

## Ethical Perspectives in AI: A Two-folded Exploratory Study From Literature and Active Development Projects

José Antonio Siqueira de Cerqueira University of Brasília(UnB) Brasília, DF, Brazil <a href="mailto:antonio.cerqueira@aluno.unb.br">antonio.cerqueira@aluno.unb.br</a>	Lucas dos S. Althoff University of Brasília(UnB) Brasília, DF, Brazil <a href="mailto:lucasa@aluno.unb.br">lucasa@aluno.unb.br</a>	Paulo Santos de Almeida University of Brasília(UnB) Brasília, DF, Brazil <a href="mailto:paulo.almeida@redes.unb.br">paulo.almeida@redes.unb.br</a>	Edna Dias Canedo University of Brasília(UnB) Brasília, DF, Brazil <a href="mailto:ednacanedo@unb.br">ednacanedo@unb.br</a>
---	---	--	---

### Abstract

**Background:** Interest in Artificial Intelligence (AI) based systems has been gaining traction at a fast pace, both for software development teams and for society as a whole. This increased interest has led to the employment of AI techniques such as Machine Learning and Deep Learning for diverse purposes, like medicine and surveillance systems, and such uses have raised the awareness about the ethical implications of the usage of AI systems. **Aims:** With this work we aim to obtain an overview of the current state of the literature and software projects on tools, methods and techniques used in practical AI ethics. **Method:** We have conducted an exploratory study in both a scientific database and a software projects repository in order to understand their current state on techniques, methods and tools used for implementing AI ethics. **Results:** A total of 182 abstracts were retrieved and five classes were devised from the analysis in Scopus, 1) AI in Agile and Business for Requirement Engineering (RE) (22.8%), 2) RE in Theoretical Context (14.8%), 3) Quality Requirements (22.6%), 4) Proceedings and Conferences (22%), 5) AI in Requirements Engineering (17.8%). Furthermore, out of 589 projects from GitHub, we found 21 tools for implementing AI ethics. Highlighted publicly available tools found to assist the implementation of AI ethics are InterpretML, Deon and TransparentAI. **Conclusions:** The combined energy of both explored sources fosters an enhanced debate and stimulates progress towards AI ethics in practice.

### 1. Introduction

There is an increasing number of software development teams building Artificial Intelligence (AI) based systems, and they are gaining popularity in our society at a fast pace [1, 2]. The use of AI techniques like Machine Learning (ML) and Deep Learning (DL) on diverse fields such as medicine, surveillance systems, business, transportation, and many other domains, have

raised great awareness about the ethical implications of the use of such systems [3, 2], becoming subject of urging interest to the industry, researchers in academia, and the population at large [4].

Whilst AI popularization is growing, incidents related to those AI-based systems are also becoming more common [3]; several notorious incidents have led to public discussion on AI ethics. One such case is the Cambridge Analytic scandal, where data from Facebook users were obtained inappropriately and used to influence the result of an election [5]. Another example is a biased ML algorithm against women by Amazon, which led to more male candidates being hired [6]. Also, new threats rise concerning ethical misuse of AI bases system such as fake news with the use of deep-fake and AI-based voice technologies, where someone's face could be superimposed on videos and political leaders can be depicted inciting violence and panic, for instance, may be used to rig elections, change political opinions and spread misinformation in general [7].

Various ethical guidelines and principles for AI have been proposed by organizations, commissions, institutes and the industry. Those propositions, however, do not meet the demands from real world development of ethical AI-based systems, as these ethical principles are often too high level, abstract and general [8, 9, 10] and pose no real evidence that they can influence ethical decision making [11]. As a result, those in charge of developing such AI-based systems who are also concerned with the ethical questions that come up have become frustrated by how little help is offered by the highly theoretical texts provided by the principles and codes available [2]. Hence, developing ethical AI is an overwhelmingly defying and complicated task [7]. Most of the studies found in the literature focus to a large extent on theoretical and conceptual principles and guidelines, not providing an effective and realistic framework on how to implement ethics in AI [3, 2]. Therefore, a deeper focus on technological details of the various methods and technologies in the AI and

ML area is needed; in other words, currently there is a need to decrease the distance between abstract values (principles, guidelines and codes) and technical implementations [8, 2].

The main objective of this work is to identify tools, methods and techniques already publicly available to assist practitioners involved in the development of AI-based systems to implement ethical principles in their work, hence shedding some light on the topic of applied ethics in AI and bridging the gap between said principles and practice. We devised five classes from our scientific database analysis in Scopus, 1) AI in Agile and Business for Requirement Engineering (RE) (22.8%), 2) RE in Theoretical Context (14.8%), 3) Quality Requirements (22.6%), 4) Proceedings and Conferences (22%), 5) AI in RE (17.8%). The combined collaboration of both scientific and open-source sources fosters a broadened debate on this topic.

