MIC 2019
Management International Conference
Opatija, Croatia • 29 May–1 June 2019

Governance of Legislative Requirements for the Development of Natural Language Processing Tools

Inara Opmane
University of Latvia, Institute of Mathematics and Computer Science, Latvia
inara.opmane@lumii.lv

Juris Balodis
University of Latvia, Faculty of Law, Latvia
juris.balodis@gmail.com

Rihards Balodis
University of Latvia, Institute of Mathematics and Computer Science, Latvia
rihards.balodis@lumii.lv

Abstract. Information technology policy makers prioritize Artificial Intelligence as a new and important area of innovation. The focus on AI development is highlighted by Microsoft, as well as the US and European governments, who believe that the development of AI should remove the existing legislative barriers. New legislation and AI regulations are needed as there are not enough for a new researcher's code of ethics. The discussion is about basic issues: BIG Data and EU Data Regulation, GDPR, the new Copyright Law, as well as the international harmonization of AI regulation. The Institute of Mathematics and Computer Science of the University of Latvia is also conducting internationally recognized research in the field of language technologies. This study describes the regulatory compliance and risk management for Law Enforcement in our case. Regulatory compliance is the process of putting in place the measures necessary to comply with the regulations, laws, and guidelines that govern the operations of a business on a day-to-day basis. Information technologies policy makers prioritize Artificial Intelligence as a new and important area of innovation.

Keywords: information technologies, artificial intelligence, regulatory compliance, research restrictions

1 Introduction

A framework of the Study:the authors identify legal regulatory challenges and the need to solve them at a scientific institute in Latvia, where one of the activities is the development of Artificial Intelligence (AI) research, focusing on Natural Language Processing (NLP) tool development. The relationship of identified problems in a global view and the reflection of international legislation on activities in Latvia is analyzed. There is a significant difference between the AI's international and national legislative frameworks.

The AI solutions at their core are based on ICT, which is now the general-purpose technology (GPT) and the most comprehensive in human life [1]. This complicates the development of global AI regulation solutions, as the conditionality formulations must be comprehensive and legally binding on general human rights convention and ethical standards of scientists.

AI is international by its nature, the development of tools is labor-intensive and business management is based on a long-term strategy which means that the Institute needs to predict AI regulation in the future when the result of development can be used.

https://doi.org/10.26493/978-961-6832-68-7.2

Research Question and Study Results: What to do for the Institute in this Case? The authors of the article suggest that management principles be applied in the legislative field, using risk analysis of compliance with laws, changing business conditions or choosing the legally lowest risk solution.

The recommendations of this study are based and argued with a case study of the literature review, which extends a discussion section.

The road map for the discussion part of this paper contains:

- restrictions of academic freedom;
- regulations models;
- leveraging of Artificial Intelligence for society;
- AI and impact of regulation in the Institute;
- regulatory compliance and risk management for Law Enforcement.

Methodology and Purpose of the Article. The source of the legal challenges to be addressed is the practical work of the authors of the article at the Institute for many years and the method of identifying the problems is based on practical experience. It is combined with an extensive literature review and is finally being reduced to recommendations for improving practical work.

The Institute's NLP Research Group employs 22 staff. Employees have enjoyed an atmosphere of academic freedom for many years. It is problematic for the Institute's administration to transform academic freedom of research into a legitimate research activity. Employees need to get acquainted with the legal framework for their research work. We provide their activities with different levels of AI regulation: conventions, directives, policy, laws, institute regulations, ethical norms. The operating framework has become very complicated, thus the authors want to demonstrate this legislative scope of AI.

2 Restrictions on Academic Freedom

The notion of "Academic freedom" has a long historical background, whose character has been discussed and improved over time and it refers to the fundamental right of human beings. "Academic freedom, the freedom of teachers and students to teach, study, and pursue knowledge and research without unreasonable interference or restriction from the law, institutional regulations, or public pressure" [2]. Academic freedom includes scientific autonomy and is vital to best serve society. Academic freedom is essential to individuals and allows academic institutions, such as universities and research centers, to support research.

Different frameworks are developed for management, funding, organization, independence, and implementation of Academic freedom.

Let us compare the terms in Latvia legislation for scientific activities. The following terms are used in [3]:

- scientific activity - a creative activity that includes science, research, and innovations;
- innovation - the implementation in a product or service of new ideas, developments and technologies of a scientific, technical, social, or cultural field or other fields;
- research - a purposeful activity for the utilization of facts, theories and natural laws obtained with scientific methods in the creation or improvement of new products, processes and methods;
- science - the sphere of intellectual activity wherein knowledge regarding natural laws existing in nature and society is acquired and compiled using theoretical or experimental methods.

