FAIR Data Spaces

ELSA Training for Data Scientists

Describing the Landscape

Maria Christoforaki

31.12.2021

Funded by the Federal Ministry of Education and Research (BMBF)

SPONSORED BY THE
Federal Ministry
of Education
and Research

The FAIR Data Spaces project is funded by the Federal Ministry of Education and Research (BMBF) from May 2021 to May 2024. In this project, the federated and secure data infrastructure Gaia-X and the German National Research Data Infrastructure (NFDI) are connected to a common, cloud-based data space for industry and research in compliance with the FAIR Principles, i.e., to share data in a findable, accessible, interoperable, reusable way.

The project will create a roadmap for the collaboration of both initiatives, clarify the ethical and legal framework for data exchange between research and industry, establish a common technical foundation and demonstrate the feasibility of cross-domain data sharing. By building on and linking services and infrastructures developed both on Gaia-X and NFDI side, a data exchange along the FAIR principles will be demonstrated in concrete applications.

This report is one of the deliverables for the FAIR Data Spaces project, specifically for Work Package “ELSA training for Data Scientists”, which aims to promote knowledge of ethical and legal frameworks and raise awareness of ethical issues among data experts through training and finally prepare an ELSA training curriculum.

The report contains a review of existing policies, guidelines and codes of conduct, and existing ELSA ethics training curricula.

Inhalt

Zusammenfassung

Der Einsatz von Künstlicher Intelligenz (KI) in immer mehr Anwendungsbereichen hat in den letzten Jahren eine Vielzahl ethischer Fragen aufgeworfen, die von der Diskriminierung gefährdeter Bevölkerungsgruppen über die Verletzung der Privatsphäre bis hin zu den Auswirkungen auf die Umwelt reichen und die ethisch-rechtlichen und gesellschaftlichen Aspekte (Ethical Legal and Societal Aspects - ELSA) der Anwendung dieser Technologien in den Mittelpunkt stellen.

Dies führte zu einem Regulierungsbedarf, der von professionellen, nationalen und internationalen Organisationen und Institutionen, zuletzt von der Europäischen Union, aufgegriffen wurde und zu einer Vielzahl von Ansätzen führte, die von Gesetzgebung bis zu Empfehlungen und Leitlinien reichen.
Ein Teil dieser Empfehlungen bezieht sich auf die Ausbildung von KI-Praktikern, ein Thema, das auch in der wissenschaftlichen Literatur weit verbreitet ist. Die ELSA-Ausbildung ist ein Mittel zur Rechenschaftslegung sowohl für Organisationen als auch für Einzelpersonen, die KI-Projekte durchführen, die Dienstleistungen für die Gemeinschaft erbringen, und reicht von der Zertifizierung von KI-Praktikern, ähnlich wie bei Ärzten, Bauingenieuren oder Anwälten, bis zur Teilnahme an Ethikmodulen und -kursen während des Studiums.

Der Bedarf an ELSA-Ausbildung wird auch durch eine kürzlich durchgeführte Umfrage (veröffentlicht 2021) unter Mitgliedern der Association for Computing Machinery (ACM) unterstrichen, die ergab, dass nur 54 % der Data-Science-Programme irgendeine Art von Ethik-Inhalten verlangen. Darüber hinaus antworteten 34 % der Branchenvertreter, dass sie bei der Qualifikation ihrer künftigen Mitarbeitenden Erfahrungen im Bereich Ethik voraussetzen, während 51 % antworteten, dass dies ein Wahlfach sei.
Zwei weitere Punkte müssen berücksichtigt werden:

- Während die KI-Gemeinschaft heute ziemlich homogen erscheint (in Bezug auf Ausbildung und Einkommensniveau), könnte dies in Zukunft nicht mehr der Fall sein, da die hohe Nachfrage nicht nur von hoch ausgebildeten und gut bezahlten Praktikern gedeckt wird, sondern auch von Personen, die quasi-akademische Wege einschlagen und sich an Stellen wenden, die weniger anspruchsvolles Wissen/Ausbildung erfordern.
- Es besteht die Tendenz, zwischen Informatik- und Computerwissenschaftsstudierende zu unterscheiden, was zu zwei Arten von Data Scientists führt, nämlich technischen Experten und solchen, die zwar über einige Kenntnisse in Data-Science-Techniken verfügen, sich aber mehr an den

Bedürfnissen der Nutzer orientieren und eine Brücke zwischen geschäftlichen und technischen Teams schlagen.

