

Article

Not peer-reviewed version

---

# Towards an ELSA Curriculum for Data Scientists

---

[Maria Christoforaki](#) \* and [Oya Deniz Beyan](#)

Posted Date: 8 January 2024

doi: 10.20944/preprints202401.0622.v1

Keywords: AI ethics; Data Science; Artificial Intelligence; ELSA; Education



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Towards an ELSA Curriculum for Data Scientists

Maria Christoforaki <sup>1,\*</sup> and Oya Beyan <sup>1,2</sup>

<sup>1</sup> University of Cologne, Faculty of Medicine and University Hospital Cologne, Institute for Biomedical Informatics; maria.christoforaki@uk-koeln.de

<sup>2</sup> Fraunhofer Institute for Applied Informatics, St. Augustin, Germany; oya.beyan@uni-koeln.de

\* Correspondence: maria.christoforaki@uk-koeln.de

**Abstract:** The use of Artificial Intelligence (AI) applications in a growing number of domains in the latest years has put into focus Ethical Legal and Societal Aspects (ELSA) of these technologies and the relevant challenges they pose. In this paper we propose an ELSA Curriculum for Data Scientists aiming to raise awareness about ELSA challenges in their work, provide them with a common language with the relevant domain experts in order to cooperate to find appropriate solutions, and finally, incorporate ELSA in the data science workflow and not be seen as an impediment or a superfluous artefact, rather than an integral part of the Data Science Project Lifecycle. The proposed curriculum uses the CRISP-DM model as a backbone to define a vertical partition, expressed in modules corresponding to the CRISP-DM phases. The horizontal partition includes knowledge units belonging to three strands that run through the phases, namely Ethical and Societal, Legal and Technical Rendering Knowledge Units (KUs). In addition to the detailed description of the aforementioned KUs, we also discuss the implementation, issues such as duration, form and evaluation of participants, as well as the variance of the knowledge level and needs of the target audience.

**Keywords:** AI ethics; data science; Artificial Intelligence; ELSA; education

---

## 1. Introduction

The ubiquity of Data Science applications in the latest years has put into the spotlight a multitude of ethical, legal, and societal aspects (ELSA) regarding data driven methods. While computer ethics is not a new subject (see for example [1]), the emergence of Big Data, Machine Learning (ML) and more recently Deep Learning (DL) applications prompted the issue of recommendations and guidelines by professional, national, international and supranational organisations and institutions (for guideline reviews see [2,3], while an inventory can be found in [4]).

Part of the guidelines' suggestions refers to ELSA education for AI practitioners. Specifically, the European Commission's high-level expert group on artificial intelligence, in their ethics guidelines for trustworthy AI [5] highlight the importance of education and awareness raising both to promote an ethical mindset for all stakeholders, including developers (pp. 23-24), and as well as a means to develop accountability practices (p. 31).

In this paper we present our suggestion for an ELSA Curriculum for Data Scientists. First, we paint with broad strokes the profile of the data scientist as us it is presented in the bibliography and emerges in surveys. Using the same sources, we also report about any ELSA training they might have gotten from their studies, the industry attitude towards ELSA, as well as their own opinions on the importance of relevant issues. After that, we present our Curriculum proposal, and we discuss the limitations and challenges for the Curriculum implementation. Finally, we close with the conclusions section.

## 2. Background

### 2.1. The profile of the data scientist

According to the people who claimed to have coined the term, a data scientist is: "a high-ranking professional with the training and curiosity to make discoveries in the world of big data." [6]. In this description of the data scientist and their work, no mention to ethical or legal aspects is made. In [7], on the other hand, the authors describe data science as the intersection of different subdisciplines, which apart from the technical and scientific domains such as, statistics, machine learning and system design, includes behavioral and social sciences for ethics and understanding human behavior. [8] adds to the technical skills stakeholder involvement.

To gain an insight on the opinions of data scientists themselves and their attitudes about ELSA challenges of their work, we turned at the surveys conducted by Anaconda Python distributor [9] and Kaggle [10], the Google subsidiary and an online community of data scientists and machine learning practitioners. These surveys are conducted yearly and mostly contain questions about technical issues. Participants (individuals and organizations) come from all over the world on a voluntary basis. Undoubtedly the sample is not rigorously defined, however, it gives us an insight on the profile of the practitioners. Looking at the profile of the data scientist as is self-reported by practitioners from all over the world, in Anaconda [11–14] and Kaggle surveys [15] we focus on two issues: the education lever of data scientists and their attitude towards ELSA challenges.