The other definition of scientific activities refers to technology readiness levels (TRL). TRL was originally conceived at NASA in 1974 and formally defined in 1989. TRL is used for R&D statistics and defined in OECD's Frascati Manual and used by science and innovation policymakers worldwide [4].

R&D intensity as a percentage of Gross Domestic Product (GDP) across OECD countries is relatively stable approximately at 2.3%, especial in 2016: Israel 4.25%, Korea 4.23%, EU 1.93%, USA 2.7%, Japan 3.1% and China 2.11% [5].

Academic freedom in Latvian Law is defined as "freely choosing the direction and methods of scientific research in accordance with his or her scientific interests, competence and the principles of humanism" [3].

Academic freedom is not absolute and not unlimited. Academic freedom is prevalent in scientific activity on the one hand, but on the other hand, there is a lot of restriction on scientific activity. Naturally, they refer also to AI. Apart from that, specific operating restrictions and prohibitions exist for AI.

2.1 Research Restrictions on Human Beings

There are historical reasons for the ethics of acceptance for research projects involving human participants. This has resulted in the adoption of international declarations and conventions that justify any research involving human participants, for example:

- the Nuremberg Code (1947) [6] sets out ten key points for responsible research with human participants;
- the Council of Europe has established the European Convention on Human Rights (1953) [7];
- the Helsinki Declaration (1964) [8], these ethical principles were further developed, with a more detailed look at issues such as informed consent.

Human studies define general principles, although there are national and regional variations in how they are expressed.

The US Belmont report is probably a brief summary of rights and obligations in relation to research and thus provides a framework for the management of research ethics in a number of places around the world [9]. The Belmont report includes three "ethical fundamentals" to assess ethical attitudes toward human participants:

- Persons: The personal dignity and autonomy of individuals must be recognized and persons with "reduced autonomy" (e.g. children and prisoners) should be particularly protected;
- Beneficence: Researchers are obliged to protect people from harm, maximizing expected benefits and reducing risk of harm;
- Justice: The benefits and burden of research should be shared fairly.

2.2 Restrictions on Biomedical Research

To implement the ethical issues of biomedicine research, the Council of Europe has developed The Convention on Human Rights and Biomedicine.[10]. The Convention states that human interests have a higher priority than science or society as a whole. It also outlines rules of medical research and detailed conditions, especially for those who cannot give their consent to participate in research.

More international guidelines in Biomedical research are outlined in [11, 12, 13, 14], but in Europe, for example, see [15, 16].

2.3 Research Restrictions on Animals

In 1966, Congress passed the Laboratory Animal Welfare Act as the first US federal law regulating animal research.

In order to prevent illegal trade of animals for research purposes, this Act established the transportation and sale of animals and licensing of animal traders, see Regulation of Animal Research [17]- The Animal Welfare Act [18], the European Communities Council Directive 86/609/EEC [19].

2.4 Research Restrictions on Nuclear Energy

Nuclear energy has historicaly been linked to national security. From 1942 to 1946, nuclear research was controlled and research was carried out by military services.

In 1946, the Atomic Energy Act control transferred to civilians, although the government retained a monopoly over nuclear energy. A 1954 amendment to this act allowed the private industry to carry out non-military applications uses of nuclear research [20].

3 Diversity of Research Regulation Models

There are several techniques and models for determining the necessary limitations in the laws and regulations for scientific activity. The regulated entity and its framework may be a sector, data, result, or a totally specific study entity.

In some cases, special teams of scientists should have greater access to research than scientists in general. Restrictions on research are determined according to the place of research, institutional affiliation, authorized supplies of goods, and the status of a scientist.

Legal and Regulatory Framework. The legal framework can be very diverse: conventions, international treaties, policy documents, national, state and sectoral agencies regulation rules or other institutional regulation.

Technological and Economic Blockade. Customs duties, export controls over strategic and military Goods, An embargo of all trade with a country or region, for example, CoCom, Comecon, Chincom, and Dot-Com, 1949-1994 [21].

State Security Restrictions and State Secrets. There are a number of federal, state, and local laws, regulations, and standards, as well as institutional policies,that pertain to emergency preparedness and laboratory safety. Several relevant laws, regulations, and standards are designated.
Research Ethics Regulation. Researcher Ethics is a daily guide for every scientist.

4 Leveraging of Artificial Intelligence for Society

AI is a field of science that explores a very wide range of manifestations of human behavior. AI is used in all areas of research, including those in respect to aforementioned limits.

AI is the theory and development of computer and hardware systems that are able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, learning, decision-making, and natural language processing.

