Автономная некоммерческая организация высшего образования
«Университет Иннополис»

АННОТАЦИЯ
НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ
(МАГИСТЕРСКУЮ ДИССЕРТАЦИЮ)
ПО НАПРАВЛЕНИЮ ПОДГОТОВКИ
09.04.01 – «ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА»

НАПРАВЛЕННОСТЬ (ПРОФИЛЬ) ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ
«АНАЛИЗ ДАННЫХ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

Тема **Обработка естественного языка для аудита соответствия нормативным требованиям. Автоматизированная проверка соответствия соглашений об обработке данных на основе обработки естественного языка**

Выполнил **Оконича Озиома Ненубари**

подпись

Иннополис, Innopolis, 2024

Оглавление

ОГЛАВЛЕНИЕ **3**

Аннотация

В современном мире соблюдение нормативных актов, касающихся защиты данных, таких как GDPR, играет центральную роль для организаций. Цель данной работы - описать способы, с помощью которых НЛП помогает обеспечить соответствие GDPR без особых усилий путем автоматического сканирования на предмет соответствия, оценки политик конфиденциальности и повышения уровня прозрачности.

Работа не ограничивается изучением применения NLP для работы с политиками конфиденциальности и облегчения лучшего понимания совместного использования данных третьими лицами, но также предполагает проведение экспериментов для оценки разницы между несколькими моделями NLP. Они внедряют и выполняют модели, чтобы выделить ту, которая работает лучше всего, основываясь на эффективности и скорости, с которой она автоматизирует процесс проверки соответствия и анализа политики конфиденциальности.

Таким образом, в данной работе подчеркивается важность использования НЛП для того, чтобы помочь организациям преодолеть трудности, связанные с соблюдением GDPR, и создать дорожную карту для перехода к более клиентоориентированному режиму защиты данных. В этой связи, сравнивая проведенные эксперименты и показывая эффективность лучшей модели, она помогает повысить эффективность мер, принимаемых для обеспечения соответствия, и способствует защите прав личности в киберпространстве.

Глава 1

Введение

Учитывая тот факт, что в современной цифровой среде люди производят множество персональных данных и обмениваются ими, защита частной жизни становится все более актуальной. Ведущую роль в этом играет Общий регламент по защите данных (GDPR) - соответствующий правовой инструмент и свод законов, принятых ЕС для смягчения обработки персональных данных и защиты прав граждан на неприкосновенность частной жизни и защиту данных. В целом GDPR решает проблему защиты доверия, налагая строгие законодательные требования на организации, обрабатывающие персональные данные, и регулируя деятельность по обработке данных на основе принципов прозрачности, подотчетности и согласия.

1.1 Постановка проблемы

Вопросы, описанные в этом разделе, связаны с проблемами, с которыми сталкиваются организации при попытке соблюдения положений GDPR и расшифровки неоднозначных правовых текстов. В нем освещаются такие

типичные условия, как как понять, что на самом деле говорят правила, законы в разработке и как донести концепцию политики конфиденциальности до неюридической аудитории? Цель этой части диссертации - объяснить эти проблемы, чтобы к концу работы мы могли попытаться ответить на следующие вопросы.

Прежде всего, необходимо отметить, что работа с правовыми нормами, предусмотренными в рамках GDPR, является весьма проблематичной для многих организаций, работающих в условиях цифровой среды. Одна из них - сложность, возникающая в связи с необходимостью осмыслить язык, используемый в законодательстве, который может быть трудно понять без юридического образования. Именно в связи с этим GDPR включает в себя множество правил и положений [1], которые иногда трудно понять, не говоря уже о том, чтобы выполнить.

Однако GDPR не является чем-то застывшим во времени; скорее, когда он вводится в действие, он меняется со временем, постоянно вносятся поправки, уточнения и новые правила. Это придает процессу еще один уровень динамизма, поскольку организация должна быть информирована о соответствующих изменениях и корректировать свои действия в соответствии с ними [2].

Кроме того, внедрение принципов защиты данных и политики конфиденциальности пользователей также представляет собой большую проблему с точки зрения того, как эти положения доносятся до пользователей. В большинстве письменных документов, особенно в политике конфиденциальности, используются юридические термины, которые многим людям трудно понять, и это заставляет таких людей не знать о своих правах и о том, как будут использоваться имеющиеся у них данные.

