Law Data Science and Ethics: the CRIKE Approach

Silvana Castano[1], Mattia Falduti[1], Alfio Ferrara[1], and Stefano Montanelli[1]

Università degli Studi di Milano
DI - Via Celoria, 18 - 20135 Milano
{silvana.castano,mattia.falduti,alfio.ferrara,stefano.montanelli}@unimi.it

Abstract. In the era of big data, research activity on data science focuses on large datasets to produce knowledge supporting decision-making processes in different application domains and contexts. Data science practices and outputs have a tremendous impact on a variety of fields by raising new ethical issues that become crucial. In this paper, we address the ethical issues related to the ethics of data, the ethics of algorithms, and the ethics of practices in the context of our data science approach for case-law decisions (CLDs) processing called CRIKE (CRIme Knowledge Extraction). In particular, we discuss the ethical issues that need to be faced when dealing with knowledge extracted from CLDs for descriptive analysis purposes and for predictive usage of data extracted from CLDs.

Keywords: case law analysis, data science ethics, ethics of data

1 Introduction

In the era of big data, research activity on data science focuses on collection, processing, and interpretation of large datasets to produce knowledge for decision-making processes in different application domains and contexts. This is stimulated, on one side, and made possible on the other side, by the continuous production of data coming from disparate data sources and locations and by the availability of web-based technologies for data storage, integration, analysis and mining, thus enabling behavior and trend prediction as well as descriptive statistics for facts and events. Ethical issues play a crucial role in data science processes, to improve the social impact and the scientific quality of data science practices and outputs. For example, in [14], a framework is proposed for the enforcement of ethical oversight over the dissemination and use of Big and Open Data. The framework is grounded on the importance of encouraging critical thinking and ethical reflection among the researchers involved in data processing practices. As discussed by Floridi in [9], the main ethical challenges in data science can be classified as follows: i) *ethics of data*, focused on collection and analysis of large dataset; ii) *ethics of algorithms*, focused on complexity and

PIE 2019, June 4, 2019, Rome, Italy.

autonomy of algorithms, and iii) *ethics of practices*, addressed to draft ethical framework to shape professional codes, strategies and policies. On this ground, in the paper we address ethical issues in the context of our data science approach for case-law decisions (CLDs) processing called CRIKE (CRIme Knowledge Extraction). The CRIKE approach has been conceived for processing large datasets of CLDs coming from diverse law sources (e.g., first grade, Court of appeal) to automatically discover applications of legal abstract term's in court's decision texts. CRIKE relies on the LATO ontology where abstract terms and decision verdicts are formally defined by means of concepts and relations. A detailed description of the LATO ontology design and of the CRIKE knowledge extraction processes is provided in [5]. The CRIKE process workflow covers all the phases of a conventional data science process: i) *data collection*, where CLDs are collected and stored in digital format for subsequent analysis, ii) *knowledge extraction*, where CLDs texts are processed to extract knowledge in form of relevant terminology corresponding to the concepts in the LATO ontology, iii)*target-oriented practices*, where knowledge extracted from CLDs can be exploited both for descriptive analysis purposes by classifying CLDs and for predictive usage of CLDs by enforcing learning procedures.

Ethical issues of different nature and different impact and implications are involved in processing CLDs using CRIKE. As a general consideration, we observe that CLDs involve individuals like judges, ascribed/accused people and possible other individuals intervening in the crime description (e.g., witnesses). Prominent ethic issues in processing of CLDs should thus avoid: i) violation of individual privacy as well as prohibited secondary uses of personal data; ii) individual classification based on data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, as well as trade union membership, genetic and biometric data, data concerning health or data concerning sex life or sexual orientation; iii) unfair of prediction algorithms concerning CLDs analytics approaches focused not only on pure data analysis, but also on court's outcomes prediction, judges profiling and automatic legal decision's making. For example, an ethical issue envisaged in [26] is propensity, that is, on the basis of prediction about what people were likely to do, what could/should be done to prevent. As discussed in [26], what if big data analytics predict that a certain person has a likelihood of 95% to being involved in domestic violence? An ethical issue here has to do with the ethical role of those setting the threshold and the data scientists writing the algorithm that calculates the chance based on the observation of certain variables available in the underlying dataset.

After describing the overall CRIKE process workflow (Section 2), goal of the paper is to provide a finer classification of ethical issues involved in the CRIKE process workflow by referring to the classification introduced in [9] and its actualization in the framework of the CRIKE (Section 3). Finally, we conclude by discussing our future work (Section 4).

