



Автономная некоммерческая организация высшего образования  
«Университет Иннополис»

ОТЗЫВ РУКОВОДИТЕЛЯ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ  
НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ  
(МАГИСТЕРСКУЮ ДИССЕРТАЦИЮ)

FINAL THESIS ASSESSMENT OF THESIS SUPERVISOR OF GRADUATION THESIS  
(MASTER GRADUATE THESIS)

ФИО обучающегося	Оконича Озиома Ненубари
Student's full name	Okonicha Ozioma Nenubari
Тема выпускной квалификационной работы	Обработка естественного языка для аудита соответствия нормативным требованиям. Автоматизированная проверка соответствия соглашений об обработке данных на основе обработки естественного языка
Thesis title	NLP for Regulatory Compliance Audit. NLP-based Automated Compliance Checking of Data Processing Agreements against General Data Protection Regulation

Уровень образования	высшее образование магистратура	Наименование направления подготовки	09.04.01 Информатика и вычислительная техника
Level of Education	Master's degree	Program track	09.04.01 Computer Science

Направленность (профиль) образовательной программы	<ul style="list-style-type: none"><li>Анализ данных и искусственный интеллект</li></ul>
Field of Study	<ul style="list-style-type: none"><li>Data Analysis and Artificial Intelligence</li></ul>

Отзыв на выпускную квалификационную работу<sup>1</sup>

<sup>1</sup> Научный руководитель указывает в произвольной форме:

Студентка провела образцовое практическое исследование по применению языковых моделей к проблеме проверки соответствия GDPR. Она эффективно определила область и выявила конкретные задачи, которые языковые модели могут практически решить. Учитывая нечеткий характер области проверки соответствия, это значительное достижение. В литературе был найден практический подход к отображению категорий правил соответствия, что заложило основу для исследования.

Одной из основных проблем было определение подходящего набора данных. Студентка успешно справилась с этой задачей, найдя несколько потенциальных наборов данных и выбрав два для экспериментов. Это свидетельствует о ее тщательности и умении подбирать данные, что является важным аспектом проекта.

Студентка оценил несколько моделей, включая семейства BERT, SBERT и GPT. Результаты оценки заслуживают внимания: самый высокий показатель F1 превысил 0,7. Несмотря на то, что это достойный результат, он все же может быть неоптимальным для практического применения, поскольку предполагает, что значительная часть информации о соответствии может быть неточно отражена в моделях. Несмотря на это, работа имеет значительный потенциал для использования в качестве помощника аудиторов-людей, обеспечивая ценную поддержку в процессе проверки соответствия.

В целом был проделан значительный объем работы, в результате которой были получены интересные и ценные результаты. Тщательный подход студента и добросовестное выполнение заслуживают похвалы. Рекомендуемая оценка за эту работу - А, что отражает высокое качество и потенциальное влияние исследования.

- 
- соответствует ли тема ВКР направлению подготовки, области, объектам, видам и задачам профессиональной деятельности;
  - является ли тема ВКР актуальной, соответствует ли современному состоянию и перспективам развития науки, техники и технологиям;
  - насколько освещена данная тема ВКР в монографиях, статьях, научных докладах и т.д.;
  - учитывает ли тема ВКР интересы и потребности индустрии;
  - четко ли поставлены цели/задачи работы, насколько соответствует их уровень теме ВКР;
  - какие методы, техники и методики использовались во время работы над темой ВКР;
  - основывается ли тема ВКР на практической работе обучающегося;
  - каков уровень сформированности компетенций показал обучающийся во время работы над темой ВКР;
  - есть ли публикации, доклады обучающегося по теме ВКР;
  - существуют ли перспективы дальнейших исследований по данной теме;
  - было ли самостоятельным выполнение работы обучающимся, ответственно ли и организовано относился обучающийся к работе, своевременно ли выполнялись все этапы индивидуального плана работы над темой;
  - является ли работа стилистически выдержанной, имеет ли смысловую законченность и оформлена ли в соответствии с требованиями;
  - рекомендуется или нет работа обучающегося к защите на заседании ГЭК;
  - какую оценку заслуживает работа.

## Justification of the grade<sup>2</sup>

The student has conducted an exemplary practical study on applying Language Models to the problem of GDPR compliance checking. She effectively defined the domain and identified specific tasks that language models can practically solve. Given the inherently vague nature of the compliance check domain, this represents a significant accomplishment. A practical approach for mapping compliance rule categories was identified in the literature, which laid the groundwork for the research.

One of the primary challenges was identifying an appropriate dataset. The student successfully navigated this by locating several potential datasets and selecting two for experimentation. This demonstrates her thoroughness and capability in handling data selection, a crucial aspect of the project.

The student evaluated several models, including BERT, SBERT, and the GPT families. The evaluation results are noteworthy, with the highest F1 score exceeding 0.7. While this is a commendable outcome, it may still be suboptimal for practical application, as it suggests that a significant portion of compliance information may not be accurately captured by the models. Despite this, the work holds substantial potential for use as an assistant to human auditors, providing valuable support in the compliance checking process.

Overall, a significant amount of work has been conducted, resulting in interesting and valuable findings. The student's thorough approach and diligent execution are commendable. The recommended grade for this work is A, reflecting the high quality and potential impact of the research.

---

<sup>2</sup> Thesis Supervisor freely states:

- Whether Thesis title corresponds to the program track, field, object, types and objectives of professional activity;
- Whether Thesis title is up-to-date and correlates with current state and prospects of science and technology development;
- The extent to which Thesis title is featured in monographs, articles, scientific reports;
- Whether Thesis title takes into consideration industry's concerns and needs;
- Whether goals/objectives of the work are clear, the extent to which their level corresponds to Thesis title;
- Which methods, techniques and approaches were used during the work on Thesis;
- Whether Thesis title is based on student's applied work;
- What level of competencies formed a student has demonstrated during the work on Thesis;
- Are there any papers or reports on Thesis title published;
- Are there prospects for any further research of the topic;
- Whether the student's working process was autonomous, did he/she do his/her work in conscientious and well-organized manner, did he/she do timely all the stages of individual work plan on the topic;
- Does Thesis stick to the style, does it have style completeness and complies with standard requirements;
- Whether student's Thesis is recommended for the Thesis Defense during State Attestation Committee hearing;
- What grade Thesis deserves.

Letter Grade / Оценка (A-D)	A
-----------------------------	---

Thesis Supervisor

Руководитель ВКР

подпись / signature



Andrey Sadovykh

Садовых Андрей  
Александрович