Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

Energy Management for AI - Opportunities and Challenges

Engineering Intelligence – Cyber Valley at University of Stuttgart

November 29-30, 2023

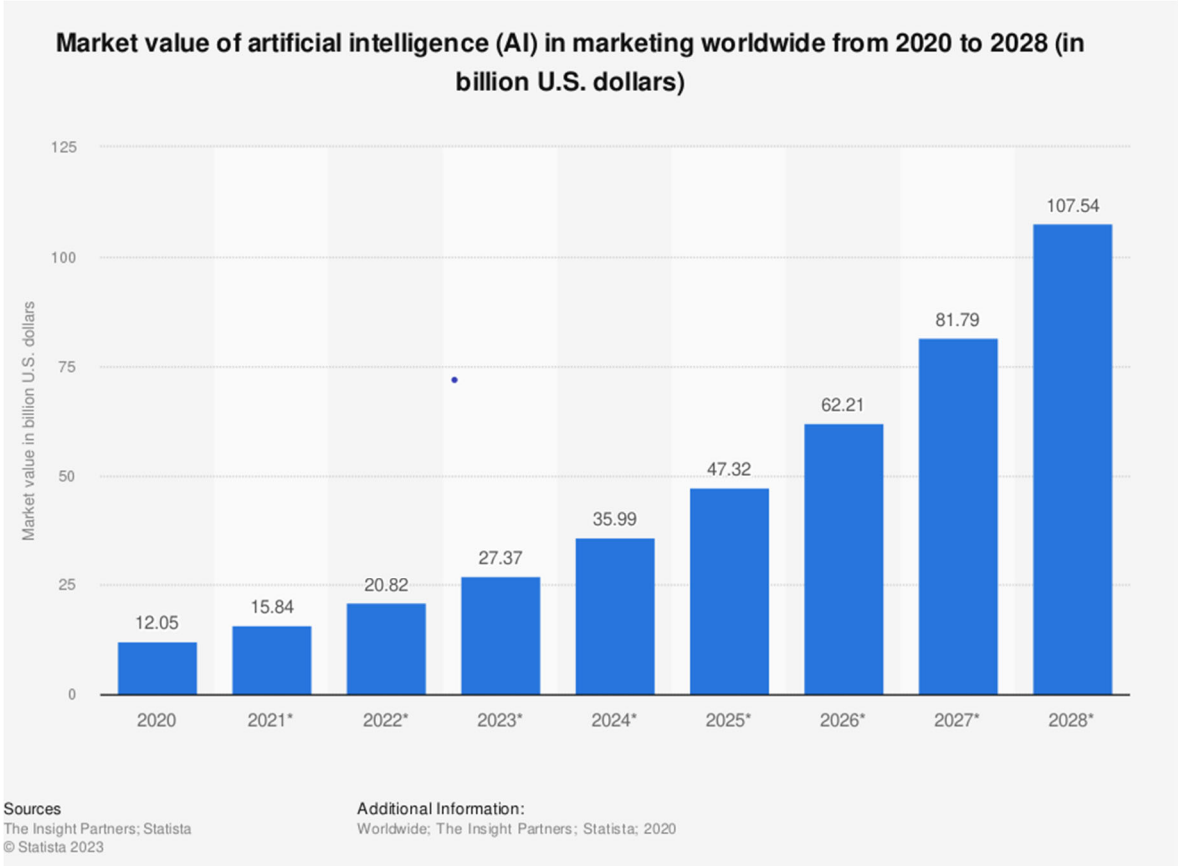
Prof. Dr. Peter Radgen

Source: HLRS, Stuttgart

Development of AI Sales – The nice market Outlook for Business

Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