Daher wird die ELSA-Ausbildung befürwortet:

a) KI-FAchleute sollten als Teil ihrer formalen Ausbildung in die Lehrpläne der Universitäten aufgenommen werden;
b) KI-Experten, die parallel zu Studenten in Nicht-Informatik-Disziplinen ein Grundverständnis für KI-Techniken erwerben;
c) die allgemeine Bevölkerung bereits in der Schule im Rahmen eines Kurses zur digitalen Kompetenz;
d) alle oben genannten Fälle.

Das Projekt „FAIR Data Spaces“ strebt den Aufbau eines gemeinsamen Cloud-basierten Datenraums für Wirtschaft und Wissenschaft durch Zusammenführung der beiden Initiativen Gaia-X und NFDI an. Während FAIR für auffindbar, zugreifbar, interoperabel und wiederverwendbar (findable, accessible, interoperable, reusable) Daten steht, umfasst das Projekt ein umfangreiches Arbeitspaket (WP2), das sich mit ethischen und rechtlichen Aspekten der Datenverarbeitung befasst. Darüber hinaus umfasst es die Entwicklung einer Reihe von Demonstratoren mit spezifischen Datenverarbeitungs- und -austauschszenarien in WP4. Ein Teil des letztgenannten Arbeitspakets ist die Aufgabe 4.5, die sich mit der Entwicklung eines Curriculums für die Ausbildung von Data Scientists zu ethischen, rechtlichen und gesellschaftlichen Aspekten (ELSA) befasst, dessen erstes Ergebnis das vorliegende Bericht ist, das sich mit der Beschreibung der aktuellen Landschaft in diesem Bereich befasst.

Dieser Bericht umfasst die Überprüfung:

- die aktuellen Ethikrichtlinien für KI, die von verschiedenen Organisationen herausgegeben werden und uns den Kontext bzw. die Anforderungen an die ELSA-Ausbildung liefern, wie sie von nationalen, internationalen und supranationalen Organisationen sowie Berufsverbänden beschrieben werden;
- die bestehenden Hochschulcurricula, die ELSA-Themen beinhalten und uns Beispiele für erfolgreiche Ansätze liefern, sowie die Identifizierung und Klassifizierung der damit verbundenen Herausforderungen.

Die wichtigsten Punkte zur Überprüfung der jüngsten Übersichten, die die Landschaft der AI-Leitlinien abbilden:

- gemeinsame Grundsätze, Fragen oder Themen zu identifizieren, die sich durch alle Bereiche ziehen: Gerechtigkeit und Fairness, Nicht-Malefizien, Sicherheit, Privatsphäre, Freiheit und Autonomie, Vertrauen, Würde, Solidarität, Menschlichkeit, Zusammenarbeit, Teilen, Rechenschaftspflicht, Verantwortung,

Überprüfbarkeit, Wohltätigkeit (im Sinne von Nachhaltigkeit, Wohlergehen und Gemeinwohl), Transparenz, Erklärbarkeit.

- Die oben genannten Begriffe sind in den Leitlinien nicht klar oder überhaupt nicht definiert, und wenn sie es sind, haben sie nicht unbedingt dieselbe Bedeutung.
- Die Leitlinien enthalten zwar viele gemeinsame Werte, lassen aber auch eine Reihe von Aspekten aus: Aspekte des politischen Missbrauchs, der staatlichen Kontrolle, der Fehlinformation und Propaganda, damit zusammenhängende soziale Fragen wie Isolation und Echokammern, die Rechte von KI-Beschäftigten auf niedriger Ebene, die Daten etikettieren oder Inhalte moderieren, die Vielfalt in der KI-Branche sowie die öffentlich-private Partnerschaft in der KI-Forschung oder die ökologischen Auswirkungen der KI (vom Bergbau und Elektroschrott bei der Hardware bis hin zum CO2-Fußabdruck der für das Training und den Einsatz des Algorithmus verbrauchten Energie).