Both of the surveys describe the majority of data scientists as having some tertiary education degree (Master's being the most common). Notably, in the Anaconda surveys, the number of responders having degrees has risen through the years (2020-2021), while in the Kaggle summary of the years 2017-2021, the opposite is true. The survey attributes this to the "increasing availability of online courses and ways to build technical skills and experience, having a degree isn't a prerequisite for getting started in data science" [15].

The difference in the findings may be attributed to the variance of the samples between the two surveys or to the fact that the time period covered is different: Anaconda surveys covers the years 2020-2023, while the Kaggle one the years 2017-2021 (the Kaggle 2022 survey results regarding the education level of the participants were yet not available at the time of writing). One might hypothesize that in the early years of 2017-2021, there were not many graduates available to work as data scientists, while that changed in the following years. In fact, according to the World Economic Forum's 2020 Future of Jobs report [16] most of the people transitioning to AI and Data Science occupations originate from a different job family (job Families are groups of occupations based upon work performed, skills, education, training, and credentials); this may be due to the fact that data science as a profession is not bound by professional associations, certifications, etc., and is also in high demand (both the 2020 and 2023 [17] Future of Jobs reports show high demand for data science practitioners).

#### *Does the industry demand ELSA skills?*

In [18] industry executives who have held leadership roles in IT business operations (e.g., CIO, CTO, and digital innovation manager), include anonymity, privacy, and ethics in core areas in which a data scientist must be knowledgeable about. Furthermore, in a 2018 ACM members survey [19], one third (34%) of the industry representatives require experience in ethics from their prospective employees, while half of them (51%) responded that this is elective.

Nevertheless, in World Economic Forum's 2023 Future of Jobs report [17], when Information and technology services organizations were asked to rank ethics as a core skill for their workers, only 12% did so for environmental stewardship and 7% for global citizenship. However, they see it as increasing in significance for the near future (2023-2027).

The latter is also reflected in the Anaconda surveys through the years: in the 2021 survey a third of the organizations replied that they have no plan to take any steps towards ensuring fairness and mitigate bias or to address model explainability and an equal percentage that they do know whether there is a plan or not [12]; in 2022, 24% said that they do not employ any measures regarding the

same issues -the “not sure” reply was around 15% [13]; in 2023 it is reported that one of the roles that companies hire or planned to hire is AI ethicist [14].

## 2.2. What do the universities teach?

Since the majority of data scientists have a tertiary degree, it is useful to examine whether they get any ELSA education during their studies.

According to the 2018 ACM members survey mentioned above [19], regarding the Computing Competencies for Undergraduate Data Science Curricula, only half (54%) of Data Science programs required ethics courses, while less of a quarter of the programs actually offered such a course. In a study conducted in Europe in 2020 about the teaching of ethics in computer science or related programs, 36% of the respondents report that their institutions did not teach ethics. However, the ones that did teach them, they did so in AI and related areas (64%), and machine learning and related topics (51%) [20].

However, in the Anaconda state of data science data science surveys, only around 20% of the students self-report that that ethics in data science/ML is covered in their courses, while a slightly higher amount (22-24%) said that bias in AI/ML/data science is taught frequently in classes or lectures [12,13].

## 2.3. What the data scientists say about ELSA challenges

Regarding the attitudes of the data scientists about ELSA issues, in the Anaconda reports they state as the biggest issue to be tackled in AI/ML the social impacts that stem from bias in data and models, being followed by impacts to individual privacy [11–13]; however, when asked about skill gaps, only 15% identified that Ethics in data science as lacking from data practitioners at their organization [12].

The 2023 survey focused specifically on the use of generative AI and the majority of participants (39%) cited the need for transparency and explainability in AI models, followed by bias and fairness in AI algorithms (27%), and balancing copyright and IP protections with innovation (15%) [14].