Today AI ensures that the technical systems created can reach the boundaries of people's capabilities and make human-equivalent decisions. The role of AI in society is described in the respectable Microsoft book "The Future Computed" [22]. In this book, Microsoft describes the necessity of AI-specific regulation and laws.

The Trump administration has set a high priority for funding fundamental AI research and removing barriers to the deployment of AI technologies [23]. The EU Commission has adopted a communication that enhances the benefits of AI in Europe, which sets out the EU strategy on AI [24]. The European Commission is going to establish a group that will discuss the challenges associated with difficult ethical, legal and societal questions for the AI [25].

To regulate AI we need new laws, not just an upgrade of the code of ethics, as global tech regulations must be created to avoid an unhealthy concentration of power in technical equipment.

Obvious issues arise in relation to the intellectual property rights harmonization of authors, the rights of collective property managers, the interests of employers, the development of business in the AI field, the protection of citizens and the protection of persons. AI regulations new paradigm must contain answers to such topics.

4.1 Scientific Activity, as Well as AI, are International in Nature

As global challenges require global solutions, a common framework for the regulation of AI technologies could be established and AI policies developed around the world could be considered [26, 27].

4.2 Regulations for Big Data Sets and BIG DATA Technologies

The EU refers to related topics: big data, open data, open science, the building of a European data economy, data-based innovation development, intellectual property and digital commerce in the era of AI, big data: ownership and use in the Digital Age.

The European Union's databases are protected in accordance with European law. The current Directive 96/9/EC on the legal protection of databases was adopted in 1996 [28]. It has been evaluated for the second time in 2018 [29]. The Directive protects original databases. Obviously, minor changes to the Directive 96/9/EC will be made with the new Copyright Directive. It is important to note that if a database is created by an employee on behalf of an employer, Member States are entitled to provide that only the employer has the right to use all ownership (economic) rights of the database.

Aggregated databases from different existing data sources, such as statutory summaries or databases for scientific publications, may also be protected if it is essential to invest in data collection, verification, and submission.

This protection is called "sui generis" rights, i.e. special ownership of databases not related copyright protection. "Sui generis" right might, in fact, apply more broadly to machine-generated data and do not cover big data situations and single-source databases.

4.3 Government-Held Data (Public Sector Information)

The Directive 2013/37/EU [30] with additional agreement on 22 January 2019 on the re-use of public sector information provides a legal framework for government-held data.

4.4 Legal Protection of Computer Programs

Computer programs are protected by a Directive 2009/24/EC [31] as Copyright requirements. Only a Computer program is protected and ideas and principles are not protected by copyright under this Directive. As for databases if a program is developed by an employee on behalf of an employer, that only the employer has the right to use all ownership (economic) rights of the program.

A software license is a legal instrument for governing the use or redistribution of software.

Despite the fact that international model license agreements (such as CreativeCommons CC-BY, CC0-the US, adopted at first reading yet in new Copyright Directive) have been developed, licensing agreements may include very specific terms of use of the software in accordance with the requirements of the Copyright and also national legislation.

Software outside the scope of copyright protection is public domain software (PD). Examples of this are unauthorized public software repositories like GitHub without a specified license (see later [43]), but voluntarily handing software into the public domain (before reaching the copyright term) is problematic in some jurisdictions, there are also licenses granting PD-like rights, for instance, the CC0.

4.5 The New General Data Protection Regulation (GDPR) in EU

In many AI supplications, real personal data are subjects of study. Many say that the EU's new data privacy rules, the General Data Protection Regulation (GDPR)[32], will have a negative impact on the development and use of AI in Europe [33,34], putting EU firms at a competitive disadvantage compared with their competitors, for example, confusion about using Anonymized and Pseudonymized, undermine De-identification Data.

Gauthier Chassang [35] indicate the impact of the EU's overall data protection framework on scientific research: strengthening the general principles for any processing of personal data for scientific research.

4.6 EU Copyright Reforms

The results of AI research are intellectual property, and a broader definition of intellectual property is needed in the context of AI application development.

The results of AI research are intellectual property and we identify that a broader definition of intellectual property is needed in the context of developing AI applications. On 26 March 2019 Position of the European Parliament adopted at first reading with a view to the adoption of Directive (EU) 2019/… of the European Parliament and of the Council on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [36]. Article 1 of this Directive noted: "Except in the cases referred to in Article 24, this Directive shall leave intact and shall in no way affect existing rules laid down in the Directives currently in force in this area, in particular Directives 96/9/EC, 2000/31/EC, 2001/29/EC, 2006/115/EC, 2009/24/EC, 2012/28/EU, and 2014/26/EU.

The Directive shall provide "for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organizations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject-matter to which they have lawful access. Copies of works … be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results."