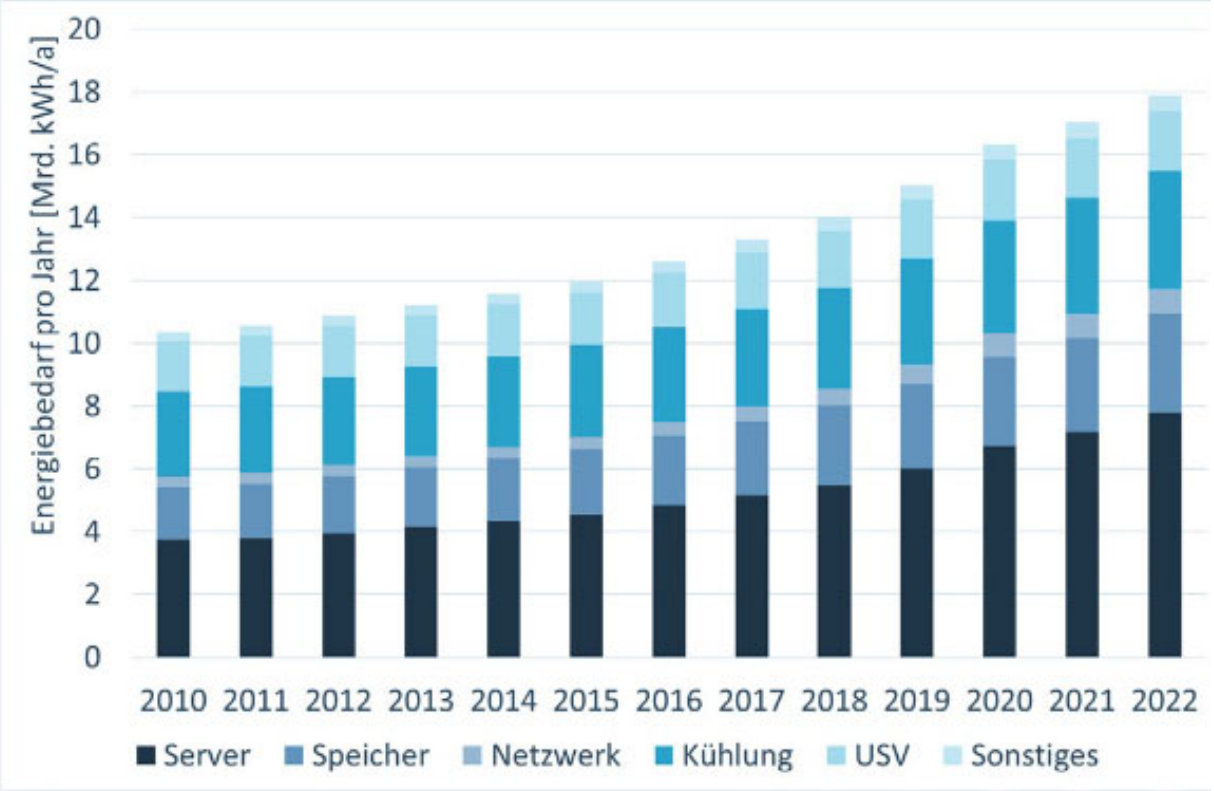
- Currently a strong upward trajectory
- Consistent acceleration in AI sales over recent years
- Forecasting an increase in AI sales until 2028, surging from 27,37 billion U.S. dollar to 107,54 billion U.S. dollar
- NVIDIA leads the AI servers market with estimated 95% share in the segment (Mehta et al., 2023)
- The company is expected to deliver 100,000 of its AI servers in 2023 (Leswing, 2023)



