LEVEL 0 SUMMARY

Name of student:

BURLON Carla

Name of your Level 1:

GONCALVES MELIM Maria Lolita

Source (e.g. scholars.google.com):

Paper available on Github

Paper title:

CHRISTOPH GRÖGER, Industry experiences on the data challenges of AI and the call for a data ecosystem for industrial enterprises, COMMUNICATIONS OF THE ACM, NOVEMBER 2021, VOL. 64, NO. 11

Keywords specific to the paper:

Machine learning, data mining, data management challenges, AI, insular AI, implementation, data ecosystem.

Summary of the main contributions:

This paper explores the current status of Artificial Intelligence (AI) in industries, specifically looking at machine learning and data mining. It points out that, despite a lot of investment in AI technologies, the actual benefits in real-world industries haven't matched the initial hype. The main problem discussed is the way AI is often implemented in isolated cases, like individual islands, rather than being integrated widely across industries.

The paper introduces the concept of "insular AI," where AI is implemented in isolated "islands" for specific use cases, leading to a heterogeneous and polyglot enterprise data landscape. This fragmentation poses various challenges to data management, data democratization, and data governance in real-world AI project.

The paper emphasizes the growing trend in traditional industries, such as manufacturing and automotive, to shift from just making physical goods to incorporating AI into processes and services as part of "Industry 4.0".

The paper suggests that complexities in managing data significantly hinder the broader adoption of AI in industries. To overcome these challenges, the authors propose the idea of a "data ecosystem" specifically designed for industrial use. This framework includes both technical and organizational elements, like data platforms and roles, aiming to make the transition frm isolated AI to a more integrated and widespread approach.

In fact, the paper describes the existence of structured and unstructured source data stored in isolated raw data stores, known as data lakes. These data lakes coexist with enterprise data warehouses, leading to various data redundancies and potential quality issues. The lack of a unified view of business objects, such as products and processes, across the enterprise complicates the development of integrated, cross-process AI use cases.

The authors provide a practical example to illustrate the challenges of insular AI. In this example, a project team of data scientists and data engineers works to predict the quality of a specific manufacturing process. The process involves extracting data from various source systems, developing customized connectors, cleansing and integrating the data, and employing machine learning tools. However, the resulting solution is described as an isolated AI island with specific data extracts, custom data models, and tailored data pipelines.

The paper argue that the same type of use case is often implemented independently across different factories, creating data redundancies, inefficiencies, and a lack of standardization. The heterogeneous data models and storage technologies used in individual factory data lakes contribute to the challenges of data management, democratization, and governance.

Then, the paper argues that the same type of use case is often implemented independently across different factories, creating data redundancies, inefficiencies, and a lack of standardization. The heterogeneous data models and storage technologies used in individual factory data lakes contribute to the challenges of data management, democratization, and governance.

Indeed, to address these challenges and move towards industrialized AI, the authors propose a systematic analysis of the underlying data challenges. THey suggest an overall solution that integrates both technical and organizational aspects to overcome the limitations of insular AI. The focus is on data quality, data management, data democratization, and data governance, with an emphasis on machine learning and data mining. The authors claim that this comprehensive approach is necessary for the successful implementation of AI in industrial enterprises.

In summary, the paper identifies three main challenges:

- managing data in a diverse landscape,

- making data accessible to everyone,

- establishing effective data governance.

> The proposed "data ecosystem" framework aims to address these challenge, highlighting the importance of data quality, management, accessibility, and governance for successful AI implementation in industries.