The rules of new copyright regulations followed by a wave of criticism from open-science developers, but the latest version of the directive resolution states that "providers of services such as open-source software development and sharing platforms, not-for-profit scientific or educational repositories as well as not-for-profit online encyclopedias should also be excluded from the definition of online content-sharing service provider. Finally, in order to ensure a high level of copyright protection, the liability exemption mechanism provided for in this Directive should not apply to service providers the main purpose of which is to engage in or to facilitate copyright piracy".

The main new Copyright Directive requirement is related to ‘online content-sharing service provider' and obligations in that case. "Providers of services, such as not-for-profit online encyclopedias, non-for-profit educational and scientific repositories, open-source software-developing and-sharing platforms, electronic communication service providers, online marketplaces, business-to-business cloud services and cloud services that allow users to upload content for their own use, are not ‘online content-sharing service providers' within the meaning of this Directive".

4.7 Collective Ownership (Economic) Right Management System

Directive 2014/26/EU [37] aims are to ensuring the operation of Collective Management Organizations (CMO) of economic right-holders, define the governance, transparency and financial management standards of CMOs. An important problem is the harmonization of CMOS and employer economic rights regulation, as, for example, is done in Ireland- Guidance Notes on Collective Rights Management Regulations [38, 39].

4.8 Changes in Patent Regulation

European patent legislation covers a wide range of regulation rules, including national patent laws, the Strasbourg Convention of 1963, the European Patent Convention of 1973 and a number of European Union directives and regulations in countries that are parties to the European Patent Convention. Often AI systems are built into different technical devices, so the patenting rules are especially emphasized. Digitalization has made changes to the patent system today [40]. Software patents include dual capabilities because the idea that is expressed can be protected by copyright law, but functional aspects of software-related inventions can be protected by patent rights. This establishes a basis for intensively discussions worldwide. Examples of patentable software are data compression, encryption algorithms, gaming systems, image processing.

Finally, the intellectual property of the database issue is about ownership of data and access to data. Do we need patent law for this? To ensure an appropriate ethical and legal framework, AI ethics guidelines will be developed [41], but a description of the legal framework for data ownership and access see in [42].

5 Short Remarks on the Regulations in Latvia

In Latvia, a number of regulatory acts are applicable to intellectual property management (The laws in force are specified, the predecessors of the law are not specified). All Latvian laws and regulations can be found in the official register of legislation www.likumi.lv.

Delay of harmonization of Latvia legislation according to EU and international requirements can be estimated for about two years: for example, in the new EU Copyright Directive (adopted proposal on 26.03.2019) appoint "Article 26, Application in time. This Directive shall apply in respect of all works and other subject matter that are protected by national law in the field of copyright on or after [24 months after the date of entry into force of this Directive]".

Latvia intellectual property rights legislation shown in Table 1.

Table 1: Latvia intellectual property rights legislation

Latvian Law	Adoption
Copyright Law	06.04.2000
Copyright Collective Management Act	18.05.2017.
Electronic Mass Media Law	12.07.2010.
Law on Scientific Activity	14.04.2005.
Patent Law	15.02.2007.
Law on Designs	28.10.2004.
Movie Law	17.06.2010.
Plant Varieties Protection Law	02.05.2002.
Law on Trade Marks and Indications of Geographical Origin	16.06.1999.
Law on Protection of Topographies of Semiconductor Products	12.03.1998.
Agreement on the Establishment of a Nordic-Baltic Regional Division of the Unified Patent Court	26.10.2017.
On the Single Patent Court Agreement	30.03.2017.
Law on Industrial Property Institutions and Procedures	18.06.2015.
About the Patent Agreement	26.11.2009.
About Singapore Trademark Agreement	17.07.2008.
The Geneva Act of the Hague Agreement Concerning the International Registration of Designs	10.03.2005.
European Patent Convention: Arrangement of 17.10.2000 on the application of Article 65 of the Convention on the Grant of European Patents and of 29.11.2000 on the Convention of 05.10.1973 on the grant of European patents	24.02.2005.
On the Agreement of 21 September 1960 on Mutual Secrecy for Inventions Related to Defense and for which Patent Applications have been Filed	02.12.2004.
About Locarno Agreement on the International Classification of Designs	25.11.2004.
The International Convention on the Protection of New Varieties of Plants of 02.12.1961	30.05.2002.
About the World Intellectual Property Organization (WIPO) Copyright Treaty	03.02.2000.
WIPO Agreement on Performance and Phonograms	03.02.2000.
The Madrid Agreement concerning the International Registration of Marks	02.09.1999.
About Trademark Agreement	02.09.1999.
Regarding the restoration of the force of the Law on Copyright and Related Rights	21.10.1998.
The Rome Convention on the Protection of Performers, Producers of Phonograms and Broadcasting Organizations	12.03.1998.
The Convention on the Protection of Producers of Phonograms against Unauthorized Reproduction of Their Phonograms	24.03.1997.
On the coming into force of the Republic of Latvia Law "On Plant Variety Protection"	06.04.1993.

