design, development, and evaluation of lightweight AI models that are optimized for energy efficiency, minimal resource consumption, and sustainable deployment. Key optimization techniques such as quantization, pruning, and knowledge distillation—will be presented, alongside practical applications in edge computing and mobile environments where real-time inference is essential.

Furthermore, the session will examine broader strategic considerations, including the trade-off between model performance and resource efficiency, the environmental implications of large-scale AI deployment, and the ethical imperatives associated with sustainable AI development. By integrating both technical and policy perspectives, the presentation will emphasize the critical role of lightweight AI in supporting the development of responsible, efficient, and scalable AI ecosystems.