Эти проблемы объясняют, почему гибкие, нестандартные, простые в использовании решения, способные создавать положения GDPR простым языком и без особых усилий, могут значительно облегчить понимание юридических текстов, изложенных в соответствии с GDPR, а также помочь повысить прозрачность практики обработки данных. Решение этих задач может помочь организациям защитить права людей на неприкосновенность частной жизни и защиту данных и должным образом выполнить требования законодательства.

1.2 Вопросы и цели исследования

В следующей части статьи мы представляем вопросы исследования, которыми мы руководствуемся, и цели, которые мы ставим перед собой. Эти вопросы помогут нам перейти к более глубокому анализу и рекомендациям в области соответствия GDPR и обработки естественного языка (NLP).

- Насколько эффективны модели НЛП, включая GPT-3, для автоматизации выявления проблем соответствия GDPR в политике конфиденциальности данных организации?
- Каковы ограничения текущих технологий НЛП в интерпретации и обеспечении соответствия GDPR, и как эти ограничения могут быть устранены?
- Какую роль могут сыграть инструменты на базе НЛП в поддержке специалистов по соблюдению нормативных требований и юристов в обеспечении соответствия GDPR?

Глава 2

Обзор литературы

Объединение возможностей NLP с деятельностью по соблюдению правовых норм - это революция в понимании и работе с нормативными требованиями и юридическими документами. Как правило, соблюдение требований по защите данных отнимало много времени и средств, что влекло за собой бумажное просеивание и анализ огромных юридических текстов [3]—[5]. Хотя информации об использовании NLP в работе с нормативно-правовым соответствием не так много, его применение принесло эффективность и масштабируемость в использовании машинного обучения и понимания естественного языка в организации для работы с нормативно-правовым соответствием [6]—[8].

Ключевые области, на которых сосредоточен данный обзор литературы, включают:

- **Проверка соответствия:** Существуют хорошо продуманные и продвинутые модели и фреймворки, которые описывают, как можно автоматизировать проверку соответствия требованиям GDPR. Эти модели действуют как юридические консультанты, которые используют страте-

гии NLP для анализа юридических текстов и определения соответствия организации требованиям правовых прецедентов [9]—[11].

- **Анализ политики конфиденциальности:** Корпорации, а также такие инструменты, как PrivaSeer, стали мощными NLP-решениями для анализа политик и значительно помогли в крупномасштабном сборе данных для извлечения, а также классификации [12]—[14]. Эти представления могут быть использованы для повышения или улучшения текущего состояния прозрачности, а также для улучшения способности пользователей делать правильный выбор в отношении их права на конфиденциальность и защиту своих данных.

- **Семантическая аннотация:** Предыдущие попытки индексирования и аннотирования юридических текстов с помощью семантики позволили добиться улучшения работы поиска, а также извлечения информации [15]—[17]. Добавляя метаданные и семантические теги в юридические тексты, ученые создали возможности для практической работы в будущем, чтобы улучшить доступность и понимание правовых норм.

2.1 Будущие направления исследований

. На основе этих исследовательских пробелов и ограничений, упомянутых в настоящей работе, для дальнейшего развития NLP в области соответствия GDPR можно наметить следующие стратегии для будущих исследований:

1. Эксперименты с генеративными моделями: Метод использования генеративных моделей должен быть расширен в будущих исследованиях за

счет современных генеративных моделей, таких как GPT-3 и T5. Эти модели обладают расширенными возможностями по обработке и генерации НЛ, которые особенно подходят для сложных приложений, таких как проверка соответствия, проверка юридических документов и подобных процессов [6], [7], [17].

2. Расширенные оценки: Очень важно проводить более комплексные исследования решений и приложений для обеспечения соответствия на основе НЛП на предмет того, насколько хорошо они работают, насколько они устойчивы и насколько легко их можно масштабировать. Система оценки должна содержать множество наборов данных, метрик и приложений, чтобы дать рекомендации по использованию подхода и информацию о его эффективности [10], [18], [19].

3. Междисциплинарные подходы: Еще одним возможным методом дальнейших исследований является фокусирование внимания как на юридических знаниях, включаемых в разработку решений по обеспечению соответствия, так и на методах НЛП для улучшения таких решений. Междисциплинарное сотрудничество со специалистами в области правоприменения, информационных технологий, комплаенса и НЛП может способствовать улучшению существующих решений путем создания инструментов, более осведомленных о специфических проблемах в сложной сфере комплаенса [20]—[22].