The Rise of Energy Consumption in German Data Centres

Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

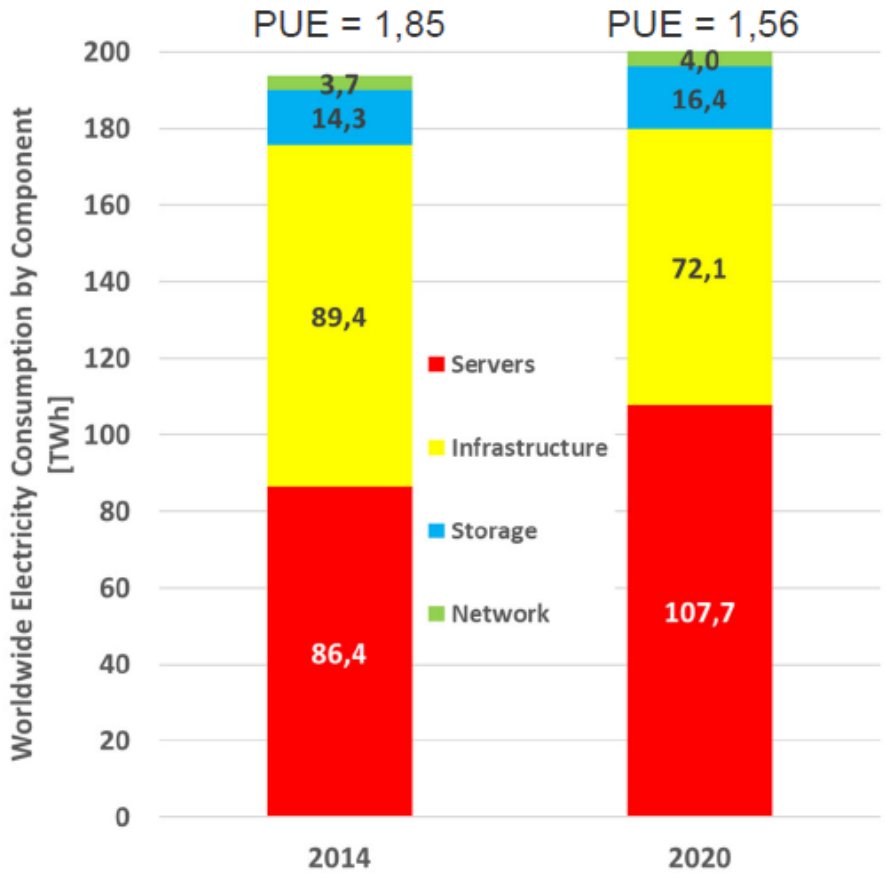
- Majority of consumption from servers and cooling
- **Power Consumption in Data Centers (2022):**
  - Total electricity consumed: 17.9 billion kWh
  - Share used by IT components: Nearly 12 billion kWh → PUE around 1.5
- **IT Component Energy Demand Growth:**
  - Since 2010, energy demand of IT components in data centers has more than doubled