Database.
Requirements about database usage are included in Copyright Law. Latvian legislation also provides "sui generis" for a regulation.

Implementation of the Public Sector Information Directive in Latvia.
Latvia has implemented in 2015 the Public Sector Information Directive as Amendments to Freedom of Information Act.

General Data Protection Regulation (GDPR).
Besides the Regulation (EU) 2016/679 in Latvia Personal Data Processing Law [47] can be applied for setting up a system for the protection of personal data also. This regulation uses terms specified in Article 4 of the Regulation (EU) 2016/679 of the European Parliament (General Data Protection Regulation).

Copyright Law.
In Latvia Copyright law is in effect with the last amendment 13.12.2018. It is obvious that provisions of the new EU Law have not yet been included in Latvian legislation. According to Copyright Law in Latvia, Authors have moral rights and economic rights.

Collective management of copyright.
Law on Collective Management of Copyright contains legal norms arising from Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market.

Patent regulation.
Patent Law with the last amendment on 01.01.2016 serves international provisions regarding patents.

In Latvia, a number of regulatory acts are applicable to intellectual property management (The laws in force are specified, the predecessors of the law are not specified). All Latvian laws and regulations can be found in the official register of legislation www.likumi.lv.

Delay of harmonization of Latvia legislation according to EU and international requirements can be estimated for about two years: for example, in the new EU Copyright Directive (adopted proposal on 26.03.2019) appoint "Article 26, Application in time. This Directive shall apply in respect of all works and other subject matter that are protected by national law in the field of copyright on or after [24 months after the date of entry into force of this Directive]".

6 International Databases and Repositories

International text and audio data repositories are widely used for NLP development, as well as repositories of NLP tools themselves, for example, GitHub [43], CLARIN [44]. Extended links can be found on [45, 46]. GitHub is built for developers. GitHub is a software development platform from open source to the enterprise.

International repositories contain sections on the use of data in accordance with the appropriate legislation and we can identify legal problems in this activity scope.

We identify a set of legal issues from GitHub including references to previously mentioned legal topics. GitHub's operation in many respects is similar to those of the Institute, so it is advisable to analyze the GitHub application and management of the legislative rules. The main fundamental difference in legislative governance, which we will have to note, is that Git is located in the US, but we in the EU and that US and EU intellectual property protection legislation are radically different. GitHub with partners in Europe will be governed by the laws of Ireland. GitHub sees that "We are proud to announce that we are compliant with the GDPR", but the situation in Latvia and Ireland in detail are different, for example, [38].

7 AI and Impact of Regulation in IMCS UL

Artificial Intelligence work in the institute [48] is carried out in applications of language technologies, including not only in technological solutions but also in solutions based on real language usage for large-volume texts.

Research in computational linguistics and machine learning is conducted at the Artificial Intelligence Laboratory in the following directions:

- speech recognition and text-to-speech synthesis for Latvian,
- syntactic and semantic parsing for multilingual information extraction and summarization,
- multilingual natural language generation,
- the controlled natural language for knowledge representation (multilingual authoring, verbalization, and querying of ontologies), machine translation (statistical, neural, rule-based),
- development of annotated Latvian language resources for NLP: lexicons, speech corpora, text corpora (incl. treebanks and embanks),
- computational lexicography and corpus linguistics,
- deep neural networks for classical algorithm learning,
- creation of Latvian language support resources.

IMCS UL research has a global character, including international project execution with many partners, for example, project on Scalable Understanding of Multilingual Media (SUMMA) together with The University of Edinburgh, UCL, University of Sheffield, BBC, Deutsche Welle, and others. In order to handle Big Data problems arising in the media monitoring business, as well as public availability of language text databases and their sale in Europe, USA.

The purpose of this study is not to analyze in detail the situation surrounding the regulation of AI but to show that global regulatory uncertainties have a real impact on a small national research institute.

Even though a new international law on the regulation of AI could soon appear, our national legislation will be delayed. What legal issues we face in the development of AI in a scientific institute.

In the case of the Institute, not all intellectual property regulations are applicable.

In accordance with Latvian legislation and the regulations of the Institute, personal copyrights belong to the employee (author) but the AI property (economic) rights to the Institute.

