FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

UNIVERSITEIT GENT

Invitation

You are cordially invited to the public defense to obtain the academic degree of

DOCTOR OF BUSINESS ECONOMICS

by Abdel-Jaouad Aberkane

Automated GDPR Compliance in Requirements Engineering

Supervisors:
Prof. Dr. Geert Poels, Prof. Dr. Seppe vanden Broucke

Thursday, 11 September 2025 at 16h00
In the Faculty Board Room (Tweekerkenstraat 2, 9000 Ghent).
Please confirm your attendance by September 1, 2025, using the following link: https://forms.gle/7Kw6nCHgegTzvdhGA.

EXAMINATION BOARD

Prof. Dr. Ferdinand Vieider
Chair, Ghent University

Prof. Dr. Geert Poels
Supervisor, Ghent University

Prof. Dr. Seppe vanden Broucke
Supervisor, Ghent University

Prof. Dr. Frederik Gailly
Ghent University

Prof. Dr. Matthias Bogaert
Ghent University

Dr. Georgios Georgiadis
Head IT Unit, Secretariat-General, European Parliament

Dr. Fatma Başak Aydemir
Utrecht University

Dr. Ben Roelens
Open University

Abstract

The General Data Protection Regulation (GDPR) is essential for safeguarding individuals' privacy rights and ensuring that data is processed responsibly within the European Union. To comply with GDPR, organizations that handle personal data must implement appropriate technical and organizational measures to ensure data protection by design and by default. These measures should be integrated from the very beginning of the software development process, specifically during the requirements engineering (RE) phase. However, the legal complexities of GDPR can make it challenging to incorporate these measures during this stage. Therefore, this dissertation explores how automated techniques, particularly those based on natural language processing (NLP) and, more recently, large language models (LLMs), can facilitate GDPR compliance in RE.

To address this objective, the dissertation adopts a three-phase research design. The first phase systematically maps the current state of the literature at the intersection of GDPR, NLP, and RE, identifying promising directions for automation. The second phase moves toward practical validation by analyzing how data processing organizations disclose GDPR compliance in privacy policies and how these disclosures relate to organizational characteristics. In the final phase, the dissertation explores whether LLMs can support GDPR compliance evaluations, focusing on their alignment with expert assessments and professionals' attitudes toward adoption. These phases are covered in Chapter 2, Chapter 3, and Chapter 4, and will be briefly discussed below.

Chapter 2 presents a systematic mapping study that analyzes 448 primary studies at the intersection of GDPR, NLP, and RE. Most of these studies (420) focused on using NLP techniques across RE activities. Nine studies addressed the application of NLP for GDPR compliance, while twenty explored RE solutions aimed at achieving compliance with GDPR. Although only one study explicitly combined GDPR, NLP, and RE, the mapping results reveal opportunities for bridging these fields. In particular, they highlight the potential for using NLP to automate manual GDPR-related tasks within RE and the feasibility of applying NLP-based machine learning techniques for compliance support. The study demonstrates that while NLP techniques are widely used in RE, their application for GDPR compliance remains limited. The study lays a foundation for future academic research and industrial innovation by uncovering these research gaps and integration opportunities.

Chapter 3 shifts the focus to practice, investigating how GDPR compliance is communicated by organizations through their privacy policies. Using NLP-based machine learning models, the study analyzes privacy policies from 8,614 EU-based companies and identifies organizational factors correlating with GDPR disclosures. The models not only serve to validate the feasibility of automated GDPR compliance assessment in practice, but also enable large-scale analysis of compliance communication across sectors and regions. The findings indicate that factors such as company size and the location of the organization's headquarters are related to the extent to which GDPR core requirements are disclosed. Small and medium-sized enterprises are less likely to communicate their compliance efforts, suggesting that these types of organizations may benefit most from supportive tools such as automated compliance assistance.

Chapter 4 explores how LLMs can assist professionals in evaluating user stories for compliance with the GDPR. The chapter outlines a two-stage study: first, an experimental comparison of assessments made by experts and those generated by LLMs; second, a survey involving 51 professionals. The results indicate that three out of four evaluated LLMs perform comparably to human experts, highlighting their potential in compliance evaluation tasks. Additionally, the survey identifies trust and perceived usefulness as key factors influencing the adoption of LLMs. This suggests that these models could effectively support GDPR compliance evaluations of user stories, provided that professionals remain actively engaged in the process.

Chapter 5 concludes by synthesizing the main findings and reflecting on their implications for both academia and industry. It discusses the scientific and practical contributions of the research, acknowledges its limitations, and outlines directions for future research. These include domain-specific fine-tuning of LLMs on regulatory and GDPR-related corpora, as well as the application of qualitative methods to further deepen the understanding of professional attitudes toward adopting LLMs in compliance contexts.

Curriculum vitae

Abdel-Jaouad Aberkane (°1993, Amsterdam) obtained his Bachelor's degree in Computing Science from the University of Amsterdam in 2016, followed by a Master's degree in Business Informatics from Utrecht University in 2018. After graduating, he worked in the public sector as a software developer, focusing on the design and implementation of rule-based systems used in the administration of social benefits. In 2019, he began a PhD at Ghent University, where his research focuses on automating GDPR compliance in requirements engineering. His work leverages requirements engineering, natural language processing, and large language models to support data protection by design in software development. He has published in journals such as IEEE Access and the International Journal of Information Systems and Project Management, and presented at international conferences including the IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, and the IEEE International Conference on Security, Privacy, Anonymity in Computation and Communication and Storage. Currently, he is preparing his fourth chapter for journal submission. His doctoral research is supervised by Prof. Dr. Geert Poels and Prof. Dr. Seppe Vanden Broucke and is conducted within the Business Informatics Research Group of the Department of Business Informatics and Operations Management. In addition to his research, Abdel-Jaouad has contributed to teaching at Ghent University, serving as a teaching assistant for the Information Systems course.