

An Ethical AI Framework to Predict and Inform Risk Classification (Zoning) for Mental Health Patients

Contact Person: Prof Thierry Chausalet, chausst@westminster.ac.uk;
and Dr Sanjib Pandey (visiting researcher), pandey1@westminster.ac.uk

In mental health care, zoning is a rating mechanism, which uses a three-tier colour coding to indicate perceived risk from the patients' conditions and to ensure they receive appropriate levels of support. It is a widely adopted manual system used across mental health settings (e.g., within UK's NHS), however it is time consuming and costly. This research proposes to automate classification, by developing a hybrid approach, which combines Temporal Abstraction to capture the temporal relationship between symptoms and patients' behaviours, Natural Language Processing to quantify statistical information from patient notes, and Supervised Machine Learning Models to make a final prediction of zoning classification for mental health patients.

In addition, the resulting models should be placed within an 'ethical AI research wrapper', by examining the available AI ethics techniques (e.g., explanatory model analysis) from existing literature, and applying a sample of them to allow users to determine why the models produced any given prediction for specific patients of interest, if such models were to be deployed. The ethical AI framework would then be evaluated on pre-development (e.g., business case, dataset building) and development (e.g., model training, building, and testing) phases where possible. This research framework could be used to produce a Mental Health Patient Zoning Decision Support Tool that meets the AI ethical standards currently being developed for healthcare AI solutions in the UK, and beyond.