## 2. Background and Related Work

Various institutions and organizations from public and private sectors presented guidelines and principles towards ethical AI-based systems. Two of them will be highlighted in this part of the study: 1) The Ethically Aligned Design [12] by the Institute of Electrical Electronics Engineers (IEEE), whose first edition was published in 2019, presenting analyses and recommendations as a guidance for governments, business and public to take as consideration when dealing with the advancement of AI for the benefit of humanity; 2) Trustworthy AI by the High-Level Expert Group on AI from the European Commission, which was presented on April 2019 [13], showing a set of 7 key requirements that AI systems should meet in order to be considered trustworthy: a) Human agency and oversight; b) Technical robustness and safety; c) Privacy and data governance; d) Transparency; e) Diversity, non-discrimination and fairness; f) Societal and environmental well-being; g) Accountability.

In the academia there are several authors who are currently researching on ethical principles and guidelines. In other words, plenty of reviews and surveys about ethical guidelines, frameworks and principles are available in the current literature. Jobin et al. [14] introduced a comprehensive mapping of the current AI ethics landscape on 84 guidelines proposed worldwide by the private and public sectors providing an overview of the most relevant principles among them. They argue that there is a fast increase in the number and variety of documents that evinces the growing interest by the international community for ethics in AI, but the proposed principles and guidelines have a significant

divergence on how to achieve ethical AI. A major cause of divergence among them is how they should be implemented.

Floridi et al. [15] provided a synthesis of six sets of guidelines, extracting 47 principles that overall have a great degree of coherence and overlap among them. The authors state that the four core principles commonly used in bioethics: 1) beneficence, 2) non-maleficence, 3) autonomy, 4) justice and a fifth one, explicability, are greatly adapted to ethical challenges in AI. They seized the significance of each of the 47 principles, forming an ethical framework, within which they offer a list of recommendations with 20 items. The work, however, is majorly conceptual and oriented to government policies, not providing technical solutions to developers. Rothenberger et al. [16] presented a survey and evaluation of guidelines for AI ethics, providing a sum of 5 principles and a ranking of these principles through interviews with 51 experts. They argue that responsibility was ranked first, but respondents asked who would be responsible for the actions of an AI.

Hagendorff [8] analyzed and compared 22 guidelines, finding that almost all guidelines suggest that technical solutions exist, yet not providing technical explanations. Moreover, he states that to deduce concrete technological implementations from very abstract ethical principles is a major problem. Therefore, he considers that there is an urge to close the gap between ethics and technical discourses. McNamara et al. [11] surveyed 63 software engineering students and 105 professional software developers to understand the impact of the Association for Computing Machinery (ACM) Code of Ethics and Professional Conduct [17]. Surprisingly, the authors concluded that there is no evidence that the ACM code of ethics influences ethical decision making in software development. Thus, we consider that discussion regarding theoretical ethics in AI is already consistent and further discussion on this theme will not be approached in this study, as there seems to be an agreement concerning theoretical ethical aspects [2]. On the other hand, there is an urge to perform the translation between the 'what' and the 'how' in AI ethics [2].

A limited number of works available argue about practical ethical AI. Most of literature do not propose a method or tool to implement ethics in AI, rather they survey available tools or perceptions from practitioners. Morley et al. [2] presented a typology and a catalog of available tools and methods to translate principles into practices, in the ML field. However, most of the tools found lack good documentation and focus on small portions of the software development process. Hence, despite being promising, they demand extra work

before use. In addition, the authors pinpointed some possible opportunities to researchers such as provide an assessment of the catalogued tools and examples of its usage.

Vakkuri et al. [3] surveyed 211 software companies in an attempt to understand the current state of practice of ethics in AI in the industry, arguing that it is still in its early stages. The authors suggest that a starting point to implement ethics in AI is by the use of the guidelines available. Nonetheless, those are not practical for developers, thus requiring additional work before they can be carried into a real system. The authors concluded that practitioners have a key role in implementing ethics in AI, once activities in AI software projects are nearly the same as in any other software project. In addition, AI ethics implementation from a software development viewpoint could be seen as a non-functional requirement of an AI based system, and that the Product Owner has the responsibility to ensure ethical User Stories in sprint backlogs. Finally, the authors indicated three issues to be avoided regarding applied AI ethics: 1) Do not outsource ethics in AI software development; 2) Do not assume that ethics can implemented without being systematically done; 3) Do not delegate ethics implementation to a single person.

Another recent research by Vakkuri et al. [18] surveyed applied ethics in AI in the start-ups context. The authors discovered that several Software Engineering practices – well established and in existence – can be used to implement AI ethics, e.g., documentation. Besides, despite practitioners having ethical concerns, AI ethics is largely not being implemented, partially being a result of a lack of formal methods and tools to implement it. Despite the existence of the catalogued tools in Morley et al. [2], the authors focused on small parts of the development process, have little usability, and for what concerns AI ethics, there is a need for it to be addressed from the beginning of development, that is, from the requirements elicitation stage [3].