Глава 3

Методология

В этой главе объясняется, как методы данной диссертации были использованы для обнаружения автоматизированного соблюдения Общего регламента по защите данных (GDPR) с помощью методов обработки естественного языка (NLP). Тот факт, что юридические тексты специфичны, а правовые нормы часто меняются, делает актуальным систематический подход к разработке, обучению и оценке моделей NLP. В данном исследовании использовались два различных набора данных: OPP-115 и ACL Coling. Они использовались для обучения и проверки эффективности нескольких моделей НЛП, включая SBERT, BERT и GPT2, которые являются одними из самых мощных технологий обработки языков.

Первое в методологии - получение и предварительная обработка набора данных, на котором будут обучаться модели. Затем проводится тщательный процесс сопоставления категорий из набора данных OPP-115 с принципами GDPR 5. Это в дальнейшем поможет в обучении моделей для классификации по нескольким меткам. Кроме того, мы анализируем различные гранулярности. Будет проведен анализ как на уровне предложений, так и на уровне всей

политики, чтобы выяснить эффективность моделей в различных ситуациях.

Также в данной методологии будет дано исчерпывающее объяснение возможностей и ограничений выбранных моделей, используемых для выявления соответствия GDPR. В частности, будут подробно рассмотрены плюсы и минусы упомянутых моделей, поскольку каждая из них подчеркивает свои уникальные возможности и причину выбора. Поэтому я подчеркну их пригодность для решения поставленной задачи.

Суть методологии заключается в процессе обучения классификации по нескольким меткам с учетом принципов GDPR, подготовке отчетов о соответствии и метриках оценки, специально созданных для оценки производительности и точности моделей. Наконец, я объясняю методы тонкой настройки, использованные для повышения эффективности моделей.

3.1 Получение набора данных

1. OPP-115 Dataset.

2. ACL Coling Dataset.

3.2 Обработка данных

Набор данных OPP-115 потребовал обширной очистки текста, особенно потому, что для получения политики с эквивалентной меткой нам нужно было просмотреть аннотированную версию и соединить предложения вместе, чтобы получить всю политику конфиденциальности. Предварительная обработка включала в себя следующие шаги:

- Очистка текста.
- Токенизация.
- Лемматизация.

Набор данных ACL Coling требует несколько иного подхода к предварительной обработке. После того как текст извлечен, он проходит несколько дополнительных этапов очистки:

- Normalization.
- Обработка пробельных символов и новой строки.
- Токенизация.
- Лемматизация.
- Удаление пунктуации.

3.3 Mapping GDPR Principles

Категории набора данных OPP-115, перечисленные в 3.1, сопоставлены с принципами GDPR, как показано в [23]. Это сопоставление выглядит следующим образом:

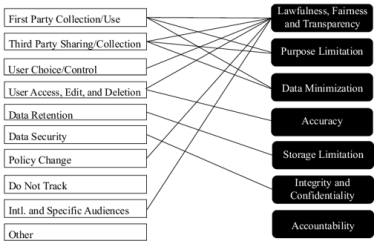

Fig. 3.1. Сопоставление категорий OPP-115 с принципами статьи 5 GDPR

3.4 Гранулярность анализа

В контексте применения НЛП для обеспечения соответствия GDPR гранулярность анализа имеет большое значение, поскольку мы применяем различные модели. Именно поэтому мы используем и сравниваем два уровня детализации: уровень предложений и уровень всей политики.

1. Анализ на уровне предложения.
2. Анализ уровня политики.

3.5 Выбор модели

Выбор подходящих моделей NLP - самый важный шаг для успешного решения любой задачи машинного обучения. Особенно в области анализа юридических текстов, где важны точность и глубинные тона интерпретации. Мы рассмотрим три различные модели, и теперь внимательно изучим каждую из них, чтобы понять, почему они были выбраны.

1. SBERT
2. BERT
3. GPT2

Глава 4

Реализация

В этой главе описываются детали реализации моделей машинного обучения, используемых в данной диссертации, включая SBERT, BERT и GPT2. Каждая модель была обучена как на уровне предложений, так и на уровне всей политики с использованием набора данных OPP-115 с метками. Затем модели были сохранены и протестированы на немаркированных политиках из набора данных ACL Coling, чтобы оценить их обобщенность и производительность. Кроме того, в ходе реализации были проведены эксперименты с вкраплениями GPT-3 с последующим обучением этих вкраплений на классификаторе.

