



Data Science Lab with University Support

Data is a key tool for policy makers to respond to changes in a society and the challenges everyday people are facing. That is why Luxembourg's National Institute of statistics and economic studies, STATEC has launched a partnership with the University of Luxembourg's Interdisciplinary Centre for Security, Reliability and Trust (SnT) to set-up an in-house Data Science Lab. This lab will focus on developing and implementing AI tools that support STATEC's mission in Luxembourg.

Artificial Intelligence (AI) and Machine Learning (ML) technology offers the opportunity to maximise the capabilities of all national statistics offices around the world. It enables statisticians to process and analyse the massive amounts of granular information now at our disposal - big data. This allows for official statistics to be produced more efficiently, which makes the roll-out of AI/ML for big data processing a priority for statistical institutes around the world. Luxembourg is part of that, and STATEC has made AI and ML integration an important element of their innovation agenda.

SnT will be supporting STATEC's teams as they go beyond their foundational AI set-up, expanding their capabilities through the Data Science Lab. The project consists of three elements:

1. **Scaling and sustaining STATEC's AI activities**

Setting-up an ML system vs. a traditional software project is best compared to the difference between raising a child and building a house. AI systems, like children, require data to learn and develop further, whereas traditional software is built according to well-established "rules" or "blueprints". Because of these fundamental differences, the same technologies and processes effectively supporting software creation and maintenance are not sufficient for developing ML systems. SnT will be supporting STATEC's IT teams with the investigation of so-named "MLOps" technologies, which will permit STATEC to scale the development and adoption of AI systems across the organization.

2. **Investigating innovative use cases**

AI is a tool and so can be used for an infinite number of use cases. Therefore, the context and needs of any organisation need to inform the direction of innovation. In the case of STATEC a potential example is the coding of text according to statistical classifications. This process is currently done manually but could be automatised using Natural Language Processing. Another use case would be using AI/ML methods for making imputations with statistical data.

3. **Developing and implementing a tool to automate statistical processing**

A normal drawback of many AI solutions is the 'black box' effect, where information is inputted and then an output response, or answer, is given without the opportunity to understand why that answer was given. This is particularly problematic when it comes to statistical analysis where transparency is key, so for the third element of the project SnT's team intends to 'break' that black box, making it possible for the statisticians to understand how the data processing was done.

"We will accompany them throughout their AI journey and help them solve the many challenges they will face now, and in the future," says Prof. Yves Le Traon, the principal investigator from SnT for the project. "Ultimately, the goal

is to help statisticians and IT teams to work in-step, bringing their specialised knowledge together in the most effective way possible. This should give STATEC the capacity to ensure quality maintenance and evolution of their AI solutions. There are so many ways AI technology can help STATEC, and we are excited to be exploring these together”

“Introducing robust AI tools into our data processing methodology will allow our statisticians to work more efficiently, promising to free up their expertise for statistical analysis,” says Dr Serge Allegrezza, Director General of STATEC. “The result of this will be stronger data for Luxembourgish policy makers to work from, and more efficient processes for the citizens donating their time to us for our data collection activities.”