Requirements Engineering (RE) is seen as the first stage of the software development life cycle that deals with the elicitation, analysis, specification, and validation of software requirements as well as the management and documentation of requirements. To start a discussion about ethical requirements in AI, first we need to address RE in AI in a general manner. Vogelsang and Borg [1], in an attempt to understand the perspectives from data scientists regarding RE for ML, stated that there is not much work on RE for ML systems, while literature suggests that RE is the most difficult activity for the development of ML-based systems. Their main findings include that requirements

engineers needs to be conscious about new requirements introduced by the ML paradigm, which are explicability and freedom from discrimination. The first type of requirement was mentioned as important quality requirement in their interviews. However, the second type seems more problematic once ML algorithms are designed to identify recurring patterns in data applying these patterns to judge about concealed data. They pointed two reasons for this last statement, first that discrimination is more implicit in ML systems and second that ML algorithms enlarge discrimination bias in the data during the training process. Moreover, all interviewees mentioned challenges concerning ethics and legal aspects.

Belani and Car [19] proposed a RE4AI taxonomy, that is, RE for AI, bringing forth an overview of challenges posed to RE towards building AI-based complex systems. We highlight that, for elicitation activity of RE, the authors defined regulation – ethics – not clear, as a problem related to the system to be. Kostova et al. [20] stated that "RE is the only place to address problems related to the use of AI based systems due to its interdisciplinary nature, with a strong technical emphasis". Their work identified two faces of RE discipline in AI, first AI tools are used more frequently during the RE process (AI4RE), second the RE process for AI based systems is different (RE4AI). In this research, our interest resides in the second aspect.

Few works discuss about RE for ethical requirements of AI. Guizzardi et al. [9] presented a definition of Ethical Requirements as the ones derived from guidelines and ethical principles. And the key concept behind it is of Runtime stakeholders, defined as persons that are using, are affected by, or influencing the results and outcomes of an AI based system while in operation. The authors argued that Runtime stakeholders are often ignored in traditional RE. More importantly, they stated that "Ethical requirements are functional and quality requirements elicited from Runtime stakeholders in accordance with the five ethical principles – beneficence, non-maleficence, autonomy, justice, explicability" [9], the same principles provided by Floridi et al. [15]. Their main goal is to use traditional RE techniques to derive ethical requirements to the case of driver-less cars, to make sure they comply to ethical principles. However, they do not present a systematic methodology employed to do so, neither a validation of the technique.

Vakkuri et al. [21] introduced a tool as a starting point for implementing ethics in AI, named ECCOLA. The proposed tool is a deck of cards to raise awareness of AI ethics in a development team, once the team produces documentation of their

ethical decision-making, as for example in the form of non-functional requirements in product backlog. However, there is no validation method of the proposed tool yet. Aberkane [10] in his Master's thesis performed a systematic literature review of ethics and requirements engineering in IEEE Xplore. Using the ACM/IEEE-CS Code of Ethics as a guide the author presented an extensive list of ethical issues identified in the literature. The author does not address in-depth analyses on ethical issues in AI through his study.

Our paper aims to broaden the discussion on ethics in AI by exploring a scientific database (Scopus), also a database for open source projects (GitHub), in order to understand the ethical requirements of AI in academia and the present state of such requirements in publicly available software projects. To the best of our knowledge, no study has approached GitHub for ethical requirements in AI. Moreover, we found that it was already conducted a bibliometric research in Web of Science database to explore sustainable requirements in AI [22]; thus we choose Scopus, as it is a large and well known scientific database, with different sets of criteria and analyses.

### 3. Methodology

Our research strategy is based on two different approaches. The first one is based on the Theory of Consolidated Meta-Analytic Approach (TEMAC) method [23], from which we only implement few steps aimed at extracting the fundamental literature inside the scope of our study, and to retain the main features about the topography of this field of research. The second one is based on mining GitHub for projects related to ethics in AI, where we explore practical implementations and relate them to the findings inside the field of research. In order to explore a scientific database to extract valuable information from bibliometric data we will use the Theory of Consolidated Meta-Analytic Approach, that provides an objective technique that allows metrics to be established between literary researches in the same field through rigid systematization of research meta-data. We will not provide a full analysis over the TEMAC method, rather we will highlight important steps used in this research.

The whole method is comprised of three stages. The first stage (Preparing the research) is to define the correct keyword, the year range, the scientific databases for extraction, and the area of knowledge, to the study. The second stage (Presenting and inter-relating the data) is to extract bibliometric information from the databases and the application of bibliometric laws to analyse relations between them. The third and last stage (Integrating

and validating models) to be used on the evidence acquired from Citation and Bibliographic Coupling mapping study. Highlighting main studies, approaches and lines of research – VosViewer and Iramuteq tools are used to present graphical analyses. In this paper, we will prune parts of TEMAC, which is advised by its creators. In other words, we will present analyses over the first stage, due to its nature of preparing the research, while in the second stage, where there is a wide set of possible analyses, we will focus on a small set suitable to the aims of this research. In addition, the third stage presents an important in-depth study over the data obtained, hence leading to a better understanding.