Database property rights belong to the institute by Latvian law:

1. For publishing an article, publishers require a copyright agreement, for example, Springer Copyright Act [49]. This contradicts the existing regulatory framework, such as the Institute's employment contracts. We are forced to change employment contracts and transfer property rights to the publisher. Likewise, the Institute's auditors often require the use of internal repositories with publications that have a public approach (and which is debatable or contrary to EU regulation). For instance, the scholarly social network ResearchGate has in recent month disabled public access to more than 1.7 million papers on its site, in compliance with take-down messages by publishers. This process of removing articles upon request, forcing all existing non-profit educational and research-data services, including university repositories.

There are a number of contradictions that need to be addressed and coordinated due to publication:

- resolve the Project sponsors, authors' personal and AI property rights,

- harmonization of the requirements of the international publisher with the legislation of Latvia,
- disclosure of commercial secrets about AI system solutions, database structure, and content in a publication.

2. Problems with EU-funded research projects which include commercial partners. Public accessibility and trade for commercial use.AI intellectual property development that has public funding should be open to public use. Like EU requirements in the case of Latvia, we can use GDPR and Copyright exemption of normal use of a Database: research exception, exemption from the obligation of citing the source, private use exception, extraction of insubstantial parts of a Database. Difficulties arise, for example, in EU-funded projects, where partners with state institute are commercial partners and, at the same time, funding institutions require to commercialize the scientific result in parallel with public use. EU requirements are often subject to specific requirements in the publication text that are not allowed by the publisher.
3. Harmonization of interests between the personal copyright of intellectual work of authors, the economic rights of the employer, the financier work finisher innovative product trade secret, research results publishers.
4. NLP tools are based on data (text, audio), where words are specifically labeled with grammatical and semantic tags. Such data bits must have a natural content, which means that the texts are selected according to natural language usage. This is the real use of language, so the database contains both fragments of literary works and real personal data. We use computer programs to extract data automatically from large numbers of texts. Scientists at public research organizations would be allowed to harvest facts and data from all sources they have legal access to read, therefore, we encountered these databases with GDPR and copyright predefined problems.
5. The Institute's activities, which make international regulation of AI important.
   The Institute cooperates with Creative Commons and uses GitHub - a development platform for language resources and tools. The Latvian text database corpora have been inserted into GitHub for noncommercial use - Non-Commercial-ShareAlike 4.0 International License. Creative Commons in halt Open access digital literature online, free of charge, and free of most copyright and licensing restrictions. Open Science is the practice of science in such a way that others can collaborate and contribute.

The Latvian text database corpora trade for Basis Technology Corporation, a Delaware corporation, is planned.

8 Regulatory Compliance and Risk Management for Law Enforcement

The Complex international and local legislative framework for Institute researchers requires special attention for administrative management. In keeping with the framework described above for the operation of NLP, it is not enough for the Institute to carry out a static assessment of the legislative situation and to make a decision, but for a legitimate NLP research, it is necessary to take the legislative review and governance continuously. That is the main governance lessons we can take from the GitHub. According to ISO 19600 (revised as ISO 37301 in 2020) [50] , Standard Compliance is the part of decision making. The Compliance based governance of legislative requirements for the development of natural language processing tools means performing several tasks.

Conventions, Directives, Laws, policies, company-defined rules we spread in three parts:

- regulatory requirements and contractual requirements arise from external sources;
- company-defined requirements;
- ethics-based requirements.

Then we will consider two possibilities: the Compliance based approach and the risk-based approach. We analyze only Compliance Risk (Legal Risk) as one type of possible risks: Strategic, Business Environment, Transaction, Investor Relations, Financial Risk, Operational Risk.

The Compliance Based Approach.
For these three documents groups, we determine the compliance - complies or does not comply with the regulation. In this task, we evaluate risk appetite and tolerance of our decision. While risk appetite is about risk probability, but risk tolerance is about the organization that can prevent it.

The Risk-Based Approach.
The risk-based approach starts weighing the risks and rewards and is the process of putting in place the measures necessary to comply with the regulations, laws, and guidelines. That govern the operations of a business on a day-to-day basis. Previously, we have shown that compliance with regulatory requirements in the subject of AI is complicated. Therefore we believe that research institution must designate a data protection officer, responsible for the practice of the data protection impact assessment, evaluate rules regarding the reuse of personal data for research purposes and assets the new provisions related to research participants' rights.

Let's list some simple suggestions for Law Enforcement according to artificial intelligence in our case:

1. Global constraints and restrictions by law (as originally referred to as restrictions on scientific activity) usually are assessed at the institutional level, establishing and opening a research direction or institute.
2. For daily risk assessment, we encourage you to use the decision risk assessment nXn matrix, collecting all the n necessary requirements for assessment.

With the nXn risk assortment matrix we check the compatibility of each requirement with all the other n-1 requirements. Thus, we find contradictory demands and try to solve the contradictions arising in the future.