2 The CRIKE approach to Case-Law Decisions Processing

The CRIKE approach to CLDs processing is articulated in six main activities as shown in Figure 1. The Collection of CLDs activity is devoted to the tasks/procedures used for acquiring and preprocessing CLDs from a qualified source, like for instance the Court of Milan. Usually, in the Italian context, CLDs are provided in form of images of the paper documents. The quality of these documents is highly variable. Thus OCR and other ad hoc solutions for data cleaning are required to obtain a pure text version of each CLD toghter with a limited set of metadata (including a CLD identifier and the date). In the Storage of CLDs activity, digital documents are stored in a database. In CRIKE, we exploit MongoDB to store for each CLD, the raw text, the available metadata, as well as the sentences and single words obtained from sentence and word tokenization of the raw CLD text.

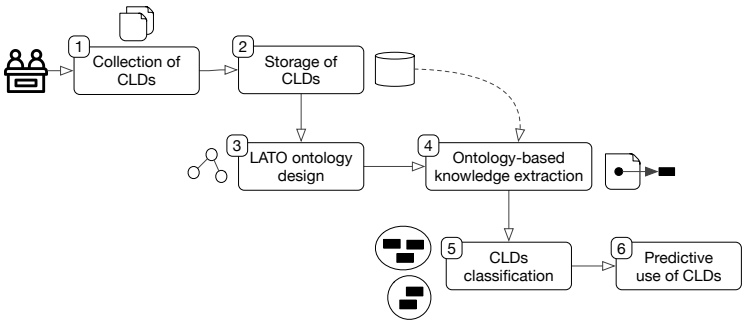

Fig. 1. The CRIKE approach to CLDs processing

CRIKE is based on the LATO ontology which drives the process of knowledge extraction from CLDs. The third activity is the LATO ontology design, with the goal of conceptualizing legal concepts and related controlled vocabulary. Then, working with LATO and with the contents of the CLD database, we extract knowledge from the CLDs (Ontology-based knowledge extraction activity). Goal of this activity is to retrieve occurrences of the legal concepts as they are defined in LATO within the CLD document collection and to extract relevant terminology used by the judge to articulate those concepts in each specific CLD. Knowledge extracted from CLDs constitutes the input for subsequent activity of CLDs classification (Fig.1.5). The goal is to classify CLDs according to the concepts of interest in LATO, to measure the relevance of terms extracted from CLDs text with respect to LATO concepts, and to associate terminology with the final decision of the judge. This activity is the basis for calculating a degree of correlation between terminology, concepts, and decisions. According to this

analysis, it is then possible to enforce learning procedures to make a predictive use of CLDs with respect to specific legal concepts (Predictive use of CLDs). Both activities 5 and 6 are target-driven in that classification and predictive use of CLDs are customized according to the final use of CLDs data (e.g., to study the interpretation given by courts to a specific legal concept, predict a decision given some facts).

3 Dealing with ethics in CRIKE

To highlight and discuss ethical issues in processing CLDs, we map the data science ethics framework proposed by Floridi in [9], on the CRIKE activity workflow resulting in the three-layer framework shown in Fig.2): i) *ethics of data*, involving ethical issues related to collection and analysis of large CLDs dataset; ii) *ethics of algorithms*, involving ethical issues related to complexity and autonomy of CRIKE algorithms, and iii) *ethics of practices*, more strictly related to ethics in target oriented classification and prediction activities of CRIKE.

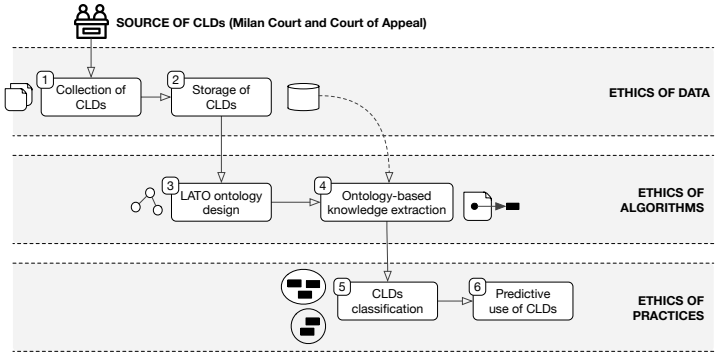

Fig. 2. Three-layer framework of CRIKE ethical issues

3.1 Ethics of data

Ethics of data primarily refers to the source providing data as well as to the procedures used for data acquisition and storage. In terms of data acquisition, working in the legal domain, in particular the Italian legal domain, imposes us to acquire data from a specific, secure and certified source. Both laws and CLDs have an institutional creator which should be accessed by directly interacting with the public administration offices in order to acquire genuine data in terms of data format and completeness. In CRIKE, we process CLDs obtained directly by the involved Courts (the Court of Milan and the Court of Appeal). The direct access to the administration databases guarantees the institutional provenance