Глава 5

Выводы

В данной диссертации основное внимание уделялось использованию НЛП для автоматизации проверки соответствия GDPR. Это начинание было продиктовано такими факторами, как растущая сложность юридических вопросов и необходимость эффективного применения законов о защите данных в организациях. В ходе комплексного анализа использовались самые современные модели НЛП, включая SBERT, BERT и GPT2, на двух уровнях гранулярности: с одной стороны, на уровне предложений, а с другой - на уровне всей политики.

5.1 Обзор выводов

Путем тщательного эксперимента и оценки было получено несколько ключевых результатов:

- **Эффективность модели:** Из оценки моделей можно сделать вывод, что, хотя все модели были достаточно эффективны в той или иной степени, SBERT лучше всего проявила себя на уровне политики по

метрикам, обеспечив высокую точность и оценку f1 по вопросам соответствия. В то время как при оценке практичности SBERT превзошла все остальные уровни на обеих гранулярностях. Модели BERT и GPT2 также получили приемлемые оценки; BERT показала лучшие результаты по метрикам, и они лучше анализировали взаимосвязи в тексте. В своих тестах на реальных политиках они, как правило, предоставляли более узкий анализ текста. Следующие оценки суммируют результаты, где максимумы на уровне предложений выделены синим цветом, а максимумы на уровне политик - оранжевым:

- **SBERT:**
  - **Уровень предложения:** Accuracy: 0.58, Precision: 0.56, Recall: 0.43, F1-score: 0.44
  - **Уровень политики:** Accuracy: 0.57, Precision: 0.78, Recall: 0.83, F1-score: 0.80
- **BERT:**
  - **Уровень предложения:** Accuracy: 0.63, Precision: 0.70, Recall: 0.50, F1-score: 0.55
  - **Уровень политики:** Accuracy: 0.26, Precision: 0.77, Recall: 0.75, F1-score: 0.73
- **GPT2:**
  - **Уровень предложения:** Точность: 0.62, Точность: 0.72, Recall: 0.48, F1-score: 0.54
  - **Уровень политики:** Точность: 0.48, Точность: 0.82, Recall: 0.75, F1-score: 0.78

- **Гранулярность анализа:** На предоставленном уровне анализа предложений были получены более точные результаты по вопросам соответствия, что позволило провести более тщательное изучение каждого пункта соглашения. Тем не менее, подход на уровне всей политики позволил получить более общие модели соответствия, которые можно проконтролировать на детальном уровне. Сравнивая результаты, полученные на двух разных уровнях детализации, можно отметить, что каждый из них особенно актуален для определенного набора мер регулирования соответствия.

- **GPT-3 Embeddings:** Благодаря встраиванию GPT-3 улучшилась контекстная осведомленность моделей, что привело к улучшению результатов прогнозирования. Тем не менее, вкрапления GPT-3 продемонстрировали оптимальное использование вычислительных ресурсов для повышения точности и побитового понимания текста, а также точности и запоминания, достигнутых при анализе довольно тонких рисков несоответствия.

5.2 Ответы на вопросы исследования

В данной диссертации ставится задача ответить на несколько ключевых вопросов, и полученные результаты дают некоторые ответы и понимание выбранных вопросов исследования:

- **Насколько эффективны модели НЛП, включая GPT-3, для автоматизации выявления проблем соответствия GDPR в политике конфиденциальности данных организации?** Результаты исследова-

ния показывают, что модели НЛП очень эффективны и точны, при этом основное внимание уделяется модели SI, включающей SBERT и BERT, для понимания соответствия требованиям. И без того приятное преобразование PDF в текст также улучшено с помощью вкраплений GPT-3, которые обеспечивают глубокое семантическое понимание и осознание контекста.

- **Каковы ограничения текущих технологий НЛП в интерпретации и обеспечении соответствия GDPR, и как эти ограничения можно устранить?** Основные недостатки заключаются в сложности вычислений и необходимости использования менее сложных с когнитивной точки зрения моделей, таких как GPT-3, а также в проблемах интерпретируемости. Преодолеть эти ограничения можно, увеличив скорость вычислений, сделав модели более доступными и придумав способы, с помощью которых модели могут быть легко поняты.

- **Какую роль могут сыграть инструменты на базе NLP в поддержке специалистов по соблюдению нормативных требований и юристов в обеспечении соответствия GDPR?** Также показано, что инструменты на базе NLP могут помочь специалистам по соблюдению нормативных требований, автоматизируя такие задачи, как правильное выявление проблем с соблюдением нормативных требований в организации, что снижает объем работы специалиста и повышает уровень надежности проверок соблюдения нормативных требований. Эти инструменты могут помочь в обеспечении соответствия, что, в свою очередь, освободит время специалистов по правовым вопросам для выполнения более сложных задач.