Die wichtigsten Punkte zur Überprüfung der ELSA-Ausbildung im tertiären Bildungsbereich:

- Lehrinhalte: Datenschutz, Sicherheit, Algorithmen und Ungleichheit/Gerechtigkeit/Fairness, Voreingenommenheit, Transparenz, Rechenschaftspflicht, philosophische Rahmenwerke und Moraltheorien. Ein wichtiger Bestandteil war auch die Vermittlung von Ethikrichtlinien eines Berufsverbands, Code of Conduct und Training für professionelle Entscheidungsprozesse, Berufs- und Führungskompetenzen für Informatikstudenten, rechtliche Rahmenbedingungen (z. B. GDPR), Menschenrechte, Risiken und Haftung, Datenschutz und bürgerliche Freiheiten, Arbeit und Beschäftigung, soziale Verantwortung, sozialer Kontext und Stakeholder, “ethics washing” und Umweltauswirkungen.
- Unterrichtsmodelle: eigenständige Ethikkurse oder Integration in den bereits bestehenden Lehrplan für bestimmte Fächer wie Data Science, maschinelles Lernen oder Cybersicherheit.
- Unterrichtsstunden: Unterschiedlich je nach Form (Workshops, normale Kurse, Sommerschulen, Online-Kurse usw.) und Niveau (Graduierte oder Studenten) der Kurse.
- Lehrmethoden: am häufigsten Vorlesungen von Fakultätsmitarbeitern und Gastrednern (entweder aus dem akademischen Bereich oder aus der Industrie), Fallstudien oder problemorientierte Studien. Weitere Methoden sind Debatten, Reflexion, Diskussion und Rollenspiele.
- Bewertung: Die ELSA-Ausbildung wird sowohl von den Auszubildenden als auch von den Ausbildern als sehr wichtig erachtet, wobei einhellig die Meinung vertreten wird, dass die Integration von Ethik in den gesamten Informatik-

Lehrplan als effektiver erachtet wird als eigenständige Kurse; es wird auch befürwortet, dass Ausbilder nicht nur aus dem Bereich der Informatik, sondern auch aus den Geistes- und Sozialwissenschaften sowie Gastdozenten aus dem akademischen Bereich und der Industrie eingesetzt werden, was den interdisziplinären Charakter der ELSA-Kurse unterstreicht

Diese Interdisziplinarität ist jedoch auch der Grund für die Herausforderungen, denen sich diese Kurse stellen müssen: Die ihnen gewidmeten Stunden werden als unzureichend angesehen, was auf den Mangel an qualifiziertem Personal und Fachwissen zurückzuführen ist; was die Lernziele betrifft, so variieren die Ergebnisse je nach den in den untersuchten Kursen vermittelten Inhalten - die mangelnde Konsistenz wird auf das Fehlen von Standards und die große Vielfalt der behandelten Disziplinen zurückgeführt; Die Finanzierung und die administrative Unterstützung werden erschwert, da sie die Zusammenarbeit mehrerer Universitätsabteilungen erfordern; während die Studierenden angaben, dass sie am meisten aus den nicht-technischen Aufgaben lernten, fanden sie die Kodierungsaufgaben entweder zu wenig anspruchsvoll, da sie auch mit Blick auf Studierende anderer Fachrichtungen erstellt wurden, oder zu vage, da viele von ihnen nicht zu den ihnen vertrauten Fachtypen gehörten (z. B. Memos). z. B. Memos).

Die Identifizierung der zugrundeliegenden Richtlinienprinzipien und die Untersuchung der bestehenden Schulungsansätze lieferten uns den Rahmen für einen ELSA-Schulungskurs und die Positionierung unseres vorgeschlagenen Curriculums. Die Berücksichtigung der bestehenden Kritik an den Leitlinien hilft uns, die Herausforderungen in unserem Vorhaben zu identifizieren, während die Untersuchung der bestehenden Schulungsansätze uns sowohl Prototypen liefert, denen wir folgen könnten, als auch eine Liste von Themen, die sich aus der Bewertung ergeben haben und die wir berücksichtigen müssen; dieser Bericht ist der Input für die folgenden Schritte, nämlich die Festlegung der ersten Version des Curriculums in einer Reihe von Workshops.