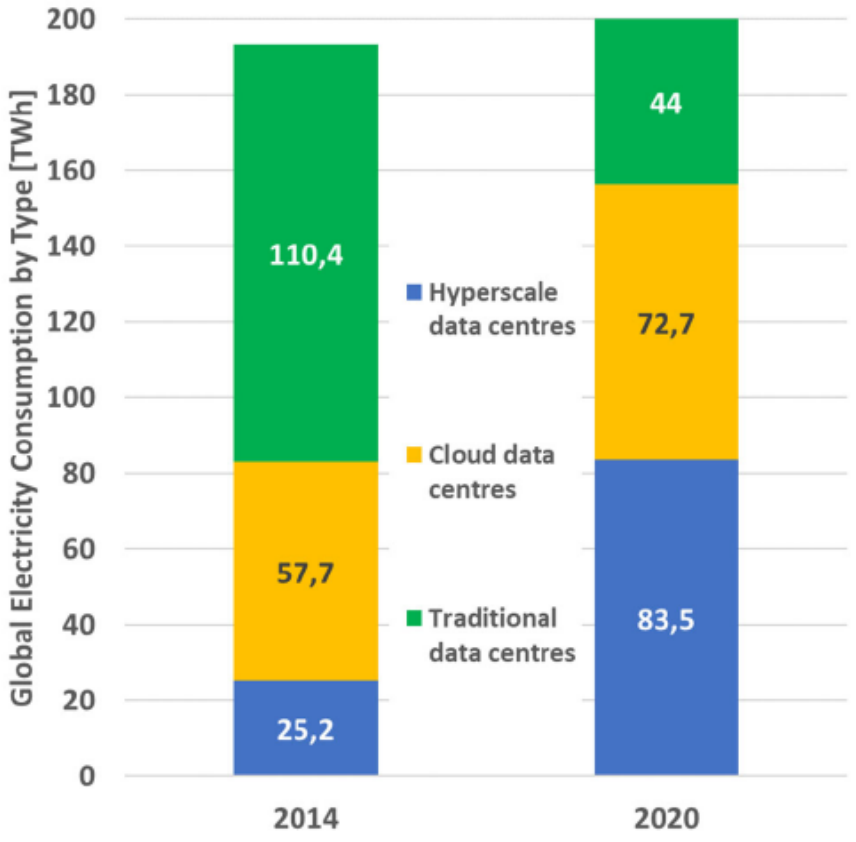


Hintemann, R. & Hinterholzer, S. (2023). Rechenzentren 2022.
Steigender Energie- und Ressourcenbedarf der Rechenzentrumsbranche. Berlin: Borderstep Institut.

Energy Management for AI – Opportunities and Challenges

Distribution of Electricity Consumption by Component and Type





Data Source: IEA, Paris, 2021. www.iea.org/data-and-statistics/charts/global-data-centre-energy-demand-by-end-use-and-data-centre-type-2014-2020

Energy Management for AI – Opportunities and Challenges

Why Energy Management Matters in AI

Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

- Open AI Chat GPT 100 million users within 2 month
- Chat GPT search estimated 15 times more energy demand than a traditional google search (Vries, 2023)
- Microsoft and Alphabet introduced Bing Chat and Board right after that (Heaven, 2023)
- Training models considered the most energy intensive
- BigScience Large Open-Science Open-Access Multilingual (BLOOM) consumed 433 MWh of electricity during training (Luccioni et al., 2023)

Data Centers – The new Energy Intensive Industry

Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

- Energy Consumption for data applications has become to play an important role to play in decarbonisation
  - In regards to its own energy consumption
  - In regards to its potential to reduce the energy consumption in other sectors
- Therefore Data Centers have come in the focus of Energy Policy
- The 2023 revised European Energy Efficiency Directive enforces for the first time efficiency reporting obligations
- The Energy Efficiency Law (EnEfG) in Germany requires the implementation of an energy management systems for DC > 300 kW non redundant power input and the use of the waste heat
- Data Centres are required in the future to report and publicly disclose energy and efficiency data

Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

NADIKI - Part of AI Lighthouses

Driving AI Innovation for Environment and Climate

- The project NADIKI is among the lighthouse projects that were awarded funding as part of the AI Lighthouse initiative.
- IER and IAAS from University of Stuttgart and SDIA are collaboratively working together on the project.

Source: www.uni-stuttgart.de/universitaet/aktuelles/meldungen/Foerderbescheid-fuer-KI-Leuchtturmprojekt-NADIKI/

- The project aims to capture and provide information on environmental impacts from digital infrastructures to enable the development of eco-friendly AI applications.

