Before you start with AI, get your data in shape!

Michael Wiedau, Evonik; *Gregor Tolksdorf, Evonik*

Today, the landscape of artificial intelligence (AI) methods is highly diverse. It varies from classical AI methods like logical rule-based reasoning to highly sophisticated machine learning methods like support vector machines, clustering or artificial neuronal networks. No matter what application is in focus or which method is applied, it is necessary to provide data to those methods. In this presentation we give an overview about the major pain points that we face using AI technologies in process systems engineering as well as maintenance and operation in the (chemical) process industry.

Data Extraction and Provisioning

Sometimes, product and consulting marketing tells us that AI methods can work on any data and that there is no need to specifically structure the data. In this contribution we state that we disagree with these specific promises from marketing and show how AI methods would fail without a good foundation of data structuring in the beginning of an AI driven data management process. We argue that methods of AI, as mentioned above, lead to much better results when using well structured input data.

Process industry specific data handling

Especially in the process industry, companies often struggle with legacy IT systems that work with unstructured data. As an example, one of the most important documents in process systems engineering is the piping and instrumentation diagram (P&ID). The international DEXPI initiative has shown the problem of unstructured data behind P&IDs and the high potential of making these diagrams computer interpretable or even computer understandable. This systematic approach of structuring engineering information in the beginning of an Asset Lifecycle Data Management process can lead to high monetary savings in later steps like maintenance and data analytics for operations.

In this presentation the myth of "we don't have to care about data structuring, the AI will do that" is de-mystified.

Furthermore, this contribution demonstrates by practical examples how structured data management in the asset lifecycle can lead to better results during AI-based optimization of production processes.