

МОДЕЛІ ТА МЕТОДИ ГЕНЕРАЦІЇ ДОГОВОРІВ ЗАСОБАМИ ГІБРИДНИХ НЕЙРОННИХ МЕРЕЖ

Волоховський В. Є.

Науковий керівник – к.т.н., доц. Назаров О. С.

Харківський національний університет радіоелектроніки, каф. ПІ,
м. Харків, Україна

e-mail: vitalii.volokhovskiy@nure.ua

Contracts play an important role in people's lives and normal functioning of companies. The correct drafting and composition of contracts in accordance with current legislation requires deep knowledge in the field of law. Manual drafting of agreements is a time-consuming process even for experienced lawyers. Modern machine learning techniques and approaches like large language models (LLM) automate the process of drafting legal contracts, increase reusability of the documents for similar cases and decrease burndown for legal professionals. Furthermore, such systems can save a significant amount of time allowing lawyers to focus on more complex tasks and provide better service to their clients.

Договори відіграють важливу роль у повсякденному житті людей та у нормальному функціонуванні компаній. Правильне складання, формулювання та оформлення договорів згідно з чинним законодавством та врахуванням інтересів зацікавлених сторін потребує вузькоспрямованих знань у галузі права. Складання угод вручну та внесення змін потребує багато часу навіть для досвідчених юристів і може призвести до помилок під час копіювання його частин з інших документів та до низького рівня повторного використання цього договору для інших випадків. Витрачаючи час на ці задачі, вони приділяють менше часу своїм клієнтам та розумінню їхніх потреб, що призводить до погіршення якості наданих послуг та неефективному використанню часу та навичок.

Визначимо проблеми, з якими найчастіше стикаються юристи та компанії, які працюють з договорами [1]:

– значна кількість часу та зусиль витрачається на складання та розуміння договорів;

– неюридичні відділи компанії, такі як відділ кадрів та продажів, залежать від команди юристів для укладання угод;

– різноманітність видів та формулювань призводить до неузгодженості між документами всередині компанії.

Автоматизація процесу генерації та аналізу договорів за допомогою сучасних методів машинного навчання може допомогти вирішити ці проблеми.

Аналіз існуючих компаній у галузі генерації договорів

Розглянемо деякі компанії у галузі автоматизації генерації договорів та підходи, які вони використовують:

– Jugo – використовує велику мовну модель GPT (Open AI) для складання угод, та надає можливості взаємодії з системою за допомогою AI-чату [2];

– Luminance – використовує власну модель Legal Inference Transformation Engine та надає можливості автоматичного ведення переговорів щодо змісту договору (Autopilot) [3];

– Icertis – використовує систему Icertis ExploreAI Service, який поєднує можливості великих мовних моделей Open AI, власні AI системи та Icertis Data Lake для аналізу та генерації документів;

– Oneflow – використовує GPT (Open AI) і надає велику кількість готових шаблонів договорів.

Серед основних функцій цих сервісів, які використовують методи машинного навчання, можна виділити наступні:

- автоматичне складання договорів різних типів;
- аналіз та резюмування;
- підтримка правил та обмежень під час складання документів.

Таким чином, ми бачимо, що компанії активно використовують сучасні підходи обробки природньої мови, надаючи нові можливості формування угод, значно спрощуючи та прискорюючи цей процес. Автоматизація цього процесу зменшує необхідну кількість юридичного персоналу компанії, дозволяючи неюридичним відділам складати угоди.

Аналіз джерел даних

Розглянемо набори даних, які можуть бути використані для навчання вузькоспрямованих моделей у галузі генерації договорів. Найбільшими наборами є Pile of Law та MultiLegalPile [4]. Pile of Law містить близько 256 ГБ правових та адміністративних даних. Для його формування було використано 35 різних джерел, включаючи аналіз юридичних документів, судові висновки, публікації державних установ, контракти, статuti, нормативні акти, журнали справ тощо. MultiLegalPile містить 689 ГБ юридичних документів 24 мовами з 17 різних юрисдикцій. Він включає у себе набір Pile of Law, а також декілька натренованих моделей на основі RoBERTa та Longformer.

Отже, ці набори даних можна використовувати для тренування як одномовних, так і багатомовних моделей, адаптуючи їх під законодавство різних країн.

Аналіз методів генерації природньої мови

Розглянемо підходи до вирішення проблеми генерації природньої мови (NLG). Серед найбільш використаних методів генерації тексту можна виділити наступні:

- рекурентні нейронні мережі (RNN);
- мережі з довгою короткочасною пам'яттю (LSTM);

- вентильні рекурентні вузли (GRU);
- варіаційні автокодувальники (VAE).

Незважаючи на їхнє широке застосування, вони мають певні недоліки:

– послідовне навчання та обмежені можливості паралелізації обчислень;

– обробка довгих тестових послідовностей призводить до втрати контексту та помилок.

Іншим підходом до генерації тексту є архітектура Трансформер [5], яка має структуру кодувальник-декодувальник. Основним принципом запропонованого підходу є механізм самоуваги, який визначає важливість частин вхідної послідовності токенів до інших слів у цій послідовності. Зазначений підхід, на відміну від наведених вище, дозволяє:

– виконувати обчислення паралельно, зменшуючи необхідний час тренування;

– обробляти послідовності тексту більшої довжини без втрати контексту.

Виходячи з отриманих метрик (benchmarks), можна побачити, що сучасна архітектура Трансформер показує кращі результати виконання певних задач обробки природньої мови, порівняно з попередніми методами, що обумовлює значний інтерес до цього типу архітектури та її значне впровадження у різних галузях.

Таким чином, в ході аналізу предметної галузі було проаналізовано компанії, які надають послуги генерації та аналізу договорів, розглянуто методи машинного навчання для генерації текстової інформації, а також набори даних, які містять угоди, законодавства різних країн та інші юридичні документи, які можуть бути використанні для тренування вузькоспрямованих моделей у зазначеній галузі.

Список використаних джерел:

1. Generative AI for legal contracts. Nasdaq. URL: <https://www.nasdaq.com/articles/generative-ai-for-legal-contracts> (дата звернення: 10.03.2024).

2. Mabey R. Unveiling our legal AI Assistant. Juro. URL: <https://juro.com/blog/legal-ai-assistant> (дата звернення: 10.03.2024).

3. Browne R. An AI just negotiated a contract for the first time ever and no human was involved. CNBC. URL: <https://www.cnbc.com/2023/11/07/ai-negotiates-legal-contract-without-humans-involved-for-first-time.html> (дата звернення: 10.03.2024).

4. Niklaus, J., Matoshi, V., Sturmer, M., Chalkidis, I., & Ho, D.E. MultiLegalPile: A 689GB Multilingual Legal Corpus. ArXiv. 2023. Abs/2306.02069.