If it is not possible to solve the problem functionally by changing the task solution, it is necessary to evaluate the potential normative changes when there is a contradiction.

Even further. Limiting geographically the task to action, for example, restricting the use of AI, for example, only in Europe or Latvia. Normative requirements and restrictions are usually inclusive, localizing to a narrower use.

We can create more complex risk assessment matrices. A risk matrix chart is a simple snapshot of the information found in risk assessment forms and is often part of the risk management process. These forms are more complex and involve identifying risks, gathering background data, calculating their likelihood and severity, and outlining risk prevention and management strategies.

Many organizations get an even clearer picture of risk by dividing the evaluation into zones, in our case, GDPR, data usage, copyright.

9 Conclusions

Academic freedom exists with research constraints. AI becomes more and more essential in technological development for society (in the human decision-making process), and we will face restrictions in defending the human rights of individuals.

The regulations harmonize the wishes of authors, the rights of collective property managers, the interests of employers, development of business in the field, the protection of citizens and the protection of individual persons.

AI research is global in nature. It is a complicated regulation at the moment, discussions about expected changes are taking place and national legislation is not harmonized with global regulation.

Even in the case of our institute, there is a need for extensive knowledge of the regulation of AI. The regulation is organized, and an employee is needed to evaluate all the risks of regulatory compliance and recommend further action.

References

Richard G. Lipsey, Kenneth I. Carlaw, and Clifford T. Bekar., ECONOMIC TRANSFORMATIONS: General Purpose Technologies and Long Term Economic Growth., OXFORD, 2005

Academic freedom. Encyclopedia Britannica, https://www.britannica.com/topic/academic-freedom

Law on Scientific Activity. https://likumi.lv/ta/en/en/id/107337

Guidelines for collecting and reporting data on research and experimental development. The 2015 edition. http://www.oecd.org/sti/inno/frascati-manual.htm

Main Science and Technology Indicators. OECD. http://www.oecd.org/sti/msti.htm

Nuremberg Code: www.hhs.gov/ohrp/archive/nurcode.html

Council of Europe: Convention of Human Rights and Fundamental Freedoms http://conventions.coe.int/Treaty/en/Treaties/html/005.htm

Declaration of Helsinki, International guidelines: www.wma.net/en/30publications/10policies/b3/

Belmont, The Belmont Report Ethical Principles and Guidelines for the Protection of Human Subjects of Research the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. DHEW Publication No. (OS) 78-0014 U.S. Government Printing Office, Washington, D.C. 20402 https://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf

Council of Europe, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine Research with human participants: Key principles. European Treaty Series-No. 164, Oviedo, 4.IV.1997, 10pages, https://www.epigeum.com/downloads/ri_accessible/uk/01_biomed/html/course_files/bi_2_30.html

Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines for Biomedical Research Involving Human Subjects: www.cioms.ch/publications/guidelines/guidelines_nov_2002_blurb.htm

UNESCO Universal Declaration on Bioethics and Human Rights: www.unesco.org/new/en/social-and-human-sciences/themes/bioethics/bioethics-and-human-rights/

The International Conference on Harmonisation's Guide to Good Clinical Practice (GCP): www.emea.europa.eu/pdfs/human/ich/013595en.pdf

World Health Organization Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants: http://whqlibdoc.who.int/publications/2011/9789241502948_eng.pdf

European Tissues and Cells Directive (2004/23/EC) www.hfea.gov.uk/2072.html

Council of Europe's Convention on Human Rights and Biomedicine: http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?NT=164&CM=7&DF=9/15/2008&CL=ENG

Regulation of Animal Research, https://www.ncbi.nlm.nih.gov/books/NBK24650/

The Animal Welfare Act https://awionline.org/content/animal-welfare-act

COUNCIL DIRECTIVE of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes, https://publications.europa.eu/en/publication-detail/-/publication/cc3a8ccb-5a30-4b6e-8da8-b13348caeb0c/language-en , 18 .12 .86 Official Journal of the European Communities No L 358/1

Act on the Peaceful Utilisation of Atomic Energy and the Protection against its Hazards (Atomic Energy Act), https://www.bmu.de/en/law/atomic-energy-act/

Christopher Leslie, South China University of Technology, Guangzhou, China:"CoCom, Comecon, Chincom, and Dot-Com: Technological Determinism in Economic Blockades, 1949-1994" 24th IFIP World Congress, IFIP WORKING GROUP 9.7 WORKSHOP " Histories of Computing in Eastern Europe" Poznan University of Technology, Poland 19-21 September 2018

The Future Computed. Artificial Intelligence and its role in society. By Microsoft Published by Microsoft Corporation Redmond, Washington. U.S.A. 2018

Brian Higgins, ARTIFICIAL INTELLIGENCE TECHNOLOGY AND THE LAW http://aitechnologylaw.com/2018/06/obama-trump-regulation-of-artificial-intelligence/, June 6, 2018

European Parliament resolution of 12 February 2019 A comprehensive European industrial policy on artificial intelligence and robotics. http://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html?redirect

EU to debate artificial intelligence regulation and legal issues https://diginomica.com/.../eu-debate-artificial-intelligence-regulation-le...