# 3. Curriculum Proposal

## 3.1. General Objectives and Vision

Taking into consideration the profile and the ELSA challenges that the data scientists face on their work, we propose a n ELSA Curriculum which has three basic objectives. Specifically, data scientists should be able to:

1. recognize Ethical Legal and Societal Aspects pertaining to their work (Awareness);
2. possess a common language with the relevant domain experts in order to cooperate to find appropriate solutions (Communication ability);
3. incorporate ELSA in the data science workflow and not be seen as an impediment or a superfluous artefact, rather than an integral part of the Data Science Project Lifecycle (Professional mentality building).

## 3.2. Curriculum structure and Overview

The objectives stated in the above section guided us in the defining the curriculum structure, which has a vertical (modules) and a horizontal (strands) partition, within which the various Knowledge Units are situated. A *Knowledge Unit (KU)* is the elemental component of the Curriculum, defined as a concise piece of knowledge pertaining to a specific ELSA subject.

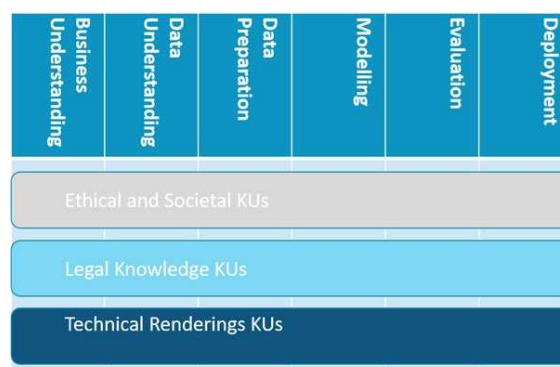
The horizontal partition is defined by the first two objectives (awareness and communication ability). In order to achieve them, data scientists should learn and be able to understand basic concepts belonging to other disciplines (law, ethics, and social sciences), as well as being aware of how the ELSA issues could be practically tackled via the use of specific frameworks, standards and techniques. Thus, we defined three main area topics, namely:

- **Ethical and Societal Knowledge Units:** Ethical and Societal aspects of data science range from incorporating community values to one's work, averting discrimination against individuals or groups, taking into consideration the environmental impact of the data science applications, and assuming responsibility and being accountable for one's decisions and actions.
- **Legal Knowledge Units:** The objective of these units is to help make data scientists cope with legal issues they might be facing in their course of work, mainly data protection and intellectual property issues, as well as basic legal terminology and concepts.
- **Technical Renderings Knowledge Units:** These units deal with technical renderings of legal or ethical desiderata like privacy, data and algorithmic bias detection and mitigation strategies, incorporation of fairness, effectiveness and explainability in the evaluation of a data science project, deployment and monitoring outside experimental/testing environments. They illustrate the way technical solutions can be employed via use cases and not aim to teach technical skills to data scientists, since: a) these might be already dealt with in specific courses during their studies or their working experience and b) vary in each application domain that might require specific techniques.

The vertical partition (modules), seeks to achieve the third objective, the seamless integration of ELSA handling in the data science workflow. For this purpose, we employ the CRISP-DM (Cross-Industry Standard Process for Data Mining), a non-proprietary, documented, and freely available industry-, tool-, and application-neutral model data mining model offered as a best practices guide. CRISP-DM comprises six phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment, which provide a road map to follow while planning and carrying out a data mining project [21].

The reasons that we chose the specific model are that: a) it is well known and often used in data science projects; b) It has been developed mainly by industry, so suited to the needs of the data science practitioners; and c) it is already used to present ethical and legal issues in the different phases of the model and to develop frameworks that ensure the application of ethical standards in the development of data science projects [22–24].

Figure 1. depicts the horizontal-vertical structure of the Curriculum. With grey we denote the ethical/societal strand, with light blue the legal strand, with and deep blue the technical renderings of the ELSA challenges.



**Figure 1.** The horizontal-vertical structure of the Curriculum. With grey we denote the ethical/societal strand, with light blue the legal strand, with and deep blue the technical renderings of the ELSA challenges. The vertical one is expressed as modules that correspond to the CRISP-DM phases, while the horizontal one comprises knowledge units belonging to three strands that run through phases.