The second approach explores GitHub – a major platform that hosts open source software projects – that became the largest open source community in the world. In each repository in GitHub a README file is present – one of the first documents that developers sees when coming into contact with a new repository – informing other people the usefulness of a project, what they can do with it, and how they can use it [24]. Looking forward to track traces of how AI projects assimilate ethical principles and concepts we perform a qualitative analyses over GitHub README files. With the study of Portugal et al. [25] as basis, in this work a similar approach is devised adapting their proposed methods to our goal. Further, we detail steps that will be performed on the Section (4) to perform our investigation and analyses of README files from GitHub:

1. Retrieve a corpus of README files from GitHub corresponding to a query using a keyword. We used the tool Corpus Retrieval, presented by Portugal et al. [26];
2. Manually categorizing the README files into different types;
3. Highlight tools found in previous step;
4. Discover the most relevant keyword in the corpus set. This is done filtering out supplementary forms and unmeaningful active forms, and through the generation and analysis of a word cloud;
5. Expand analysis through the use of POS-Tagging technique in the whole corpus, retrieving frequency of verbs that are commonly related to requirements – RE patterns candidates;
6. Create a separate sub corpus for each RE pattern;
7. Extract information concerning requirements, by manual reading, and listing, for each sub corpus produced;
8. Categorize found requirements into ethical principles in AI. In order to provide a better visualization, Figure 1 shows how we approach exploring GitHub's README files.

### 4. Results and Analyses

An initial step to perform bibliometric analyses over a scientific database is the preparation and definition

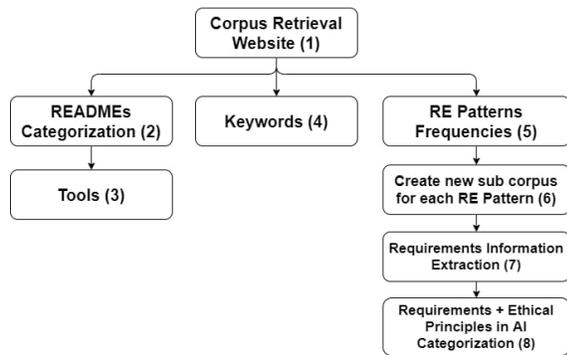


Figure 1. Methodology for exploring GitHub

of parameters for the research. To reduce meaning suppression as well as to capture the majority of works associating AI, ethics and software requirements we designed a search pattern divided in these three branches in the Scopus database: (“ethic” OR “ethics” OR “ethical” OR “moral” OR “code of conduct” OR “transparency” OR “security”) AND (“requirement analysis” OR “requirements specification” OR “requirement elicitation” OR “software development” OR “requirement engineering”) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR “predictive model” OR “DL” OR “ML” OR “AI”).

Each branch is composed by a set of interchangeable words selected to retain all aimed subtopics inside the database. The conduction of this research was from 10th of May to 14th of June. The year range was limited to the last ten years. Neither area of knowledge nor kind of document were limited, since we do not want to filter the applicability of the research. The query retrieved 182 results.

To have a broader view from the evolution of the research area we expand the year range to 15 years with a total of 209 publications, as shown in Figure 2. Regarding the evolution of publications over time, it can be seen that the last three year contains 69% of the scientific production in the area. Citation evolution also reflects the attention this topic is gaining in the last years, as shown in Figure 3. Different groups of research in a wide range of countries and institutions are active in research areas that expose ethics requirements within AI context. Regarding publications by country, as shown in Figure 4, the USA stand out with over two times India’s and UK production; we also highlight the presence of Brazil with the 10th production. Even though USA has the aforementioned high production, Italy has a higher concentrated, since the top institution in production, as show in Figure 5. Figure 4 and 5 only depict the top 10 countries and institutions, respectively.

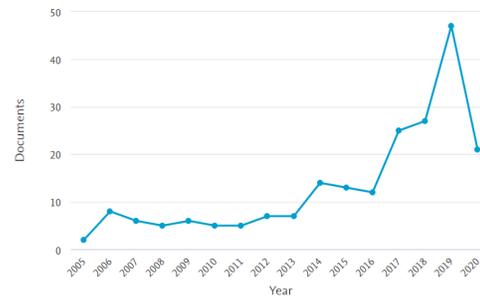


Figure 2. Number of publications in Scopus by year.

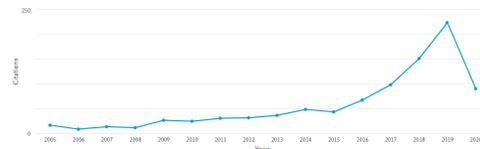


Figure 3. Evolution of amount of citations.

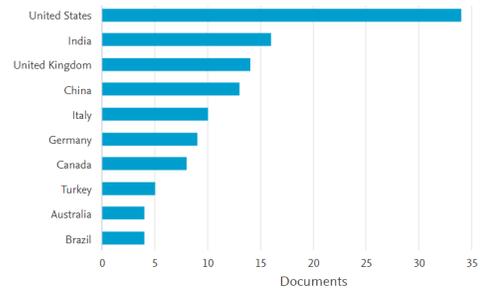


Figure 4. Records by country.