Список использованной литературы

[1] Ž. Spalević и K. Vićentijević, «GDPR and challenges of personal data protection,» en, *The European Journal of Applied Economics*, т. 19, № 1, с. 55—65, 2022, ISSN: 2406-2588, 2406-3215. DOI: 10.5937/EJAE19-36596.

[2] S. Sirur, J. R. Nurse и H. Webb, «Are We There Yet? Understanding the Challenges Faced in Complying with the General Data Protection Regulation (GDPR),» в *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*, сер. MPS '18, New York, NY, USA: Association for Computing Machinery, янв. 2018, с. 88—95, ISBN: 978-1-4503-5988-7. DOI: 10.1145/3267357.3267368. url: https://doi.org/10.1145/3267357.3267368.

[3] Z. S. Li, C. M. Werner, N. A. Ernst и D. Damian, «GDPR Compliance in the Context of Continuous Integration,» *ArXiv*, февр. 2020. url: https://www.semanticscholar.org/paper/71e16573d39360b98306b3bfa5482c10b4e73746.

[4] K. Mori, T. Nagai, Y. Takata и M. Kamizono, «Analysis of Privacy Compliance by Classifying Multiple Policies on the Web,» en, в *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, Los Alamitos, CA, USA: IEEE, июнь 2022, с. 1734—1741, ISBN: 978-1-66548-810-5. DOI: 10.1109/COMPSAC54236.2022.00276. url: https://ieeexplore.ieee.org/document/9842614/.

[5] A. Qamar, T. Javed и M. Beg, *Detecting Compliance of Privacy Policies with Data Protection Laws*. февр. 2021.

[6] S. Sousa и R. Kern, «How to keep text private? A systematic review of deep learning methods for privacy-preserving natural language processing,» en, *Artificial Intelligence Review*, т. 56, № 2, с. 1427—1492, февр. 2023, 12 citations (Semantic Scholar/DOI) [2023-12-08], ISSN: 1573-7462. DOI: 10.1007/s10462-022-10204-6.

[7] M. Srinath, S. Wilson и C. L. Giles, «Privacy at Scale: Introducing the PrivaSeer Corpus of Web Privacy Policies,» № arXiv:2004.11131, апр. 2020, 28 citations (Semantic Scholar/arXiv) [2023-12-08] arXiv:2004.11131 [cs]. DOI: 10.48550/arXiv.2004.11131. url: http://arxiv.org/abs/2004.11131.

[8] P. Silva, C. Gonçalves, C. Godinho, N. Antunes и M. Curado, «Using natural language processing to detect privacy violations in online contracts,» в *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, сер. SAC '20, 12 citations (Semantic Scholar/DOI) [2023-12-08], New York, NY, USA: Association for Computing Machinery, март 2020, с. 1305—1307, ISBN: 978-1-4503-6866-7. DOI: 10.1145/3341105.3375774. url: https://doi.org/10.1145/3341105.3375774.

[9] H. T. Alattas, F. M. Almassary, N. R. AlMahasheer и др., «Extract Compliance-Related Evidence Using Machine Learning,» в *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, 0 citations (Semantic Scholar/DOI) [2023-12-08], дек. 2022, с. 537—542. DOI: 10.1109/CICN56167.2022.10008324. url: https://ieeexplore.ieee.org/abstract/document/10008324.

[10] O. Amaral Cejas, S. Abualhaija и L. Briand, «ML-based Compliance Verification of Data Processing Agreements against GDPR,» English, сент. 2023. url: https://orbilu.uni.lu/handle/10993/55408.

[11] D. Torre, S. Abualhaija, M. Sabetzadeh и др., «An AI-assisted Approach for Checking the Completeness of Privacy Policies Against GDPR,» в *2020 IEEE 28th International Requirements Engineering Conference (RE)*, 33 citations (Semantic Scholar/DOI) [2023-12-08], авг. 2020, с. 136—146. DOI: 10.1109/RE48521.2020.00025. url: https://ieeexplore.ieee.org/abstract/document/9218152.

[12] S. Arora, H. Hosseini, C. Utz и др., «A Tale of Two Regulatory Regimes: Creation and Analysis of a Bilingual Privacy Policy Corpus,» в *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, Marseille, France: European Language Resources Association, июнь 2022, с. 5460—5472. url: https://aclanthology.org/2022.lrec-1.585.

