STARTUP INNOVATION CHALLENGE
powered by Energie AG

Challenge 2

Efficient Data Integration for Sustainable Decisions

What are the four main tasks of this challenge?

- **Data scraping:** Building scrapers to collect product and partner data.
- **Data harmonization:** Structuring data into a standardized format.
- **Data analysis:** Identifying gaps and inconsistencies.
- **Data enrichment:** Generating or filling in missing data, potentially using AI.

Is the focus on consumer goods or energy data?

The challenge primarily targets **consumer goods** such as fashion, cosmetics, and electronics, with a strong emphasis on Brand and Product data as well as sustainability metrics, including water consumption, carbon emissions, and certifications. Energy data is not the main focus.

What types of data are being targeted?

Sustainability metrics for consumer products (e.g., water usage, carbon emissions, certifications) for integration into a marketplace.

What level of development is expected? Does the solution need to be fully developed?

The team is looking for startups that can demonstrate their capabilities from day one. The solution should be a fully functional product that can be tailored to meet our requirements. At a minimum, it must be capable of automated web scraping from multiple partner websites. It is an advantage if the system can already read out sustainability reports and similar information.

What challenges might arise during data harmonization?

Partners often use different formats, making it difficult to ensure consistency across datasets.

Is this challenge only for energy-related products?

No, the focus is on consumer goods like fashion, cosmetics, and electronics, but with an emphasis on sustainability aspects.

What is the POC process like?

The budget of up to €50,000 is allocated in stages based on milestones. Continuous collaboration and feedback loops will be part of the Proof of Concept (POC) / MVP (Minimum Viable Product) phase

What is the key focus of this challenge?

To build tools that improve data accessibility, cleanliness, and structure, supporting sustainability efforts. The challenge has four components: scraping data, harmonizing it, analyzing it, and enriching it.

STARTUP INNOVATION CHALLENGE
powered by Energie AG

What industries does the data integration target?

The focus is on E-Commerce - so consumer goods, such as fashion, cosmetics, and electronics. Sustainability metrics like water consumption, carbon footprint, and certifications are central.

What if a startup's solution is only partially developed?

We will evaluate whether the partial solution provides a sufficient foundation for further development and aligns with our needs. It depends on what your solution offers and how well it addresses the core requirements of the challenge.

What types of gaps are analysed in the data?

Many products have poor or inconsistent data structures, or no structure at all. For example, a product might not have its color listed in a dedicated field but only in the product title. Other gaps could include missing certifications, manufacturing locations, or other key details that can be found on the website but not in the product data. The goal is not only to identify these gaps but also to analyze how the data is structured and develop a unified schema for consistency.

Are advanced tools like LLM-based scrapers required?

These are not mandatory but are encouraged for tasks like generating and validating missing data.