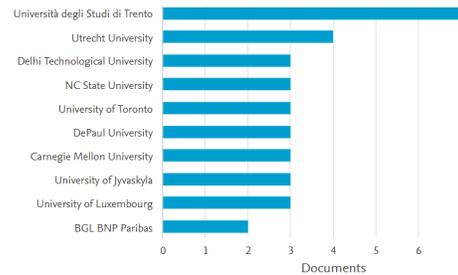
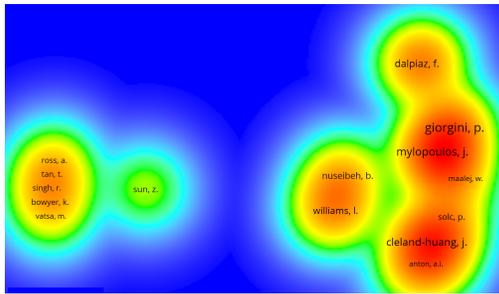


Figure 5. Records by institution.

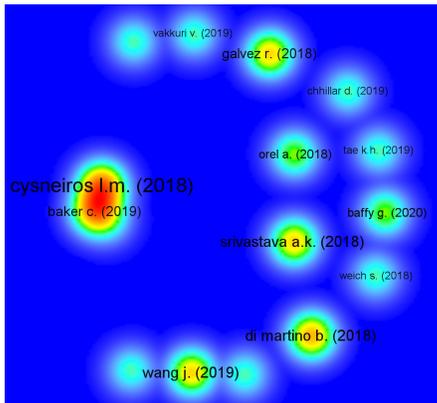
To detail the topography of the research area we applied the author co-citation map analysis, as shown in Figure 6. This map has roughly 3 clusters, indicating three group of authors sustaining this research area. The most critical author are Giorgini, P. who proposed the “STS-tool” in 2012, which helps engineers to specify Socio-technical Security Requirements through social commitments. This kind of work are directly related to Mylopoulos, J.’s works in security and privacy requirements specifying ownership, permission, delegation and others. Other important cluster are

represented by Cleland-huang, J. with studies in automated tracing, product listing and recommendation, such works are related to data security and transparency. Finally, the cluster represented by Tan, T., Singh, R. and Sun, Z. can be considered as the biometric branch of the research topography, with studies in iris recognition and localization, closely related to ethics and policy of biometric systems because such systems are associated with spoofing attacks and have implications in security requirements.



**Figure 6. Co-citation density map. Source: Scopus. Map generated by VOSViewer.**

To answer what are the tendencies of this research area we applied the Bibliographic Coupling map analysis. This analysis, as shown in 7, is based on the number of citation counted within the set of selected articles, mapping the currently most important sources for research considering only the last three years. In overall, 15 clusters can be observed.

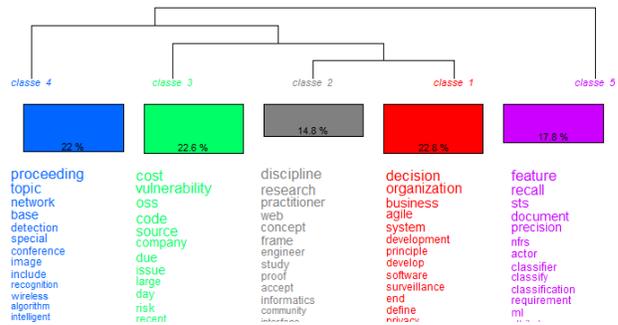


**Figure 7. Bibliographic coupling of authors density map. Source: Scopus. Map generated by VOSViewer.**

The most prominent cluster represented by Cysneros (2018) and Backer (2019), in general, presenting approaches for classification and identification of security requirements. This reveals the great concern of the research in privacy threats and security risks, eventually related to applications such autonomous

driving, linguistic analysis. This cluster is followed by Di martino's (2018) cloud services research and Wang (2019), each of them with 4 accumulated citations.

To perform the integrating model we conducted an analysis based on the Descending Hierarchic Classification that proposes to identify main classes on requirements in ethics in AI research. This analysis examined 182 abstracts and found 1019 text segments. The text segments were organized into five classes: Class 1 with 22.8%, Class 2 with 14.8%, Class 3 with 22.6%, Class 4 with 22% and Class 5 with 17.8%, as shown in Figure 8. In Class 1 the



**Figure 8. Descending hierarchic classification dendrogram. Generated with the use of Iramuteq.**

most representative work is Intelligent Software Mining with Business Intelligence Tools for Automation of Micro services in SOA: A Use Case for Analytics [27], where authors explored the automated process of mining software engineering data for useful business applications. Analysing the words that represent the class such as Decision, Organization, Business, Agile, System, Development and Principle, it is possible to note that they point to the use of AI-based systems as a decisive factor in organizations. Correspondingly the class is called AI in Agile and Business for RE.