of data as well as their integrity. A second relevant issue concerning ethics of data involves personal data. In particular, criminal CLDs may contain three different categories of personal data, namely (i) identification data, (ii) special categories of personal data and (iii) criminal records. Identification data are defined by the General Data Protection Regulation (GDPR) as those data describing an identifiable person [7]. Special categories of personal data are described at paragraph 9 of the GDPR as "those data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, as well as trade union membership, genetic and biometric data, data concerning health or data concerning sex life or sexual orientation". Criminal records are the records concerning a person's criminal history. This last category of personal data is protected at paragraph 10, where GDPR specifies that "access to those data is permitted only under the control of an official authority or when the processing is authorized by European Union or Member State law". The aim of the regulation is to protect personal data against illicit handlings. In particular, main ethical issues related to CLDs acquisition and storage concern the risk associated both to the privacy of groups of people and to re-identification of individuals. Specifically, the risk associated with groups regards the possibilities to combine data and groups of individuals, for example, by committed crime, by race or nationality, by spoken language or dialect, by age or gender. These activities could violate groups privacy and could permit re-identification through inference [8]. Concerning re-identification of individuals, the main risk is to violate the right of being forgotten, as drafted in [4]. These issues are faced in different research fields. For example, [3] presents an estimation of re-identification risk for data sharing policies of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, as well as an evaluation of the risk of a specific re-identification attack using voter registration lists. In general, uncontrolled re-identification risks can conduct to a dangerous information control loss and privacy violation, due to the fact that information privacy concerns specifically the capacity of an individual to maintain control of his or her information [25]. Since privacy regulation is based on the notion of meaningful consent, having trust in data acquisition and processing is a crucial issue [22]. In particular, the topic of privacy in accessing individual criminal history information is addressed in [12], where the authors define policies for providing public access to individual criminal records in Spain and the USA, considering access to court records, protection of honor, privacy and personal data, free speech and rehabilitation. In this context, CRIKE is compliant with the privacy regulation in that it is conceived to detect exclusively legal concepts inside the CLDs and to group the CLDs by legal concepts and their application. Secondly, we want to extract legal knowledge by automatically considering the verdicts. In other terms, we consider only legal terminology and crime argumentation. Our goals are not related to personal data, directly or indirectly. Knowledge extraction and text mining activities are only related to find legal concepts application and how they are expressed by judges inside various CLDs. Moreover, due to the particular type of data and the agreement we signed with the involved Court administrations, our dataset is closed and it cannot be

shared nor published. The CLDs database is protected against external attacks, in that it is stored on stand alone machine accessed only by a restricted number of authorized researchers with given time restrictions. These restrictions were mandatory to sign the agreement with the involved public administration offices, for CLDs acquisition and use. We note that our dataset avoids the group privacy issues in that we obtained a whole set of CLDs, rather than only selected CLDs targeted to a specific topic/objective to be analysed, like for instance all CLDs related to a specific crime or to a specific group of crimes.

3.2 Ethics of algorithms

The ethical issues related to design and implementation of algorithms that elaborate criminal data are transparency, accountability and discrimination. First, in terms of transparency, the risk is to use or implement processes and algorithms that are unclear, incomprehensible and unrepeatable [24]. Transparency is related to the concepts of accessibility and comprehensibility of information, as reported in [21, 24]. Real-world algorithmic decision-making processes designed to maximize fairness and transparency are described in the Open Algorithm (OPAL) project [15]. Transparency itself is insufficient, on one side, because companies would not reveal and disseminate proprietary algorithms not to lose their competitive edge, and on the other side, because of the so-called transparency paradox [19]. This refers to the fact that, it is clear what machine learning algorithms do in taking decisions about, for example, credit, medical diagnose, personalized recommendations, advertising or job opportunities, but it is still less clear how these decisions are taken [23]. This issue is directly related to accountability, which is the problem of associating the blame for problems and errors of very complex systems to specific individuals [13, 17].