1 Introduction

The FAIR-DS project aims at building a cloud based Data Space with the cooperation of science and business by bringing together two initiatives, namely Gaia-X and NFDI. While FAIR stands for findable, accessible, interoperable, reusable data, the project includes an extensive Work Package (WP2) dealing with ethical and legal aspects of Data Processing. Additionally, it includes the development of a series of Demonstrators involving specific data processing and exchange scenaria in WP4. Part of the latter is the Task 4.5 regarding development of a curriculum for Ethical Legal and Societal Aspects (ELSA) training for Data Scientists[1], the first deliverable of which is the present document dealing with the description of the current landscape in the area.

The structure of this report is as follows: in section 2 we describe the problem and the need for ELSA training, in section 3 we present the current landscape with respect to guidelines and Principles and in section 4 we review the approaches to ELSA teaching so far.

2 Problem description -The need for ELSA training

The explosion of data driven applications using Artificial Intelligence (AI) in the latest years has given rise to a variety of of ethical issues that range from discrimination against vulnerable population to privacy invasion (O'Neil 2016) to the environmental impact of these algorithms (Lacoste et al. 2019), puting into focus the ever present domain of AI ethics. This raised a call for regulation that was answered by professional, national and international organisations and institutions, most recently by the European Union (European Commission 2021) as well, resulting in a variety of approaches ranging from legislation to recommendations and guidelines.

Part of the guidelines' suggestions includes education for AI practitioners. Specifically, The (High-Level Expert Group on AI (AI HLEG) 2019) mentions regulation, codes of conduct, standardisation, certifications, accountability via governance frameworks,

[1] In this report the terms Data Science/AI and Data Scientist/AI practitioner are used interchangeably. Although Data Science and AI do not overlap, they intersect in what we can describe as data-driven, non-symbolic AI, characterised by the development of data driven, self learning, intelligent systems, involving the collection, processing and analysis of big datasets as well as the usage of mostly opaque algorithms.

education and awareness to foster an ethical mind-set, stakeholder participation and social dialogue, diversity and inclusive design teams.
Similar suggestions can be found in the literature. (Martin 2019) advocates that firms that undertake AI projects that provide services to the community, for example, criminal justice, medicine, or education, should be held responsible and accountable for the algorithms they create and sell. The reason being twofold: firstly, they are the ones and usually the only ones, that are knowledgeable on how the algorithm is designed and implemented, and secondly, the values they incorporate in them by design, must respect the norms of the said community.

Additionally, since AI systems are used for decisions regarding access to social goods and have an impact on citizen's rights, AI practitioners should be required to be certified like doctors or civil engineers, lawyers, (i.e., requiring the definition of Data Scientist as a formal profession), and attend ethics education during their studies (Garzcarek and Steuer 2019), (Mittelstadt 2019).

Education on ethics can be a way of raising awareness and identifying potential morally critical situations, and is advocated for

a) AI practitioners as a part of their formal education to be included in university curricula (Burton et al. 2017; Garzcarek and Steuer 2019; Martin 2015; Leonelli 2016),
b) AI experts, in parallel with students in non-computing disciplines acquiring a basic understanding of AI techniques (Eaton et al. 2017)
c) the general population even starting from school as part of a digital literacy course (Abiteboul and Stoyanovich 2015),
d) or all the above (Taguma, Feron, and Hwee 2018; Floridi et al. 2018).

Additionally, (Moore 2020) supports a greater incorporation of politics in the ethics curricula in computer science while (Borenstein and Howard 2021) view the need to make developers aware that the technology they are building is intertwined with ethical dimensions as a first step on establishing an authentic professional mindset.
As is obvious from the above, the main focus is on education as part of a university (or even middle school) curriculum, while the already existing practitioners are expected to either self-educate and react through workplace activism (Belfield 2020), or follow regulation and acquire a formal certification.
In FAIR-DS we aim to develop a curriculum for already practising Data Scientists, which has a lot in common with a university ethics curriculum but also will take into consideration the more practical aspects that data scientists deal with at the course of their profession.