In Class 2 the most representative work is Requirements We Live By [28] where the author called researchers attention for reflections upon RE as a discipline in light of new technologies as AI. Analysing the words that represent the class such as Discipline, Research, Practitioner, Web, Concept, Frame and Engineer, it can be seen that they all point towards RE research as a discipline and a concept, along with practitioners and engineers in this field with a theoretical approach. Accordingly the class is called RE in Theoretical Context. In Class 3 the most representative works are: 1) Enhancing Offshore Safety Culture Through Continuous Management of Barriers and Success Paths [29], in this study decision support systems are explored to assess safety culture and control room management in the context of offshore

operations. 2) The correlation between OSS project and organizational performance [30] in this research authors proposed policy directions to improve awareness of Open Source Software in their company. Analysing the words that represent the class such as Cost, Vulnerability, OSS, Code, Source and Company, it can be seen that they all point towards quality factors as cost, vulnerability, of source code, in the context of companies using OSS or not. Accordingly the class is called Quality Requirements.

In Class 4, the most representative work is actually an abstract of a Proceeding focused on the research of sensors, and, analysing the words that represent the class such as Proceeding, Topic, Network, Base, Detection, Special, and Conference, it can be seen that they all point towards Proceedings and Conferences without much connection to RE and AI, mainly due to the fact that no area of knowledge was filtered. Accordingly the class is called Proceedings and Conferences.

In Class 5, the most representative works are: 1) Hidden in plain sight: Automatically identifying security requirements from natural language artifacts [31] where the authors used ML techniques to develop a tool-assisted process taking as input a set of natural language artifacts to aid requirements engineers in producing a more comprehensive and classified set of security requirements. 2) Extracting Quality Attributes from User Stories for Early Architecture Decision Making [32] aimed to automatically identify quality attributes from user stories (functional user requirements). 3) Automatically Classifying Functional and Non-functional Requirements Using Supervised Machine Learning [33], in this study authors used supervised ML to automatically classify requirements as functional and non-functional. Analysing the words that represent the class such as Feature, Recall, STS, Document, Precision, NFRS and Requirement, as well as the most representative works, it is observed that they converge to automatically classify requirements by using AI techniques. Thus, the class is called AI in Requirements Engineering.

#### 4.1. GitHub

In order to discover a satisfactory string to be used in the query for GitHub repositories, a preliminary research was conducted. The number of results for each search pattern, are presented in Table 1.

With the use of the string “*ethical artificial intelligence*” with quotation marks produced a set of 19 repositories only about curated lists of courses, books, video lectures and papers about AI. While using the string *ethic artificial intelligence* without

**Table 1. Number of repositories found with different search strings.**

String	Number of repositories
“ethical artificial intelligence”	19
ethic artificial intelligence	36
ethical artificial intelligence	461
artificial intelligence ethics	501
ethics in ai	589

quotation marks resulted in 36 repositories, it is too small for proper analyses. Thus, exploration of GitHub README files was performed using the string *ethics in ai* without quotation marks in the search field of the Corpus Retrieval web based tool, providing us with a set of 589 README files, the largest amount of repositories retrieved between the tested search strings.

Further, a manual reading of the selection of all 589 README files from the corpus was performed. We found out the following filtered categories: 486 (82.5%) Reference lists (e.g., curated lists, lectures and course materials, assignments, conferences materials, software lists), 78 (13.2%) AI applications for end users, 21 (3.5%) tools for implementing AI ethics. Only 4 (0.7%) were not found. Analyzed data is available in <https://zenodo.org/record/4284782>.

From the tools found in GitHub that actually assist ethical AI implementation we highlight. 1) InterpretML: a package that incorporates ML interpretability techniques that explains blackbox systems, thus it is possible to understand the reasons behind individual predictions. 2) Deon: is a command line tool to assist in easily adding ethics checklist to a data science projects and 3) TransparentAI: wraps a mature tool named SHAP to give simple visualization solutions for AI-based systems, in face of TrustworthyAI requirements by the European Union, associated to transparency of model and datasets applied in some project. 4) XAI: designed based on the 8 Principles for Responsive Machine Learning, enables analysis and evaluation of data models, having AI explainability as main player.

To get a qualitative insight from GitHub corpus, we filtered out supplementary forms and unmeaningful active forms (e.g., td, https, href, nbsp, javascript and format types). Following, a word cloud is generated to visualize most frequent words present in the filtered corpus, as shown in Figure 9. From Figure 9 we can observe that most of repositories available are related to learning aspects of ethics in AI (e.g., courses and books), and concerning file repositories (e.g., assignments and courses materials). With the totality of README files found (589), an analysis on Iramuteq is performed



explicability, serving as a tool to show a black box ML algorithm, or other AI technique, as a white box. Revealing, again, that this principle is an enabler for the others, that is, removing it would destroy all the concept of ethical AI. Being able to audit, detect anomalies, understand how things are working behind the curtains is the main aspect involving ethics in AI that should be protected and broadened.