Transparency in Energy Consumption and Environmental Impact

Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

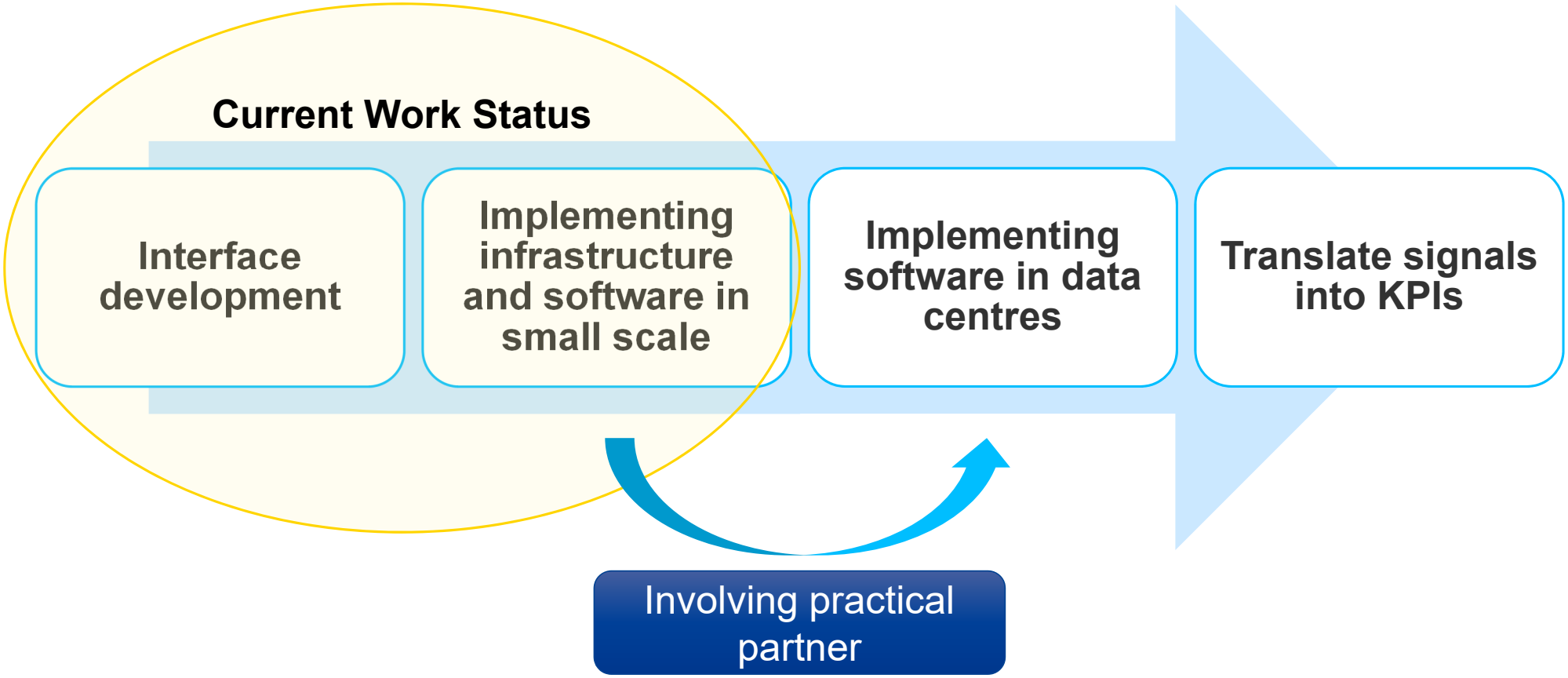
The NADIKI project aims to capture and provide information on environmental impacts from digital infrastructure to develop eco-friendly AI applications.

Developers should see and understand the impact of their use of compute resources to ensure that code is optimized, calculation are made when sufficient renewable energy is available and can try to reduce training efforts in AI application based on better conceptualization and structuring

Development of an open interface	Provision of signals	Translation of the signals
Aggregation of resource consumption data	From physical infrastructure and software level	Into metrics for: Power consumption Utilized server hardware cooling systems

Current Work Status

We are just at the beginning



How the Results for the NADIKI Project could be used

Universität Stuttgart
IER Institut für Energiewirtschaft
und Rationelle Energieanwendung

How to use the results

- Use NADIKI for real-time measurement and reporting of AIs environmental costs, including energy resources and emissions.
- The measured signals enable developers of AI applications to include energy and resource consumption in the development process.
- Adjust AI training in response to infrastructure signals, optimizing resource usage by aligning with renewable energy availability and adjusting training speed.
- Deploy AI applications strategically based on resource availability and cooling considerations, ensuring optimal efficiency and minimizing environmental impact.

How to get involved?

Implementation of the interface in data centre and getting real-time energy consumption information on you computing tasks in your compute environment.
If you are interested to **become a first mover and be engaged from the beginning**, please contact us!

Prof. Dr. Peter Radgen

Institute für Energy Economics and Rational Energy Use

peter.radgen@ier.uni-stuttgart.de

Nicola Schuckert

Institute für Energy Economics and Rational Energy Use

nicola.schuckert@ier.uni-stuttgart.de

Prof. Dr. Marco Aiello

Institute of Architecture of Application Systems (IAAS)

marco.aiello@iaas.uni-stuttgart.de

In cooperation with our project partner SDIA SUSTAINABLE DIGITAL INFRASTRUCTURE ALLIANCE