

Developing a national flood impact database using newspaper archives and AI

Flooding is one of the most damaging natural hazards affecting society. It can disrupt transport, damage homes and critical infrastructure, impact businesses, and place strain on emergency services. As climate change is expected to increase the frequency and severity of extreme weather, it becomes even more important to understand where floods have happened in the past — and what consequences they had.

While information about flood impacts is essential for improving flood risk assessments and planning effective climate adaptation measures, gathering this information is often very difficult. Evidence of flood damage is often scattered across many different systems – such as insurance records, municipal reports, infrastructure operators, and emergency response logs – and much of it is not publicly available. This makes it hard to build a reliable national overview of flood events and their impacts in Denmark.

Researchers have increasingly explored written sources such as digitized newspaper archives as a valuable information source. Newspapers often contain detailed local reporting on when and where floods occurred, what caused them, and what consequences they had for people, buildings, and infrastructure. Recent advances in artificial intelligence (AI), especially modern language models, now make it possible to analyze these large text archives far more efficiently and systematically than before.

This project will develop an end-to-end AI-based workflow that scans large historical collections of news articles and identifies those that describe real flood events, while filtering out irrelevant content. The system will then extract practical details such as location, timing, flood type, affected buildings and infrastructure, and wider societal disruption. To strengthen reliability, the reported flood events will be cross-checked against meteorological observations and other supporting data.

The project is a collaboration between the Danish Meteorological Institute and COWI, combining expertise in flooding and AI from both organizations. It will run from March 2026 to February 2028. The outcome will be a national, quality-checked flood impact dataset for Denmark that can support engineering analyses, risk mapping, planning, design, and climate adaptation. In addition, the project will generate transferable knowledge on how AI can turn complex text sources into robust decision-support data – including methods for validation, handling uncertainty, and responsible use.