## 5. Final Remarks

There is an increasing discussion in academia and in industry on ethics in AI. Several researchers have an agreement on theoretical principles that guide ethics in AI. However, such principles are not easy to implement, thus, there is a need to translate ethical principles into practice. This study attempted to provide an overview analysis on the topic of ethics in AI both in a scientific database and a repository of open source projects. Although it is seen from our analysis that the scientific community focus on AI methods or tools to assist requirements engineers on requirements analysis, in fact GitHub projects benefits from scientific community, and vice-versa, as published papers commonly open-source their codes, and scholars explore available tools in their researches. It was produced five classes from our scientific database analysis in Scopus, 1) AI in Agile and Business for RE (22.8%), 2) RE in Theoretical Context (14.8%), 3) Quality Requirements (22.6%), 4) Proceedings and Conferences (22%), 5) AI in Requirements Engineering (17.8%).

The requirements found within GitHub through the used method may not reflect only results for AI projects: a considerable number of projects in GitHub are not focused on AI-based systems, many being actually courses, references or such that reference the area. Projects found were not necessarily built with the considerations of those principles, but we classified the requirements according to them because they are already being used in the literature to address ethical requirements in AI as in [9]. It is uttermost that the main focus of regulations should be situated on the work of the software developer, in consonance with Antonov and Kerikmäe [35]. Implement ethics in AI is not an easy task, sustained dedication is needed. However, it is crucial to internalize ethical AI development urgency. Highlighted tools found in GitHub such as InterpretML, Deon, XAI and TransparentAI are some examples of active development projects from the Open Source Community that can enable transparency, a highly recurrent principle from guidelines. The combined energy of both scientific and open-source sources fosters an enhanced debate and stimulates progress towards AI

ethics in practice.

## References

- [1] A. Vogelsang and M. Borg, "Requirements engineering for machine learning: Perspectives from data scientists," in *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, pp. 245–251, IEEE, 2019.
- [2] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "From what to how: An initial review of publicly available ai ethics tools, methods and research to translate principles into practices," *arXiv preprint arXiv:1905.06876*, 2019.
- [3] V. Vakkuri, K.-K. Kemell, J. Kultanen, and P. Abrahamsson, "The current state of industrial practice in artificial intelligence ethics," *IEEE Software*, 2020.
- [4] A. Hagerty and I. Rubinov, "Global ai ethics: A review of the social impacts and ethical implications of artificial intelligence," *arXiv preprint arXiv:1907.07892*, 2019.
- [5] T. N. Y. Times, *Cambridge Analytica and Facebook: The Scandal and the Fallout So far*, Apr. 2018. <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>.
- [6] Reuters, *Amazon scraps secret AI recruiting tool that showed bias against women*, Oct. 2018. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.
- [7] K. Siau and W. Wang, "Artificial intelligence (ai) ethics: Ethics of ai and ethical ai," *Journal of Database Management (JDM)*, vol. 31, no. 2, pp. 74–87, 2020.
- [8] T. Hagendorff, "The ethics of ai ethics—an evaluation of guidelines," *arXiv preprint arXiv:1903.03425*, 2019.
- [9] R. Guizzardi, G. Amaral, G. Guizzardi, and J. Mylopoulos, "Ethical requirements for ai systems," in *Canadian Conference on Artificial Intelligence*, pp. 251–256, Springer, 2020.
- [10] A. Aberkane, "Exploring ethics in requirements engineering," Master's thesis, 2018.
- [11] A. McNamara, J. Smith, and E. Murphy-Hill, "Does acms code of ethics change ethical decision making in software development?," in *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pp. 729–733, 2018.
- [12] I. S. Association, *The IEEE global initiative on ethics of autonomous and intelligent systems*, Apr. 2019. <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>.
- [13] E. C. High-Level Expert Group on Artificial Intelligence, *Ethics guidelines for trustworthy AI*, Apr. 2019. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.
- [14] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.

