

AI For Traffic Safety: Combining Language Models and Topic Modelling to Understand Traffic Crashes

Mitta Hage (mitta.hage@gmail.com), Paula Granlund (paulagranlund@hotmail.com), Carlos M. Lima Azevedo (climaz@dtu.dk), & Mette Møller (mette@dtu.dk)
The Technical University of Denmark (DTU), Department of Technology, Management and Economics, Intelligent Transportation Systems Z

Background And Purpose

This study is conducted as part of a Master's thesis at the Technical University of Denmark (DTU), Department of Technology, Management and Economics, expected to be completed by June 26. Therefore, only part of the results are final at this stage, while the remaining results will be available before the conference.

Traffic crashes remain a major societal challenge, with an annual average of 21,397 crashes recorded by the police each year in Denmark alone [5]. Understanding the causes of these crashes is essential for improving road safety and supporting targeted, data-driven interventions, such as policy evaluation and infrastructure improvements. A central source of information in traffic safety analysis is police-written crash reports [4], which contain both structured information about the crash and involved parties, including the crash type ("uheldssituation"), and free-text narratives describing how the crash occurred. While the structured data provides descriptive information such as location, time, and involved parties, it does not fully capture contextual factors such as driver behaviour and interactions between road users. The narratives contain this richer context, but their unstructured form makes them difficult to analyse systematically at scale, limiting their use in traffic safety research [6]. Recent advances in artificial intelligence, particularly large language models (LLMs), enable structured extraction of information from such text while preserving contextual meaning [7]. At the same time, topic modelling can identify latent patterns and themes across large collections of narratives, providing interpretable insights into underlying structures [3].

The purpose of this study is to demonstrate the applicability of an AI framework for analysing Danish traffic crash reports. Specifically, we (i) classify crash narratives using LLM-based methods and (ii) extend the framework with topic modelling to identify underlying patterns, such as driver distraction and other contributing factors across crash types. By combining these approaches, the study moves beyond predefined crash categories and supports a more nuanced, data-driven understanding of traffic safety and targeted interventions.

Methods

The analysis is based on a dataset of Danish traffic crashes obtained from Vejdirektoratet (VD), containing both structured variables (e.g., time, location, and crash type) and unstructured police narratives providing a detailed description of each crash [5]. The dataset spans the period from 1985 to the end of 2025 and consists of 898,677 crash records. A preprocessing step was conducted to clean the texts and remove non-informative elements, ensuring high data quality and suitability for both language modelling and topic analysis. To analyse the police crash narratives, a two-stage framework is applied. First, a supervised classification approach based on large language models is used to assign predefined crash types. Second, a semi-supervised topic modelling approach will be employed to identify underlying patterns and themes within the narratives.

Training and Tuning Setup

For the classification task, the labelled crash narratives were divided into stratified training, validation, and test sets. The training set was used to fine-tune the BERT model, the validation set was used for model selection, and the final performance was reported on the held-out test set. For topic modelling, preparatory analysis included the definition of stop words and the tuning of UMAP and HDBSCAN hyperparameters on a subset of the embedded narratives. The selected settings were then used in the final BERTopic framework.

LLM-based Classification

LLMs were used to classify crash narratives into predefined crash types. The study follows the architecture proposed by Wang et. al. (2025) [7]. Specifically, a pre-trained BERT model was fine-tuned on the crash narratives to enable domain-specific classification. Fine-tuning allows the model to adapt from general language understanding to the specific terminology and context of traffic crashes [1]. BERT uses bidirectional encoding, meaning it considers both the words before and after each word to understand its meaning. This helps the model capture important details, such as behavioural patterns, situational dynamics, and other contextual factors that are not included in traditional structured data. As a result, the approach enables a more detailed and context-aware categorisation of crashes. This approach is used to validate the implementation in the Danish context and to compare the classification results with those reported in an existing study from the US.

Topic Modelling

To identify latent patterns and themes in the crash narratives, topic modelling was applied. Specifically, BERTopic was used, leveraging the same transformer-based embeddings as the BERT model [2]. This allows the topic modelling to build directly on the representations learned during classification, enabling a coherent analysis of the narratives. BERTopic combines embeddings of the police narratives with clustering and class-based TF-IDF (c-TF-IDF) to generate interpretable topics. The model groups semantically similar narratives and extracts representative keywords for each topic, enabling the identification of recurring crash patterns and contributing factors.

A seeded topic modelling approach is applied to help group the data into domain-relevant topics. This is achieved by first extracting topics using an unsupervised BERTopic model and subsequently incorporating these insights as prior knowledge in a semi-supervised BERTopic model. This approach improves the interpretability and relevance of the resulting topics. This second step extends the existing method by extracting relevant topics from the narratives to complement the structured data. The framework is illustrated in Figure 1.

