HYPER-AI

Revolutionising big data applications with autonomous cloud-to-edge resources

Core Goals and Vision of HYPER-AI

Thanasis Moustakas, CERTH

HYPER-AI: Building a Hyper-Distributed AI Platform

Main Goal:

Create an innovative AI platform that enables hyper-distributed computing across Cloud, Edge, and IoT layers.

Core Idea:

- Shift computation closer to data sources using intelligent swarms of nodes.
- Leverage decentralized AI/ML to achieve efficiency, scalability, and interoperability.

Why It Matters:

- Handle the ever-growing data demands of modern industries.
- Support real-time, reliable, and efficient decision-making at scale.

Key Features of HYPER-AI

Decentralized Intelligence:

- Enable swarms of autonomous nodes to collaborate in real-time.
- Reduce latency and optimize resource usage by processing data locally.

Scalable, Adaptive Ecosystem:

- Seamlessly integrate diverse computational layers: Cloud, Edge, IoT.
- Dynamically adapt to changing workload, energy, and network conditions.

Robust Security & Privacy:

- Build trust with distributed ledger technologies.
- Protect sensitive data using advanced encryption and federated learning.

Strategic Benefits

Performance Boost:

- Accelerated processing by bringing computation closer to the data source.
- Enhanced energy efficiency and reduced communication overhead.

Industry Applications:

- Supports critical sectors: healthcare, manufacturing, agriculture, energy, and mobility.

European Leadership:

- Reinforce Europe's role in the global data economy through cutting-edge AI innovation.

Vision:

To lead the future of adaptive, decentralized computing, setting new standards in scalability and efficiency.

The Challenge in Data Processing

Modern Challenges in Data Processing Applications

Data abundance : Growing volume and complexity of data from IoT, AI, and edge devices.

Centralized Systems Limitations:

- High latency due to distant data centers.
- Bandwidth bottlenecks and high communication costs.
- Inefficient energy consumption and resource utilization.

Need for Innovation:

•A paradigm shift to decentralized, intelligent, and adaptive processing systems is critical to overcome these barriers.

Optimized Computing Swarms for Data Processing

Core Concept:

Distributed swarms of intelligent nodes collaboratively process data closer to its source, minimizing latency and optimizing resource usage.

Key Features:

- **Dynamic Resource Allocation:**

  Real-time workload distribution across nodes based on availability and efficiency.
- **Localized Intelligence:**

  AI/ML algorithms execute at the edge, ensuring faster insights.
- **Autonomous Functionality:**

  Self-healing and self-optimizing mechanisms for continuous operation.

Advantages:

- Reduced dependency on centralized data centers.
- Improved energy efficiency and cost savings.
- Enhanced reliability and real-time responsiveness.

Benefits to Data Processing Applications

Enhanced Performance

- Faster processing with reduced latency.
- Real-time analytics for critical applications like healthcare and manufacturing.

Scalability and Flexibility

- Seamless handling of large-scale applications across diverse environments.

Energy Efficiency

- Reduced energy consumption by limiting data movement.

Industry Use Cases:

- Precision Agriculture: Automated irrigation and monitoring systems.
- Smart Mobility: Real-time AI for autonomous vehicles.
- Green Energy: Efficient simulations for environmental stability.

Impact: Revolutionizing data processing for modern industries by combining speed, efficiency, and adaptability.

Vision for Cloud, Edge, and IoT Optimization

Unified Computational Ecosystem

Objective: Seamlessly integrate Cloud, Edge, and IoT layers to deliver scalable, efficient, and adaptive services.

Challenges Addressed:

- **Fragmentation:** Lack of interoperability between layers.
- **Latency and Bandwidth:** Inefficiencies in moving data to centralized cloud systems.
- **Resource Utilization:** Suboptimal use of computational and storage capacities.

Goal: Enable a cohesive continuum of computational, storage, and network resources.

Key Approaches for Improvement

Distributed Resource Management

- Real-time allocation of resources based on workload, energy, latency needs, etc.
- Autonomous optimization of computation and storage at the edge to minimize dependency on the cloud.

Seamless Data Flow

- Integration of IoT devices with intelligent edge nodes for localized decision-making.
- Reduced data transmission to the cloud, conserving bandwidth and improving response times.

Intelligent Networking

- Use of predictive models for traffic management and congestion reduction.
- Optimized pathways for low-latency communication across layers.

Project Impacts on Service Layers

Cloud Services:

- Offload non-critical tasks to edge and IoT layers, freeing up computational power.
- Efficient resource pooling for large-scale data processing.

Edge Services:

- Real-time processing and analytics near the data source.
- Adaptive decision-making with reduced latency and energy consumption.

IoT Services:

- Enhanced device coordination through edge intelligence.
- Efficient use of limited device resources for prolonged operation.

Outcome: A robust, scalable, and efficient ecosystem enabling faster, smarter, and greener applications across industries like healthcare, manufacturing, and energy.

HYPER-AI

Revolutionising big data applications with autonomous cloud-to-edge resources

Thank you