Additionally, we take into consideration that Data Science/AI professionals are not as homogeneous a group as university students, as they may come from different education levels, a variety of social and national backgrounds. (Belfield 2020) has noted that while the AI community is pretty homogeneous (as of education and earnings level), this might not be the case in the future as the high demand will be met not only by highly educated and well paid practitioners but also by people following quasi-academic routes addressing job positions demanding less high end knowledge/education. Even in academia there is a tendency to differentiate between Information Science and Computer Science students resulting in two types of Data Scientists, nameny, technical experts and the ones who have some knowledge of a of Data Science techniques, but are more oriented towards user needs, bridging business and technical teams'(Bates et al. 2020).
A survey of ACM members, representing academic institutions and industry organisations conducted by the ACM Data Science Task Force in 2018 revealed that only 54% of Data Science programmes required content in which ethics might have been considered. In the same survey, 34% of the industry representatives answered that, regarding their prospective employees qualifications, require experience in ethics, while 51% responded that this is elective (Danyluk and Paul Leidig 2021). As a result, we cannot take for granted that all AI practitioners have a tertiary education background and even those who do, received an ethics training.

In the next section we present what the main principles that the various existing guidelines prescribe, since we regard them as the context of an ELSA training curriculum.

3 Guidelines and Principles

A series of national and international organisations, initiatives, companies, NGOs, governments, professional associations, etc, have developed a number of guidelines, principles, or codes of contact, in order to delineate the ethics issues presented in AI/Data Science practice.
The NGO AlgorithmWatch created a global inventory of such guidelines comprising 173 contributions at the time of writing ('AI Ethics Guidelines Global Inventory by AlgorithmWatch' 2020).
They differentiate between binding agreements, voluntary commitments and recommendations, while they do not include legislation. The earlier publication dates back to 2010[2], the majority coming from government and civil society organisations of

[2] Some contributions are not dated

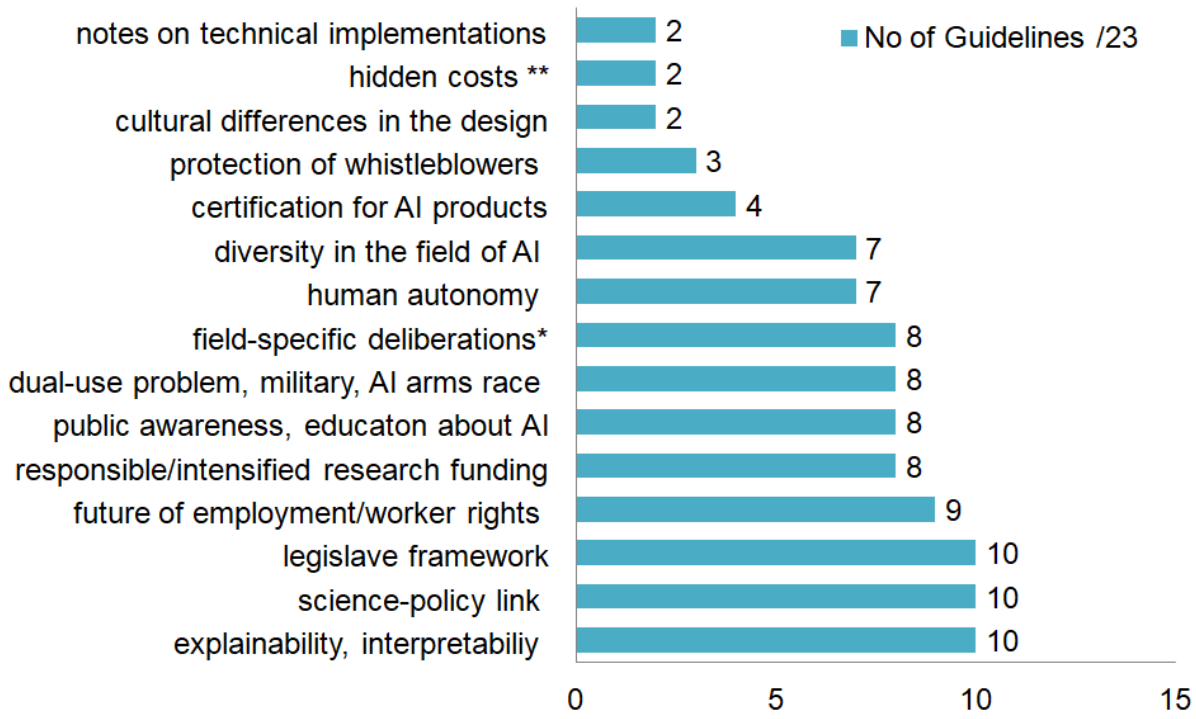
northwestern Europe and North America, 115 of them characterised as recommendations, i.e., non-enforceable guidelines.