A further issue to be addressed is how and to whom to enforce accountability for discriminatory outcomes of data analysis. Handling criminal data means in fact to face the risk of associating a criminal behavior with groups of individuals on the basis of their race, religion, cultural background, language, age or gender. An example of data mining discriminatory outcome in ranking job candidates is described in [2]. Authors demand caution in the use of data mining techniques and they advocate that this should be part of a comprehensive set of strategies for contrasting discrimination in the workplace and for promoting fair treatment and equality. Other interesting contributions on this issue are the idea of Classification with No Discrimination (CND) [10] and the proposal of a guideline for researchers and anti-discrimination data analysts on concepts, problems, application areas, datasets, methods, and approaches from a multidisciplinary perspective, as presented in [20]. A discussion of algorithm fairness issues on criminal data analysis and racial disparities, in particular focusing on the problem of designing an algorithm for pretrial release decisions, is given in [6]. Since CRIKE knowledge extraction enforces an ontology-based approach with LATO, we comply with the need of transparency in terms of comprehensibility and human intervention. In particular, we decided to base the process of knowledge extraction mainly on quite simple functionalities for searching LATO

terminology within the CLDs documents in order to guarantee a transparent and easily repeatable process. We handle CRIKE accountability issues by arguing that LATO can be changed and modified directly by the designer, to influence CRIKE results. Moreover, the system is open and still under definition. Our goal is to preserve human intervention and direct control over the system behavior and over the achieved results. Furthermore, in order to avoid the reported discriminatory risks, we base knowledge extraction and classification processes only on general legal concepts and application, by considering for instance crime paragraph, article, verdict and the related terminology.

3.3 Ethics of practices

The issues concerning the ethics of practices are related to the use of the outcomes of data analysis. In particular, we need to face risks concerning anonymity and informed consent, secondary use, and data protection. Informed consent appears insufficient to solve ethical problems related to individuals privacy as discussed in [1], where authors point out how privacy and big data are simply incompatible without a definition of new approaches having anonymity has one of their primary goals since the design. In particular, they point out how anonymity is different from nameless and reachability. About secondary use, the aim is to ensure ethical practices fostering both the progress of data science and the protection of the right of individuals and groups, as pointed in [14]. An example of the question of privacy and secondary use of data in health research is given in [16] by considering three different levels: informed consent, anonymity, and public interest mandate. In health research, the reuse of clinical data is a fast-growing field, recognized as essential to: i) realize the potentials for high-quality healthcare, ii) improve healthcare management, iii) reduce healthcare costs, and iv) perform effective clinical research ([18]). In particular, one of the main issues in this field is the trade-off between the need of keeping personal data anonymous and the need of exploiting data to achieve results that could be useful for the citizens, according to the notion of public interest. An example is available in [11], where authors describe two court cases (appeared in US and UK) about selling prescription data and the related questions of what constitutes privacy and what public interest. Balancing privacy, public interest and open access raises ethical and juridical questions in the legal field as well, because Criminal Courts declare in their decisions what is forbidden and what is allowed. Thus, according to the European Court of Human Rights, criminal argumentation reported in CLDs has to be published, accessible, and known by individuals. CRIKE's results achieved so far are completely anonymized and do not report any personal or identifying data, because CRIKE works exclusively with legal concepts formalized in LATO. The CRIKE system has a scientific research aim only and it respects the GDPR rules for scientific research purposes. We mine CLDs in order to extract the legal argumentation and the juridical terms application, by considering the diffusion of the legal knowledge as a positive element. For these reasons, we aim at facilitating the access to legal knowledge without pursuing goals of judge profiling or similar.

4 Future work

Our work on CRIKE is ongoing. So far, we achieved first and promising results in automatically extracting and classifying CLDs terminology concerning drug-related crimes. In particular, we focused on legal abstract terms formalization in LATO in thus context. In law articles, legal abstract terms represent something indeterminate that need a concrete application to be defined; examples of abstract terms are *good faith*, *long-term cohabitation*, or *minor offense case*, where what should be considered *good*, *long-term*, or *minor* requires a concrete interpretation by the Court in order to be defined. In this context, we defined a CRIKE process to detect concrete applications of legal abstract terms in CLDs and to determine how and where considered legal abstract terms are applied by judges in their legal argumentation. As discussed in the paper, CRIKE has been designed from the very beginning to be compliant with guidelines and regulations concerning the ethical issues in the field of data science. Our future work will keep this as a primary goal of CRIKE. In particular, we aim at evolving the LATO ontology to include further legal concepts and related terminology by systematically testing the capability of the system to detect and classify CLDs against them. Moreover, we aim at exploiting the use of machine learning techniques to automatically enrich LATO starting from the training set composed by the CLDs that have been classified through the ontology-driven approach, thus enforcing a bootstrapping mechanism where each cycle of knowledge extraction and classification is used to improve the ontology and the subsequent extraction cycle. Finally, we aim at studying the correlation between the concrete application of legal abstract terms and the final Court decision, in order to apply a predictive approach for determining an expected verdict given the concrete facts that are related to each specific legal concept of interest.