Oxford-style Debate: Should Artificial Intelligence be more regulated? Leon Strous, past President of IFIP. The 24 th IFIP World Computer Congress September 17-21, 2018

Olivia J. Erde´lyi, Judy Goldsmith Regulating Artificial Intelligence Proposal for a Global Solution, School of Economic Disciplines University of Siegen 57076 Siegen, Germany olivjoan@yahoo.com, Department of Computer Science University of Kentucky Lexington, KY 40506, USA goldsmit@cs.uky.edu]

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, https://eur-lex.europa.eu/search.html?qid=1557592457307&text=directive%20data%20base&scope=EURLEX&type=quick&lang=en

Evaluation of Directive 96/9/EC on the legal protection of databases. 26. Executive Summary, https://ec.europa.eu/digital-single-market/en/protection-databases

DIRECTIVE 2013/37/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013L0037

DIRECTIVE 2009/24/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCILof 23 April 2009on the legal protection of computer programs(Codified version), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0024&from=EN

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), https://eur-lex.europa.eu/search.html?qid=1557594184050&text=General%20Data%20Protection%20Regulation&scope=EURLEX&type=quick&lang=en

Nick Wallace, Daniel Castro. The Impact of the EU's New Data Protection Regulation on AI. March 27, 2018 CENTER FOR DATA INNOVATION

EU copyright reforms draw fire from scientists. Nature 556, 14-15 (2018) doi: 10.1038/d41586-018-03837-7 https://www.nature.com/articles/d41586-018-03837-7

Gauthier Chassang. The impact of the EU general data protection regulation on scientific research. ecancer 2017, 11:709 DOI: 10.3332/ecancer.2017.709 https://ecancer.org/journal/11/full/709-the-impact-of-the-eu-general-data-protection-regulation-on-scientific-research.php

European Parliament legislative resolution of 26 March 2019 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market. http://www.europarl.europa.eu/doceo/document/TA-8-2019-0231_EN.html

DIRECTIVE 2014/26/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0026&from=EN

Statutory Instrument No. 156/2016 - European Union (Collective Rights Management) (Directive 2014/26/EU) Regulations, Ireland, 2016.

The Minister for Jobs, Enterprise and Innovation signed into law statutory instrument number 156 of 2016, Guidance-Notes-Collective-Rights-Management-Regulations.pdf, https://dbei.gov.ie/.../Guidance-Notes-Collective-Rights-Management-Regulations.pdf

Edited by Xavier Seuba, Christophe Geiger, and Julien Pénin. Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data A CEIPI-ICTSD publications series. Issue Number 5 June 2018

Brussels, 8.4.2019 COM(2019) 168 final COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. Building Trust in Human-Centric Artificial Intelligence http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2019/0168/COM_COM(2019)0168_EN.pdf

Nestor Duch-Brown, BertinMartens, FrankMueller-Langer. JRC Digital Economy Working Paper 2017-01, The economics of ownership, access and trade in digital data. https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf

GitHub, www.github.com

CLARIN - European Research Infrastructure for Language Resources and Technology, www.clarin.eu

A Word Database for Natural Language Processing. Brigitte Barnett, Hubert Lehmann, Magdalena Zoeppritz. IBM Scientific Center, Heidelberg. https://www.aclweb.org/anthology/C86-1103

Meiryum Ali, The Best 25 Datasets for Natural Language Processing. https://lionbridge.ai/datasets/the-best-25-datasets-for-natural-language-processing/

Personal Data Processing Law, www.likumi.lv

Institute of Mathematics and Computer Science, University of Latvia. http://lumii.lv

Springer Copyright Act https://www.springer.com/authors/book+authors/helpdesk?SGWID=0-1723113-12-799504-0

ISO 37301 Compliance management systems - Requirements with guidance for use. https://committee.iso.org/sites/tc309/home/projects/ongoing/ongoing-3.htmlBeech, M. H. 1982. 'The Domestic Realm in the Lives of Hindu Women in Calcutta.' In *Separate Worlds: Studies of Purdah in South Asia*,edited by H. Papanek and G. Minault, 110–38. Delhi: Chanakya.