Recent surveys that attempt to map the AI guidelines landscape identify common principles, issues or themes that run through them as:

- transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity, (Jobin, Ienca, and Vayena 2019);
- accountability, privacy and fairness, transparency, auditability, security, beneficence (in the terms of sustainability, well-being and common good) and explainability (Hagendorff 2020);
- privacy, accountability, safety and security, transparency and explainability, human control of technology, professional responsibility, promotion of human values (Fjeld et al. 2020);
- beneficence, non-maleficence, autonomy, justice and explicability (Floridi et al. 2018);
- humanity, collaboration, sharing, fairness, transparency, privacy, security, safety, accountability, AGI/ASI (Zeng, Lu, and Huangfu 2018).

The number of identified principles or issues varies according to both the number and selection criteria of the documents surveyed, as well as the granularity of the definition for each respective principle or issue. For example, (Jobin, Ienca, and Vayena 2019) and (Zeng, Lu, and Huangfu 2018) include transparency and explainability under the general notion of *transparency*, while for (Hagendorff 2020) they are examined as different issues, (Floridi et al. 2018) include accountability in explicability, while for (Hagendorff 2020) is a separate issue and (Jobin, Ienca, and Vayena 2019) and (Zeng, Lu, and Huangfu 2018) include it in responsibility. Table 1. summarises the principles named in the above surveys and includes the principles proposed by (Richards and King 2014) influential paper.

While there are a lot of common values in all the guidelines, there are also a number of omissions observed by (Hagendorff 2020). Specifically, about aspects of political abuse, governmental control, misinformation and propaganda, related social issues such as isolation and echo chambers, the rights of low level AI workers who do data labelling or content moderation, diversity in the AI industry as well as the private-public partnership in AI, research or the ecological impact of AI (from mining and e-waste regarding the hardware to the carbon footprint of the energy consumed for the algorithm training and deployment). Figure 1 presents the least frequently mentioned issues in the guidelines examined in (Hagendorff 2020).



*health, military, mobility etc, ** labeling, clickwork, content moderation, energy, resources

Figure 1. Guidelines Omissions according to (Hagendorff 2020). Number of mentions of issues out of 23 guidelines examined .

4 Teaching AI ethics

In this section we present an overview of how ELSA courses have been introduced and taught in tertiary education[3]. While most of these courses are usually referred to as "Ethics", a more detailed examination of their curriculum reveals that they also include legal and societal impacts as well. As a result, in this report we will treat the two terms as interchangeable.

We reviewed two sources of information in the bibliography, namely surveys of ethics curricula and pilot programs in the field of Computer Science and related domains. Our

[3] While our purpose in FAIR-DS is to create a curriculum for data science professionals, and not students, it was not possible to find adequately documented curricula addressed to professionals.

examination took into consideration the categories of evaluation presented in (Avci 2017), a review of 34 scholarly articles on ethics teaching in a variety of domains (mainly in health care[4]), from countries all over the world[5], specifically:

1. **Teaching scope:** the basic features of content (e.g., ethical theories, professional codes, etc.) and teaching hours.
2. **Teaching method:** the ways of actualizing the curriculum (e.g. lectures, discussions, case studies, etc.).
3. **Classroom model:** separate ethics courses or integration of ethics into the entire curriculum, online, in person, or mixed.
4. **Perception:** The perception of educators and students whether ethics courses are a positive influence and useful later in their practice.
5. **Performance:** How well the programs work.
6. **Effectiveness :** The results of an implemented ethics program.