[13] M. Bokaie Hosseini, P. K C, I. Reyes и S. Egelman, «Identifying and Classifying Third-party Entities in Natural Language Privacy Policies,» в *Proceedings of the Second Workshop on Privacy in NLP*, O. Feyisetan, S. Ghanavati, S. Malmasi и P. Thaine, ред., 7 citations (Semantic Scholar/DOI) [2023-12-08], Online: Association for Computational

Linguistics, нояб. 2020, с. 18—27. DOI: 10.18653/v1/2020.privatenlp-1.3. url: https://aclanthology.org/2020.privatenlp-1.3.

[14] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin и K. Aberer, «Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning,» en,

[15] Y. Ling, K. Wang, G. Bai, H. Wang и J. S. Dong, «Are they Toeing the Line? Diagnosing Privacy Compliance Violations among Browser Extensions,» в *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, сер. ASE '22, 7 citations (Semantic Scholar/DOI) [2023-12-08], New York, NY, USA: Association for Computing Machinery, янв. 2023, с. 1—12, ISBN: 978-1-4503-9475-8. DOI: 10.1145/3551349.3560436. url: https://dl.acm.org/doi/10.1145/3551349.3560436.

[16] D. Sánchez, A. Viejo и M. Batet, «Automatic Assessment of Privacy Policies under the GDPR,» en, *Applied Sciences*, т. 11, № 4, с. 1762, февр. 2021, 10 citations (Semantic Scholar/DOI) [2023-12-08], ISSN: 2076-3417. DOI: 10.3390/app11041762.

[17] P. Silva, C. Gonçalves, C. Godinho, N. Antunes и M. Curado, «Using NLP and Machine Learning to Detect Data Privacy Violations,» в *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 14 citations (Semantic Scholar/DOI) [2023-12-08], июль 2020, с. 972—977. DOI: 10.1109/INFOCOMWKSHPS50562.2020.9162683. url: https://ieeexplore.ieee.org/abstract/document/9162683.

[18] P. A. Bonatti, S. Kirrane, I. M. Petrova и L. Sauro, «Machine Understandable Policies and GDPR Compliance Checking,» en, *KI - Künstliche Intelligenz*, т. 34, № 3, с. 303—315, сент. 2020, 27 citations (Semantic Scholar/DOI) [2023-12-08], ISSN: 1610-1987. DOI: 10.1007/s13218-020-00677-4.

[19] J. M. Del Alamo, D. S. Guaman, B. García и A. Diez, «A systematic mapping study on automated analysis of privacy policies,» en, *Computing*, т. 104, № 9, с. 2053—2076, сент. 2022, 10 citations (Semantic Scholar/DOI) [2023-12-08], ISSN: 1436-5057. DOI: 10.1007/s00607-022-01076-3.

[20] A. Sleimi, N. Sannier, M. Sabetzadeh, L. Briand и J. Dann, «Automated Extraction of Semantic Legal Metadata using Natural Language Processing,» в *2018 IEEE 26th International Requirements Engineering Conference (RE)*, 48 citations (Semantic Scholar/DOI) [2023-12-08], авг. 2018, с. 124—135. DOI: 10.1109/RE.2018.00022. url: https://ieeexplore.ieee.org/document/8491129?denied=.

[21] T. A. Rahat, M. Long и Y. Tian, «Is Your Policy Compliant? A Deep Learning-based Empirical Study of Privacy Policies' Compliance with GDPR,» в *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, сер. WPES'22, 1 citations (Semantic Scholar/DOI) [2023-12-08] 1 citations (Crossref) [2023-12-04], New York, NY, USA: Association for Computing Machinery, нояб. 2022, с. 89—102, ISBN: 978-1-4503-9873-2. DOI: 10.1145/3559613.3563195. url: https://dl.acm.org/doi/10.1145/3559613.3563195.

[22] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin и K. Aberer, «Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning,» en, 2018, с. 531—548, ISBN: 978-1-939133-04-5. url: https://www.usenix.org/conference/usenixsecurity18/presentation/harkous.

[23] E. Poplavska, T. B. Norton, S. Wilson и N. M. Sadeh, «From Prescription to Description: Mapping the GDPR to a Privacy Policy Corpus Annotation Scheme,» в *International Conference on Legal Knowledge and Information Systems*, 2020. url: https://api.semanticscholar.org/CorpusID:229377855.