HYP_Link

EuroHPC to Establish AI Factories to Advance EU AI Capabilities

Mark Nossokoff and Bob Sorensen
September 2024

RECENT DEVELOPMENT

The EuroHPC Joint Undertaking (JU) recently initiated two calls seeking interest from hosting sites to house AI Factories. These calls stem from the July 2024 appended JU regulation to expand the JU's AI objectives. Funding includes up to €400M for new or upgraded AI EuroHPC supercomputers and up to €180M for the establishment and operation of the AI Factories as well as the development and deployment of advanced experimental AI-optimized supercomputing platforms. The EU funds will be matched with an equal amount of budget by participating JU countries. The AI Factories are intended to bring together key elements required for AI success (computing power, data, and talent) and make the resources available to a broad range of European users, including start-ups, industry, and researchers.

ANALYST COMMENTARY

A recent Hyperion Research global HPC-AI user site survey indicated that about 70% of the study participants currently employ AI to augment their traditional HPC modeling and simulation workloads. In the same survey, availability of in-house AI expertise and data-related items including quality, scope, and diversity were cited as the top barriers for participants in adopting AI even more broadly.

The EU has taken a measured and patient approach in its response to the increasing demand for AI infrastructure. Each of the JU's pre-exascale machines (LUMI, Leonardo, and MareNostrum 5) provide accelerated computation capabilities to support AI workloads, but AI wasn't the primary architectural focus for those machines. By expanding the JU's regulation to include AI objectives and elevating data and talent in the new calls, the EU is signaling its intent to be global leaders in provisioning AI-centric infrastructure and removing complexities for European users to adopt AI. The JU's approach also suggests recognition and balance of EU member states' individual strategic AI initiatives. For example, Germany's AI efforts include responsible development and deployment of AI and societal integration, while France's strategy highlights data policy as well as transparent and fair use of AI applications.

The JU's AI direction, with its emphasis on start-ups and industry, is differentiated from the more government-centric US efforts. For example, the US DOE has an aggressive focus on "AI for Science" with its AI Initiative at ORNL on Frontier (the current fastest supercomputer on the Top 500), Aurora (architected for AI for Science and anticipated to surpass Frontier on the Top500 when it is fully accepted), and the proposed FASST (Frontiers in AI for Science, Security and Technology) program.