Big data analysis

L2: Asma DHOUIB and Coline DEKEYSER
L3: Schéhrazade BESOMBES

In the following document I will summarize the ideas and concepts discussed by my classmates based on the summaries done by the L2 members.

1. Business process modelling

Based on the summaries previously done by my L2 colleague, we can conclude these ideas and statement regarding the impact of AI in Business Process modelling:

AI has a growing influence in several sectors like recruitment, finance, marketing healthcare, education, fashion... For example, in recruitment, it has be showed that AI can optimize recruitment processes and roughly reduce discrimination. In the FinTech industry, AI brings a lot of innovation that is referred as "Smart FinTech." In marketing AI enables to generate precise customer segmentation and sales predictions, facilitating a lot the work of marketers. The AI offers a very precise customization and personalization in data processing, which is very beneficial for companies regardless of their sector and activity. The AI can generate high revenue growth, innovation, and customer satisfaction, making it very interesting to use for companies.

However, despite helping business in the management and analyse of data, some aspects are to be taken into account very seriously. Indeed, while using AI some considerations are to consider. Some of the challenges are data privacy, ethical considerations in qualitative data, and contextual adaptation. Some of these challenges can me contained by meticulous planning, execution, and ethical considerations to maximize the benefits of AI and data modeling in specific application domains. AI algorithms are also mentioned to enhance data analysis capabilities.

To conclude, we can say that AI integration can revolutionize processes and decision-making in plenty of sectors. Its capacities can facilitate a lot of the work of companies. However, challenges such as privacy concerns and the need for effective system integration are to be considered and corrected in the future.

2. Data modelling and source code generating and AI

The articles presented by the previous levels deal with the impact of AI in data modelling and source code generating.

The summaries mainly focused on AI-generated code, Large Language Models (LLMs), and their implications in software engineering. These discussed the integration of emerging technologies, such as neural architectures and AI-driven code generation tools. According to the articles studied, AI can help resolving challenges in translating natural language into source code, optimizing software development processes, and detecting code vulnerabilities.

The LLMs have a huge potential in scientific research, especially in software engineering tasks. LLMs have been proven to be useful in research processes and, to some extent, could enhance productivity.

Some security concerns, once again, associated with AI, are addressed. This time, it is regarding the code generators with the cases of data poisoning attacks. Methods to undermine system integrity are discussed, and a data poisoning attack is proposed to evaluate the security of NL-to-code generators (Natural Language – to-code).

Regarding software development, the articles studied were mainly about code generation, data management, software evolution, and model-driven engineering. These highlighted the use of models and metamodels in different contexts and explore the integration of model-driven approaches in engineering and software development. They also emphasized the importance of adapting the methodologies according to the context in which it is used.

As for the AI-based code generation in programming education, AI use is shown as a tool with highly transformative potential. AI-based code generation in programming education is explored in one of the article, evaluating ChatGPT's performance and users' attitudes. It stresses the ethical, pedagogical, and technical challenges of it.

Finally, aspects of software engineering and artificial intelligence, with a focus on the IDM (Intelligence Driver Model) are discussed. The IDM is usually used in traffic simulation and control systems. A catalog of refactoring for model-to-model (M2M) transformation in IDM is proposed and limitations of metamodels are discussed. A “Data Ecosystem” to overcome the current data management challenges is proposed.

To conclude, these articles provide an insightful overview of various areas of software engineering and AI research and practice, covering topics from model maintenance to prediction vector-borne diseases, data management and comparison.