- [15] L. Floridi, J. Cows, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, *et al.*, “Ai4peoplean ethical framework for a good ai society: Opportunities, risks, principles, and recommendations,” *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018.
- [16] L. Rothenberger, B. Fabian, and E. Arunov, “Relevance of ethical guidelines for artificial intelligence - a survey and evaluation,” in *27th European Conference on Information Systems - Information Systems for a Sharing Society, ECIS 2019, Stockholm and Uppsala, Sweden, June 8-14, 2019* (J. vom Brocke, S. Gregor, and O. Müller, eds.), 2019.
- [17] A. for Computing Machinery, *ACM code of ethics and professional conduct*, Aug. 2018. <https://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-booklet.pdf>.
- [18] V. Vakkuri, K. Kemell, M. Jantunen, and P. Abrahamsson, ““this is just a prototype”: How ethics are ignored in software startup-like environments,” in *Agile Processes in Software Engineering and Extreme Programming - 21st International Conference on Agile Software Development, XP 2020, Copenhagen, Denmark, June 8-12, 2020, Proceedings* (V. Stray, R. Hoda, M. Paasivaara, and P. Kruchten, eds.), vol. 383 of *Lecture Notes in Business Information Processing*, pp. 195–210, Springer, 2020.
- [19] H. Belani, M. Vukovic, and Z. Car, “Requirements engineering challenges in building ai-based complex systems,” in *27th IEEE International Requirements Engineering Conference Workshops, RE 2019 Workshops, Jeju Island, Korea (South), September 23-27, 2019*, pp. 252–255, IEEE, 2019.
- [20] B. Kostova, S. Gurses, and A. Wegmann, “On the interplay between requirements, engineering, and artificial intelligence,” in *Joint Proceedings of REFSQ-2020 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 26th International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2020), Pisa, Italy, March 24, 2020* (M. Sabetzadeh, A. Vogelsang, S. Abualhajja, M. Borg, F. Dalpiaz, M. Daneva, N. Condori-Fernández, X. Franch, D. Fucci, V. Gervasi, E. C. Groen, R. S. S. Guizzardi, A. Herrmann, J. Horkoff, L. Mich, A. Perini, and A. Susi, eds.), vol. 2584 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020.
- [21] V. Vakkuri, K.-K. Kemell, and P. Abrahamsson, “Eccola—a method for implementing ethically aligned ai systems,” *arXiv preprint arXiv:2004.08377*, 2020.
- [22] S. Larsson, M. Anneroth, A. Felländer, L. Felländer-Tsai, F. Heintz, and R. C. Ångström, “Sustainable ai: An inventory of the state of knowledge of ethical, social, and legal challenges related to artificial intelligence,” 2019.
- [23] A. M. Mariano, A. C. B. Reis, L. dos Santos Althoff, and L. B. Barros, “A bibliographic review of software metrics: Applying the consolidated meta-analytic approach,” in *International Joint conference on Industrial Engineering and Operations Management*, (Cham), pp. 243–256, Springer, Springer International Publishing, 2018.
- [24] G. A. A. Prana, C. Treude, F. Thung, T. Atapattu, and D. Lo, “Categorizing the content of github readme files,” *Empirical Software Engineering*, vol. 24, no. 3, pp. 1296–1327, 2019.
- [25] R. L. Q. Portugal, M. A. Casanova, T. Li, and J. C. S. do Prado Leite, “GH4RE: repository recommendation on github for requirements elicitation reuse,” in *Proceedings of the Forum and Doctoral Consortium Papers Presented at the 29th International Conference on Advanced Information Systems Engineering, CAiSE 2017, Essen, Germany, June 12-16, 2017* (X. Franch, J. Ralyté, R. Matulevicius, C. Salinesi, and R. J. Wieringa, eds.), vol. 1848 of *CEUR Workshop Proceedings*, pp. 113–120, CEUR-WS.org, 2017.
- [26] R. L. Q. Portugal, H. Roque, and J. C. S. do Prado Leite, “A corpus builder: Retrieving raw data from github for knowledge reuse in requirements elicitation,” in *Proceedings of the 3rd Annual International Symposium on Information Management and Big Data - SIMBig 2016, Cusco, Peru, September 1-3, 2016* (J. A. Lossio-Ventura and H. Alatrística-Salas, eds.), vol. 1743 of *CEUR Workshop Proceedings*, pp. 48–54, CEUR-WS.org, 2016.
- [27] D. P. Wangoo, “Intelligent software mining with business intelligence tools for automation of micro services in soa: A use case for analytics,” in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 98–101, IEEE, 2020.
- [28] B. Nuseibeh, “Requirements we live by,” in *2019 IEEE 27th International Requirements Engineering Conference (RE)*, pp. 1–1, IEEE, 2019.
- [29] W. R. Nelson, A. I. Ahluwalia, *et al.*, “Enhancing offshore safety culture through continuous management of barriers and success paths,” in *SPE Health, Safety, Security, Environment, & Social Responsibility Conference-North America*, Society of Petroleum Engineers, 2017.
- [30] J.-B. Kim and H. Park, “The correlation between oss project and organizational performance,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 4S2, pp. 72–78, 2019.
- [31] M. Riaz, J. King, J. Slankas, and L. Williams, “Hidden in plain sight: Automatically identifying security requirements from natural language artifacts,” in *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, pp. 183–192, IEEE, 2014.
- [32] F. Gilson, M. Galster, and F. Georis, “Extracting quality attributes from user stories for early architecture decision making,” in *2019 IEEE International Conference on Software Architecture Companion (ICSA-C)*, pp. 129–136, IEEE, 2019.
- [33] Z. Kurtanović and W. Maalej, “Automatically classifying functional and non-functional requirements using supervised machine learning,” in *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pp. 490–495, IEEE, 2017.
- [34] R. L. Q. Portugal and J. C. S. do Prado Leite, “Extracting requirements patterns from software repositories,” in *2016 IEEE 24th International Requirements Engineering Conference Workshops (REW)*, pp. 304–307, IEEE, 2016.
- [35] A. Antonov and T. Kerikmäe, “Trustworthy ai as a future driver for competitiveness and social change in the eu,” in *The EU in the 21st Century*, pp. 135–154, Springer, 2020.