References

1. Barocas, S., Nissenbaum, H.: Big Data's End Run around Anonymity and Consent, p. 44–75. Cambridge University Press (2014)
2. Barocas, S., Selbst, A.D.: Big data's disparate impact. Californi Law Review **104**, 671 (2016)
3. Benitez, K., Malin, B.: Evaluating re-identification risks with respect to the hipaa privacy rule. Journal of the American Medical Informatics Association : JAMIA **17**, 169–77 (03 2010). https://doi.org/10.1136/jamia.2009.000026
4. Bennett, S.C.: The right to be forgotten: Reconciling eu and us perspectives. Berkeley Journal of International Law **30**, 161 (2012)
5. Castano, S., Falduti, M., Ferrara, A., Montanelli, S.: Crime knowledge extraction: An ontology-driven approach for detecting abstract terms in case law decisions (2019), 17th International Conference on Artificial Intelligence and Law (ICAIL)
6. Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., Huq, A.: Algorithmic decision making and the cost of fairness. In: Proc. of the 23rd ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining. pp. 797–806. ACM (2017)
7. EU Parliament and Council of European Union: General Data Protection Regulation (May 2016), http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

8. Floridi, L.: Open Data, Data Protection, and Group Privacy. Philosophy & Technology **27**(1), 1–3 (2014)
9. Floridi, L., Taddeo, M.: What is data ethics? Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences **374**, 20160360 (12 2016). https://doi.org/10.1098/rsta.2016.0360
10. Kamiran, F., Calders, T.: Classification with no discrimination by preferential sampling. In: Proc. 19th Machine Learning Conf. Belgium and The Netherlands. pp. 1–6. Citeseer (2010)
11. Kaplan, B.: How should health data be used?: Privacy, secondary use, and big data sales. Cambridge Quarterly of Healthcare Ethics **25**(2), 312–329 (2016)
12. Karst, K.L.: "The Files": Legal Controls over the Accuracy and Accessibility of Stored Personal Data. Law and Contemporary Problems **31**(2), 342–376 (1966)
13. Kraemer, F., Van Overveld, K., Peterson, M.: Is There an Ethics of Algorithms? Ethics and Information Technology **13**(3), 251–260 (2011)
14. Leonelli, S.: Locating Ethics in Data Science: Responsibility and Accountability in Global and Distributed Knowledge Production Systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **374**(2083), 20160122 (2016)
15. Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P.: Fair, Transparent, and Accountable Algorithmic Decision-Making Processes. Philosophy & Technology **31**(4), 611–627 (2018)
16. Lowrance, W.: Learning from Experience: Privacy and the Secondary Use of Data in Health Research. Journal of health services research & policy **8**(1_suppl), 2–7 (2003)
17. Matthias, A.: The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. Ethics and information technology **6**(3), 175–183 (2004)
18. Meystre, S., Lovis, C., Bürkle, T., Tognola, G., Budrionis, A., Lehmann, C.: Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. Yearbook of medical informatics **26**(01), 38–52 (2017)
19. Nissenbaum, H.: A Contextual Approach to Privacy Online. Daedalus **140**(4), 32–48 (2011)
20. Romei, A., Ruggieri, S.: A Multidisciplinary Survey on Discrimination Analysis. The Knowledge Engineering Review **29**(5), 582–638 (2014)
21. Rubel, A., Jones, K.M.L.: Student Privacy in Learning Analytics: an Information Ethics Perspective. The Information Society **32**(2), 143–159 (2016), https://doi.org/10.1080/01972243.2016.1130502
22. Schermer, B.: The Limits of Privacy in Automated Profiling and Data Mining. Computer Law & Security Review **27**(1), 45–52 (2011)
23. Spice, B.: Carnegie mellon transparency reports make ai decision-making accountable. Tech. rep., Carnegie Mellon University School of Computer Science (2016), https://www.cs.cmu.edu/news/carnegie-mellon-transparency-reports-make-ai-decision-making-accountable
24. Turilli, M., Floridi, L.: The Ethics of Information Transparency. Ethics and Information Technology **11**(2), 105–112 (2009)
25. Van Wel, L., Royakkers, L.: Ethical Issues in Web Data Mining. Ethics and Information Technology **6**(2), 129–140 (2004)
26. Zwitter, A.: Big Data Ethics. Big Data & Society **1**(2) (2014)