# Open Discussion #2

LISBON
March 2025

# Theme

## AI, Water, and Energy

## InfoSys 2025 & InfoWare 2025

# Open Discussion

LISBON
March 2025

## Coordinators

> **Prof. Dr. Petre Dini**, IARIA, USA/EU

> **Dr. Martin Zinner**, Technische Universität Dresden, Germany

# Items on the Table

- **Size of languages models**
  - Large, small, tiny
  - Software, embedded (SoC)
- **Computation complexity**
  - LLMs of trillion of parameters
  - Highly repetitive computation
  - Redundant cycles in AI factory
- **Huge data**
  - Finding the best patterns (temporal aspects for dropping data)
  - Accepting adapted accuracy (aspects in stopping computation)

- **Needs for Computation and Data**
  - Computation power > **Energy**
  - Data Storage/Processing & Cooling of DataCenters > **Water + Energy**

**Examples of conditions**

Lenses/dioptre (what data?)
Syslog messages (all data?)

# Not only Metrics

**LargeLMs - SmallLMs - TinyLMs**

**Model Size** >1 billion parameters || 100 million - 1 billion parameters || <100 million parameters)

**Training Data** (size and context-based datasets; unbalanced input)

**Computation Needs** (hardware, energy; tiny: run on edge devices/phones)

**Cost** (high due to compute and storage requirements; moderate; minimal resources)

**Computation** - **Energy**

**Data and Datasets** – **Water + Energy**

1. **Cooling Systems**

**Water-Cooled Systems**: Water cooling vs. air cooling at managing the heat generated by servers.
**Types of Water Cooling**: Direct water cooling (water circulating close to or through components to absorb heat directly)
Indirect cooling uses water (to cool air or a secondary coolant).

2. **Water Consumption Metrics**

**Water Usage Effectiveness (WUE)**: This metric is used to measure the amount of water used by a data center relative to its computing power.

Liters per kilowatt-hour (L/kWh). A lower WUE indicates more efficient water use.

**Average Consumption**: F(cooling technology, climate), data centers can consume from thousands to millions of gallons of water per day.

A large data center might use up to 3 to 5 million gallons of water per day in a hot climate.

3. **Innovations to Reduce Water Use**

**Air-Side Economizers**: Outside air to cool buildings when the outside temperature is sufficiently low, significantly reducing the need for water-cooled systems

**Use of Non-Potable Water**: Some data centers are shifting to using non-potable water for cooling purposes to reduce their impact on freshwater resources.

**Advanced Cooling Technologies:** New technologies (immersion cooling, where servers are submerged in a non-conductive liquid that absorbs heat more efficiently than air or water, can drastically reduce water usage.

4. **AI and Increased Demand**

**Higher Power and Cooling Needs**: Substantial computational power generates more heat. This increases the demand for effective cooling, potentially leading to higher water use.

**Efficiency Improvements**: AI can also be used to optimize the operation of cooling systems, predicting the optimal times to use different cooling methods and reducing overall water and energy consumption.

# Energy / Computation & Cooling

## 1. Large Language Models (e.g., GPT-3, BERT Large

**Power Usage**: Large language models / training phase.

For example, training a model like GPT-3, with its 175 billion parameters, can consume millions of watt-hours of electricity.

**Infrastructure**: The models are trained on clusters of GPUs or specialized hardware like TPUs.

The operational phase (inference) consumes significantly less energy..

## 2. Medium Language Models

**Power Usage**: Medium-sized models require less energy for training and inference; still needing multiple GPUs.

**Balancing Cost and Performance**: Balance between performance and resource usage

Where the costs and environmental impacts of larger models are prohibitive.

## 3. Tiny Language Models

**Power Usage**: Tiny models are designed to be extremely efficient, capable of running on edge devices like smartphones and IoT devices.

Their power consumption is minimal, especially in inference mode.

**Efficiency and Application**: These models are ideal for applications requiring low latency and low power, such as mobile apps, wearable devices, and embedded systems.

# Energy / Optimization

- **Energy Efficiency Innovations**
  - **Quantization and Pruning**: Techniques like quantization, which reduces the precision of the numbers used in computations, and pruning, which removes unnecessary weights, can significantly decrease the power required both during training and inference.
  - **Hardware Optimization**: Using more efficient hardware architectures, such as those specifically designed for AI computations, can reduce energy consumption. For instance, TPUs and next-generation GPUs are optimized for deep learning tasks and can perform more computations per watt than general-purpose CPUs.
  - **Algorithmic Efficiency**: Advances in algorithms and model architectures continue to improve the energy efficiency of AI systems. Techniques such as transfer learning and distillation allow smaller models to leverage knowledge captured by larger models without the need for extensive retraining.

# AI - Environment

## 1. Sustainability and Environmental Impact

**Resource Conservation**: AI can contribute to more sustainable water and energy management by optimization.

The use of AI in optimizing water usage in agriculture and energy consumption in industries.

**Environmental Monitoring**: AI for tracking environmental changes and pollution levels in water bodies, aiding in the proactive management of water resources and energy production impacts.

**Renewable Energy Integration**: Potential of AI to enhance the integration and reliability of renewable energy sources, which can reduce reliance on non-renewable resources and decrease water usage in energy production.

## 2. Economic and Infrastructure Challenges

**Cost Implications**: Economic impacts of implementing AI solutions in water and energy sectors, including the cost of AI systems and the potential for cost savings through improved efficiency.

**Infrastructure Adaptation**: Challenges and necessities of adapting existing water and energy infrastructures to leverage AI technologies effectively.

**Access and Equity:** Implications of AI on equitable access to water and energy resources, especially in underdeveloped and developing regions.

## 3. Policy, Ethics, and Social Implications

**Regulatory Frameworks**: The need for robust regulatory frameworks to manage the deployment of AI in critical sectors like water and energy to ensure safety, efficiency, and fairness.

**Data Privacy and Security**: Concerns related to data privacy and security in AI applications, considering the sensitive nature of water and energy data.

**Long-Term Sustainability Goals**: AI can be aligned with long-term sustainability goals, including ethical considerations around prioritizing technologies that could potentially displace traditional jobs in these sectors.

# More…

# Open Discussion