### 3.3. ELSA Curriculum KUs

In this section we are going to present a more detailed, albeit not exhaustive description of the KUs. For each module (corresponding to each CRISP-DM phase as described in [21]), we make a summary description of the KUs assigned to it.

The interested reader can find the complete description in the **Supplementary Materials**.

### Module I: Business understanding

This phase concerns understanding the project objectives from a business perspective and converting the problem to a data science problem, and devising the preliminary plan to achieve these objectives.

The corresponding KUs involve stakeholder identification; incorporating community values, such as professional codes of ethics and national and international guidelines, as well as “-by design” guidelines and frameworks for Ethics, Privacy, Data Protection, etc; organisational culture meaning an organisation’s systems, procedures, and practices for guiding and supporting ethical behaviour; basic legal concepts, such as domains of law and specifically of cyberlaw, law stratification (e.g. national vs international and EU law, and how these relate), as well as concepts regarding accountability measures such as audit, impact assessment, compliance, risk assurance.

### Module II: Data understanding

This phase concerns data collection, exploration and quality verification, gaining insights and identifying possible issues.

This module addresses many issues common in the following one (Data Preparation), and in a specific curriculum implementation the two modules can be integrated (this approach is also followed in the classification of ethics issues (collectively name *data challenges*) by [22], which had a major influence in our Curriculum structure). However, we create two different modules for the following reasons: a) there are a number of challenges that are to be faced before any pre-processing takes place and that has to be underlined in the curriculum; b) this phase can alter the previous one of business understanding, since data understanding can alter the initial problem definition; c) it might be the case that phases 2 and 3 are not performed by the same people, as data scientists may be provided by already existing datasets and are not directly connected with dataset creation.

That said, depending on the implementation of the Curriculum, some of the KUs that are dealing with common issues can be addressed in either or both Module, as time and purpose constraints vary. For more details, see Section 4. for the Implementation strategies.

The KU’s in this Module deal with Data Protection and Intellectual Property issues (basic concepts and issues that have to do with data collection); ways of detecting and mitigating data bias in this phase; the use of synthetic data; ethics dumping; and methods of dataset documentation.

### Module III: Data preparation

This phase aims to the construction of the final dataset that will be used by the model, a process that includes selection, cleaning, quality checking, integration and formatting of the collected data.

While the main subjects are same as in the previous model (the KUs bear the same names), we try to differentiate between the relevant issues. That said, depending on the implementation, the respective KUs can be taught independently or not (e.g. in order to avoid repetitions).

The KUs belonging to this module also deal with data protection (here with the application of a specific standard as a framework for a case study), and intellectual property focusing on the different kinds of training data according to their licences when specifically considering generative AI issues. Also, data challenges in pre-processing like annotation, cleaning, creating synthetic data, to choosing features and proxies. such as various kinds of bias (representation, measurement and aggregation bias), or even data validity issues (for example, a gap between the way annotation results between crowdsourcing and scholar annotation, error induced by automatic machine learning annotation, in the creation of synthetic data depending on the problem). Societal issues like employing low cost solution for data annotation (e.g. by outsourcing to countries in the Global South) are also addressed here. Dataset documentation is a continuation of the respective KU of the previous module.

### Module IV: Modelling

This phase concerns the selection and parameter calibration of models according to the specific problem definition and specific data requirements.

KUs in this module address Model bias and possible mitigation techniques, transparency and explainability, environmental impact of model training, Intellectual property issues that pertain to the algorithm itself and its products (the use of pretrained models, commercial secrets, property of the trained model and the outcome of the algorithm) and model documentation frameworks.

#### Module V: Evaluation

This process includes the review of the model construction and evaluation of whether it achieves the objectives set in the business understanding phase.

The KUs here deal with model evaluation beyond accuracy, i.e., taking also into consideration also factors like fairness, efficiency (in terms of resource allocation), explainability, trustworthiness (robustness outside the experimental settings), and whether the selected features and proxies actually solve the initial problem, although they provide accurate results. Specific focus is given to fairness, since the evaluation phase gives us a first tangible result we can use to assess whether there are bias and discrimination issues. Additionally, we cover here the various kinds of fairness, legal provisions regarding it, as well as the human perceptions of it and how these impacts the trustworthiness of a product, and insights from other disciplines (e.g., ethical philosophy) that go beyond technical solutions. Finally, there is a KU regarding model documentation as a continuation of the previous module KU with the same name.

