Review on Carbon Aware Cloud Computing for AI

**Abstract**: Climate change crisis is an urgent global threat, and the rapidly increasing carbon footprint of AI workloads in cloud computing infrastructures necessitates immediate action. This survey paper investigates the field of carbon-aware cloud computing for AI, emphasizing the integration of carbon efficiency into AI cloud services. This paper mainly focuses on key challenges highlighted in the literature, including but not limited to the lack of transparency and disaggregated control over carbon emissions between cloud providers and users, the significant effort required for developers to implement carbon-aware optimizations, and the growing complexity of resource management due to renewable energy variability and hardware heterogeneity. By reviewing techniques like carbon budgeting, enhanced service-level emissions visibility, and configurable centralized resource management optimizations, this paper provides a detailed overview of current approaches and their effectiveness. This survey highlights the necessity for integrated carbon efficiency practices in cloud computing and encourages collaborative action among all stakeholders to promote sustainable AI development.  
  
**Keywords:** Carbon-Aware Computing, Sustainable AI, Carbon and Efficiency-Aware Compute Management, Cloud Computing for Carbon Reduction, Sustainability in Computing.

.