

## **МОДЕЛІ ТА МЕТОДИ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ РЕДАГУВАННЯ ДОКУМЕНТІВ**

**Боцула О.О.**

**Національний технічний університет  
«Харківський політехнічний інститут», м. Харків**

Штучний інтелект (ШІ) постійно вдосконалює можливості автоматизації у різних сферах, включаючи обробку текстової інформації. Редагування документів, яке полягає у видаленні або приховуванні конфіденційної інформації, є однією з важливих задач, де застосовуються моделі та методи ШІ.

Моделі обробки природної мови (NLP) є основою для автоматизованого редагування. Однією з найбільш популярних архітектур є трансформери, такі як T5 (Text-to-Text Transfer Transformer) і RoBERTa (Robustly Optimized BERT Approach). Вони демонструють високу ефективність у задачах розпізнавання тексту, класифікації даних та вилучення сутностей. Такі моделі можуть автоматично визначати чутливу інформацію, як-от імена, номери рахунків або адреси.

Глибокі нейронні мережі (DNN), такі як згорткові нейронні мережі (CNN) і рекурентні нейронні мережі (RNN), є важливими компонентами для аналізу текстових даних у реальному часі. RNN, зокрема їх вдосконалені варіанти LSTM і GRU, забезпечують можливість обробки послідовностей тексту з урахуванням контексту. Вони застосовуються для вилучення семантичних особливостей і створення адаптивних методів редагування.

Методи редагування включають розпізнавання конфіденційної інформації за допомогою моделей розпізнавання сутностей (NER) та заміну таких даних на узагальнені або приховані значення. Застосування гібридних систем, які поєднують машинне навчання та правила, підвищує точність і гнучкість процесу.

Мультимодальні підходи забезпечують обробку документів, які включають не лише текст, а й графічні елементи. Моделі, такі як Donut (Document Understanding Transformer), спеціалізуються на роботі з документами у різних форматах, включаючи PDF та зображення. Вони поєднують функції обробки природної мови та комп'ютерного зору, що робить їх ефективними для аналізу складних документів.

Незважаючи на значний прогрес, автоматизація редагування документів має низку викликів. Точність ідентифікації залишається важливим завданням, оскільки помилки можуть призвести до витоку конфіденційної інформації або неправильного редагування. Масштабованість систем також викликає труднощі через обчислювальні витрати, необхідні для обробки великих обсягів даних. Крім того, етичні питання та дотримання правових норм, таких як ССРА чи GDPR, є важливими аспектами для розробки та використання таких систем.

Таким чином, моделі та методи ШІ відкривають нові горизонти для автоматизації редагування документів. Проте для їх успішного впровадження необхідно враховувати технологічні, етичні та правові аспекти, забезпечуючи захист даних та відповідність сучасним стандартам безпеки.