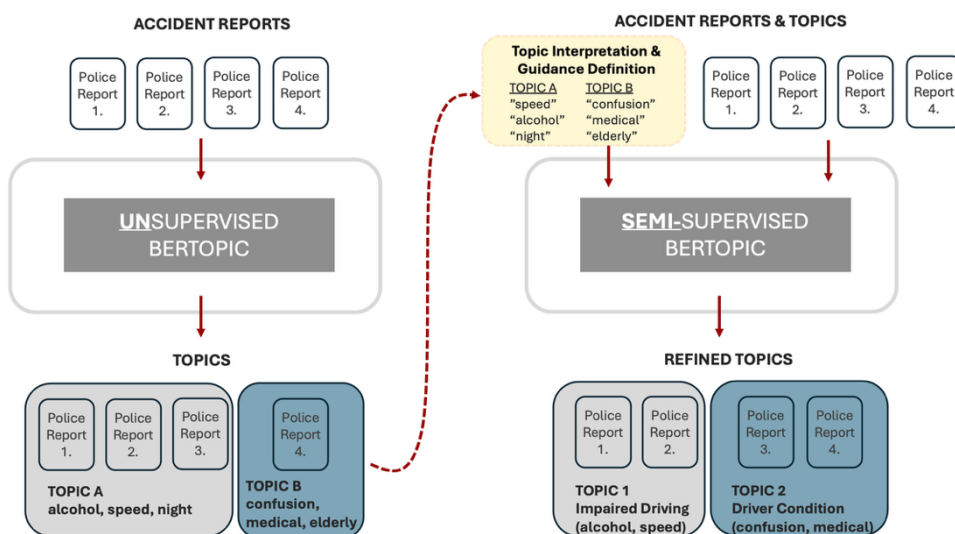


Figure 1. Illustration of seeded topic modelling using BERTopic. On the left, an unsupervised approach identifies topics from crash reports. On the right, seed words guide topic formation, changing how some topics are formed and how reports are assigned.

Early Results LLM Classification Performance

The initial results from the BERT-based classification demonstrate that LLMs can effectively process Danish police narratives and assign predefined crash types. The model achieved an accuracy of 86%, meaning that it correctly classified 86 out of every 100 crash reports. This indicates that the model is generally able to capture and interpret the contextual information contained in the narratives. The classification results shown per crash category can be seen in a confusion matrix in Figure 2.

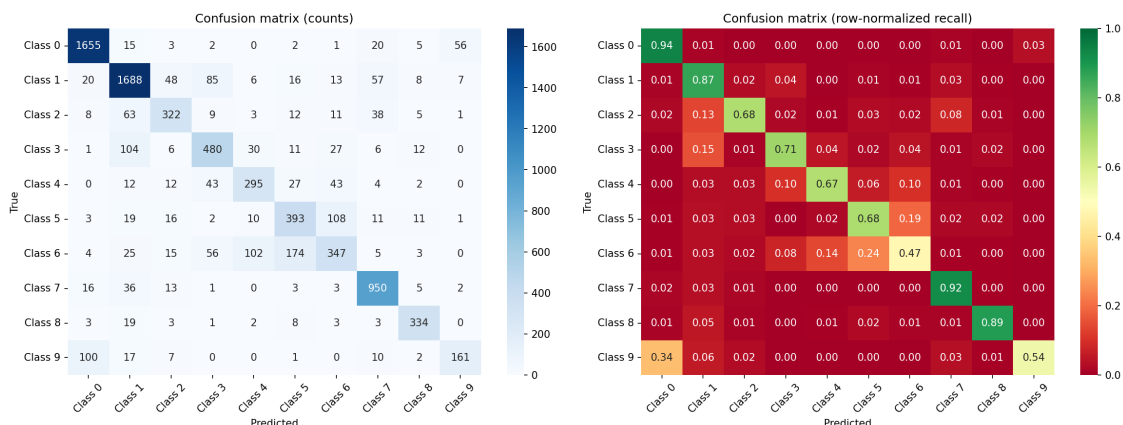


Figure 2 - Confusion matrices for the BERT classifier across all ten crash situation classes. The left panel shows raw prediction counts. The right panel shows row-normalised recall scores, values close to 1.0 on the diagonal indicate strong per-class recall.

Expected Results for Topic Modelling and Insights

Topic modelling is currently in progress and is expected to reveal distinct patterns within the predefined crash categories by identifying recurring themes in crash narratives. These themes are anticipated to reflect road users, behaviours, and environmental conditions associated with different types of crashes, providing insights into underlying causes. The identified topics are expected to capture patterns such as congestion-related incidents and interactions between specific road users. For example, a topic related to rear-end collisions may include keywords such as “bremseade”, “kø”, and “bagfra”, indicating congestion and sudden braking. By linking these topics to the crash categories predicted by the BERT model, the analysis aims to provide a more detailed understanding of how different crash types are associated with specific behaviours and situational factors.

Furthermore, the study aims to analyse how these patterns evolve over time. By incorporating temporal information, it becomes possible to investigate how changes in traffic regulations and policies influence crash characteristics. For instance, shifts in crash patterns may be observed following policy changes such as modifications to driving age regulations or requirements for medical checks among elderly drivers. Overall, the expected outcome is a more interpretable and data-driven understanding of traffic crashes, where structured classification is complemented by insights into underlying causes. This can support the development and evaluation of targeted traffic safety interventions.

References

- [1] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint. <https://doi.org/10.48550/arXiv.1810.04805>
- [2] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint. <https://arxiv.org/abs/2203.05794>
- [3] Janstrup, K. H., Kostic, B., Møller, M., Rodrigues, F., Borysov, S., & Pereira, F. C. (2023). Predicting injury-severity for cyclist crashes using natural language processing and neural network modelling. Safety Science, 164, 106153. <https://doi.org/10.1016/j.ssci.2023.106153>
- [4] Statistics Denmark (2021). Documentation of statistics: Road traffic crashes. <https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/road-traffic-crashes> (Accessed: 2026-03-26).
- [5] Vejdirektoratet (VD). Danish traffic crash dataset, 1985–2025. Restricted access provided under agreement with Vejdirektoratet.
- [6] Vejdirektoratet (2025). Personal communication via email. Received: March 15, 2026.
- [7] Wang, X., Kovaceva, J., Costa, M., Wang, S., Pereira, F., & Thomson, R. (2025). Domain-Adapted Pre trained Language Models for Implicit Information Extraction in Crash Narratives. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.5589453>