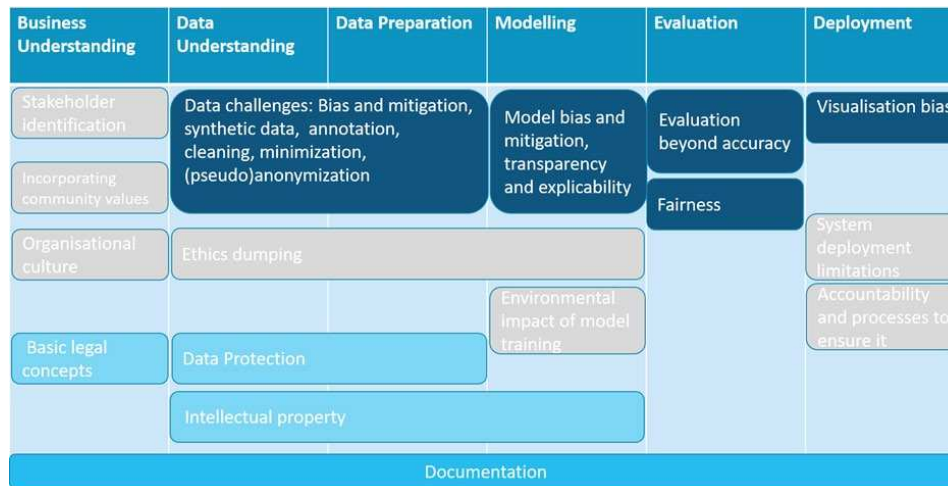
#### Module VI: Deployment

This phase includes deployment, monitoring and maintenance of the system. Even though it is often the customer who carries out the deployment steps, it is important for the customer to understand how they actually use the system and its limitations. As [22] point out, the data scientist's ethical responsibilities do not end with the completion of a project. The data scientist also has a duty to explain their choices and the implications, using language that non-data scientists, such as managers, can understand.

So, there is a KU regarding the System deployment limitations both with respect to known issues and with what the system is actually able to do or not, and the chosen level of automation and the possible impact that both might have, especially in the case of adverse outcomes. There is also a special KU on Visualisation bias and the well-known pitfalls in that area; data scientists are not especially familiar with them as they are usually thought of as UX issues, and this KU aims to help them in their cooperation with UX designers in order to avoid as much as possible misinterpretation of results. Finally, there is a KU on Accountability and processes to ensure it, being a mirror image of some of the subjects dealt in Module I: Business understanding, such as auditing frameworks and organisation culture and processes, the extent of personal responsibility and the case for certification for data scientists in conjunction with professional codes and the creation of a formal profession.

Figure 2. Shows a more detailed, albeit not exhaustive, presentation of the Curriculum content. Each of these blobs may correspond to more than one knowledge unit; some of the topics addressed in each module can be seen in the figure. To the original three strands we added a documentation knowledge unit that runs through all modules, emphasising the need for documenting each action taken during the various phases of the project. The way of doing this depends on the phase; we do not advocate for a specific documentation or auditing system.

A detailed presentation of the Curriculum. Not all KUs are represented, rather the main subjects that are addressed. As in Figure 1., with grey we denote the ethical/societal strand, with light blue the legal strand, with and deep blue the technical renderings of the ELSA challenges



**Figure 2.A** detailed presentation of the Curriculum. Not all KUs are represented, rather the main subjects that are addressed. As in Figure 1., with grey we denote the ethical/societal strand, with light blue the legal strand, with and deep blue the technical renderings of the ELSA challenges.

#### 4. Discussion - Implementation Strategies

Regarding the Curriculum implementation, we encounter the issues of program duration and form. Specifically, how much time could be dedicated for each module and whether this could be done either for a longer period of time, (for some hours a week or for example), or in an intensive one-week summer school, as a series of workshops, etc.

This is also connected with the target audience level of knowledge and needs. The curriculum is mainly aimed at entry level data scientists, giving them a bird's eye view, so to speak, of some the ELSA challenges they might encounter during their work. In this case, the program might fill in the knowledge gap of the missing university courses.

