

Leveraging machine learning for analysing text-based data and assessing human factors identified in occupational incident reporting

Maria Tiikkaja, Henriikka Kannisto, Akseli Nurmi, Vuokko Puro, Tarja Heikkilä

Human activity has an impact on most occupational incidents. According to systems approach, it should be viewed as part of a socio-technical system and in the context of the time, place, and conditions in which the activity takes place. In organisations, large text datasets on safety incidents are collected yet underutilized. Machine learning based methods, such as large language models (LLMs) enable the processing of large data efficiently and are a promising tool to develop safety management.

The aim was to identify information gaps concerning causal factors behind human activity in text descriptions of safety incident reports in the framework of human factors (HF).

Three kinds of HF analyses were conducted with textual safety data from four industrial organisations

- 1) Investigation reports of 67 occupational accidents were analysed by experts using theory-oriented classification of the causal factors identified on the reports by classifying factors according to the top levels of a HF Tool™ which are work characteristics, individual factors and actions, organisational factors, and group factors.
- 2) 1582 accident and near-miss reports from the safety management systems of the organisations were analysed using LLM which was fine-tuned with expert-annotated-data to classify safety texts according to the four top levels of the HF Tool.
- 3) Text data on industrial occupational accident descriptions (331 00) obtained from the Finnish Workers' Compensation Center were analysed using the same LLM.

Of the HFs identified in the expert analysis of accident investigations almost half were related to work characteristics, ca 35% to individual factors and actions, ca 10% to organizational factors, and less than 10% to group factors. In the analysis of safety incidents using LLM nearly a fifth of the HFs identified were related to individual level, almost 80% to work level, with both organizational and group level factors forming 1%. Respectively, the results of the descriptions of the compensation center are that over a tenth belonged to individual level, with work level totaling up to nearly 90%, while organizational and group level factors were detected only a few times. The results give insight on how the top levels of the HF Tool™ are expressed in organisational data.