However, when sketching the data scientist profile in Section 2., we identified two kinds of data scientists:

- one with background in mathematics, statistics and computer science and programming skills, acquired either by working experience and or by relevant studies;
- a second one, comprising people from other disciplines, who have transitioned to data science. These may come from a different background (for example, social sciences, economics, etc), and have a different set of skills, such as: be more versed in raising appropriate hypotheses regarding a business problem; soft skills associated with communication and teamwork; specialized domain knowledge; or, business and strategy competences.

We also have to take into consideration the fact that future, data scientists do not form a homogeneous group, such as being university graduates, which is the target group of today's ethics curricula. They also come from a variety of educational, social and national backgrounds, and quite a lot of them come in data science following quasi-academic routes (like on-line courses), especially when being employed in job positions with less demanding high-end knowledge or education. We might as well expect their number to arise, if we take into consideration the use of generative AI to automate computer programming tasks and the high demand in the industry.

Thus, there is another way that the curriculum can be implemented: with the objective to fit the job roles of the participants, the application domain special demands, or both.

In the first case, this may be accomplished by expanding specific KUs (or contracting others) so that they are aimed to, for example, project managers or developers. In the second one, the content of the KUs can be adapted to fit specific application domains, for example NLP, image processing, or health care.

## 5. Conclusions

In this paper we presented a proposed ELSA Curriculum for Data Scientists. We tried to draw the profile of the data scientist, starting with the ideal form, as provided in the bibliography as a multidisciplinary and multi-talented individual, and augmenting it with the results of empirical studies and surveys, as a person coming from a variety of educational and professional backgrounds. This profile reveals that, while ELSA issues are considered either by organizations issuing guidelines, the industry and by the data scientists themselves as very important, there is a lack of knowledge about the respective challenges in the practitioners.

The proposed Curriculum comprises knowledge units from three domains (ethical legal and technical) that belong in modules which themselves correspond to the six phases of the well-known CRISP-DM Model. The objectives of the Curriculum are, to raise awareness about ELSA challenges in data science applications; enhance the communication ability of the data scientist by providing a common language with domain experts (for example legal scholars) that they will have to cooperate in order to successfully tackle the above mentioned challenges; and finally, to foster a professional mentality that treats ELSA issues as an integral part of the Data Science Project Lifecycle by embedding them into the Data Science Project Lifecycle.

We recognize that the implementation of such a Curriculum requires a considerable amount of resources, both regarding the duration of a training program and the its multidisciplinary nature. We propose a flexible implementation strategy, that expands, contracts or conflates KUs in the various Modules, as to fit the roles and the specific application domains of the participants. Also, the depth of the subjects treated can vary: for novices, a more holistic but less in-depth program is proposed; for more experienced practitioners, it might be better to focus on specific areas of interest.

We have published the first version of the Curriculum (see **Supplementary Materials**) and we are welcoming comments and suggestions from the community. For this purpose, we have already contacted a series of workshops and currently we request feedback via a specifically designed survey. This will lead as to a second version of the Curriculum. Naturally, the best way to assess the curriculum, will be to actually implement it and evaluate it via the experiences of both the instructors and the participants.

**Supplementary Materials:** The following material can be downloaded from the FAIR Data Spaces Zenodo community (<https://zenodo.org/communities/fair-ds>): 'ELSA Training Curriculum for Data Scientists - Version 1.0'. FAIR-DS Project Deliverable. Cologne: UzK, 31 August 2023. <https://zenodo.org/record/8318726>; 'ELSA Training for Data Scientists-Describing the Landscape'. FAIR-DS Project Deliverable. Cologne: UzK, 31 December 2021. <https://doi.org/10.5281/zenodo.7233569>; presentation slides and video recordings of the ELSA Workshops conducted in 2022 and 2023. —

**Author Contributions:** Conceptualization, Maria Christoforaki and Oya Beyan; investigation, Maria Christoforaki; writing—original draft preparation, Maria Christoforaki; writing—review and editing, Maria Christoforaki and Oya Beyan; supervision, Oya Beyan. All authors have read and agreed to the published version of the manuscript

**Funding:** his research was partially funded by the German Federal Ministry of Education and Research (BMBF) FAIR Data Spaces project by grant number FAIRDS15.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest

## References

1. Moor, J.H. What Is Computer Ethics? *Metaphilosophy* **1985**, *16*, 266–275.
2. Jobin, A.; Ienca, M.; Vayena, E. The Global Landscape of AI Ethics Guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399.
3. Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; Srikumar, M. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*; Berkman Klein Center for Internet & Society: Rochester, NY, 2020;
4. AI Ethics Guidelines Global Inventory by AlgorithmWatch Available online: <https://inventory.algorithmwatch.org> (accessed on 10 February 2021).
5. High-Level Expert Group on AI (AI HLEG) *Ethics Guidelines for Trustworthy AI*; European Commission: Brussels, 2019;
6. Davenport, T.H.; Patil, D.J. Data Scientist: The Sexiest Job of the 21st Century. *Harv. Bus. Rev.* **2012**, *90*, 70–76.
7. van der Aalst, W.M.P. Data Scientist: The Engineer of the Future. In *Enterprise Interoperability VI*; Mertins, K., Bénaben, F., Poler, R., Bourrières, J.-P., Eds.; Springer International Publishing: Cham, 2014; pp. 13–26 ISBN 978-3-319-04947-2.
8. Luna-Reyes, L.F. The Search for the Data Scientist: Creating Value from Data. *ACM SIGCAS Comput. Soc.* **2018**, *47*, 12–16, doi:10.1145/3243141.3243145.
9. About Anaconda Available online: <https://www.anaconda.com/about-us> (accessed on 14 December 2023).
10. Kaggle: Your Machine Learning and Data Science Community Available online: <https://www.kaggle.com/> (accessed on 14 December 2023).
11. Anaconda 2020 State of Data Science Report; 2020;
12. Anaconda 2021 State of Data Science Report; 2021;
13. Anaconda 2022 State of Data Science Report; 2022;
14. Anaconda 2023 State of Data Science Report; 2023;
15. Kaggle Kaggle's State of Machine Learning and Data Science 2021; 2021;
16. Zahidi, S.; Ratcheva, V.; Hingel, G.; Brown, S. *The Future of Jobs Report 2020*; 2020;
17. Di Battista, A.; Grayling, S.; Hasselaar, E. *Future of Jobs Report 2023*; World Economic Forum, Geneva, Switzerland: Geneva, 2023;
18. Mikalef, P.; Giannakos, M.; Pappas, I.; Krogstie, J. The Human Side of Big Data: Understanding the Skills of the Data Scientist in Education and Industry. *503-512* **2018**, doi:10.1109/EDUCON.2018.8363273.
19. Danyluk, A.; Paul Leidig *Computing Competencies for Undergraduate Data Science Curricula-ACM Data Science Task Force*; ACM, 2021;
20. Stavarakakis, I.; Gordon, D.; Tierney, B.; Becevel, A.; Murphy, E.; Dodig-Crnkovic, G.; Dobrin, R.; Schiaffonati, V.; Pereira, C.; Tikhonenko, S.; et al. The Teaching of Computer Ethics on Computer Science and Related Degree Programmes. a European Survey. *Int. J. Ethics Educ.* **2021**, doi:10.1007/s40889-021-00135-1.
21. Shearer, C. The CRISP-DM Model: The New Blueprint for Data Mining. *J. Data Warehous.* **2000**, *5*, 13–22.
22. Saltz, J.S.; Dewar, N. Data Science Ethical Considerations: A Systematic Literature Review and Proposed Project Framework. *Ethics Inf. Technol.* **2019**, *21*, 197–208, doi:10.1007/s10676-019-09502-5.
23. Rochel, J.; Evéquo, F. Getting into the Engine Room: A Blueprint to Investigate the Shadowy Steps of AI Ethics. *AI Soc.* **2020**, doi:10.1007/s00146-020-01069-w.
24. Morley, J.; Floridi, L.; Kinsey, L.; Elhalal, A. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Sci. Eng. Ethics* **2020**, *26*, 2141–2168, doi:10.1007/s11948-019-00165-5.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.