The distinction between performance and effectiveness is defined by the authors as follows: '*the term performance points out the application of a program regardless of its consequences, whereas effectiveness looks into the outcomes of a program. In other words, performance focuses on whether a program is applied successfully, while effectiveness concentrates on whether the results of an implemented program are effective*', (Avci 2017, 10)**.**

Teaching scope, methods and classroom model definitions are adopted as defined above, while perception, performance and effectiveness are grouped under the general concept of evaluation, since we could not identify a specific and widely accepted way to measure each one of them, while in some papers they were not specified at all (an observation also made in (Avci 2017)).

Regarding the evaluation criteria, we considered Kirkpatrick's framework as presented in (Steele et al. 2016) as the dominant model of training evaluation**.** The model is hierarchical and consists of four levels: reaction, learning, behaviour, and results. Specifically:

- Reaction level is based on the trainees' perceptions about the usefulness of a program.

[4] The full list if scientific domains is : Nursing (7), Medicine and Health Science (6), Psychology (3), Business (2), Education (2), Ethics (1), Science and Engineering (1), Social Work (1), Public Relations (1), Information Systems (1), and Accounting (1) .

[5] Specifically, United States (12), Australia (2), Taiwan (1), Belgium (1), Egypt (1), South Korea (1), Ireland (1), New Zealand (1), and Switzerland (1).Furthermore, 3 research studies were international, and 2 others were about NorthAmerica (the United States and Canada).

- Learning, a category ranging from assessing declarative knowledge to acquisition of skills for solving ethical issues, or even to the overall change of the trainees' overall attitude towards ethics in their professional context.
- Behaviour, referring to transferring performance, or training-relevant behaviours outside of the training context. It is measured by assessing the impact of the training regarding targeted behavioural outcomes.These can be used to predict the desired behaviours in a professional context.
- Results, i.e., measuring the impact that the training has on broader organisational or institutional outcomes.

The rest of this section is structured as follows : the first subsection presents the findings of surveys of ethics courses as taught mainly in European and North American Universities in the discipline of Computer Science (CS) or other relevant disciplines such as Engineering, but also Business Studies and Social Sciences, or in specific CS domains such as Data Science and Machine Learning (ML).

Then, we continue by examining specific pilot programs, namely the ones realised in Harvard and Stanford universities, an Human Computer Communication (HCC) summer course by University of Colorado, concluding with a proposed ML and ethics integrated course. These programs were chosen because they are well documented in literature via peer-reviewed publications and also provide public access to all curricular material they use. All of them are integrated courses, since we regard them as more appropriate models for ELSA training for professionals, an opinion supported by the evaluation of similar courses as will be presented later in the report.

4.1 ELSA training in review articles

4.1.1 Teaching scope

An increasing number of universities have included Ethics in their Computer Science or relevant disciplines (e.g. Engineering) curricula. A number of surveys analyse the curricula of mainly European and North American Universities. We review 5 comprehensive and relatively recent ones (from 2017 to 2021), specifically:

1. (Stavrakakis et al. 2021) outline the results of a large-scale survey of European academics about existing competencies in the teaching of computer ethics in Computer Science and related disciplines (Ethics4EU 2021). Having as a subject the ethical topics taught in the respective Computer Science and/or Computer Science related programmes, the study was completed by respondents from 61 universities across 23 European countries.
2. (Fiesler, Garrett, and Beard 2020) present an in-depth qualitative analysis of 115 syllabi from university technology ethics courses located mainly in the US.

3. (Saltz et al. 2019) focused on Machine Learning courses and the integration or not of ethics as part of their coursework. The authors examined the syllabi of top twenty computer science programs at U.S. universities, as ranked by their graduate programs in U.S. News and World Report in 2018, identifying 186 ML-related courses.
4. (Bielefeldt et al. 2017) research how ESI (Ethical and Societal Impact) topics are taught to engineering and computing students in the United States, using a survey of faculty members. Overall, there were 1448 respondents who completed at least 30% of the survey, including 1242 who finished the survey
5. (Mulhearn et al. 2017) conducted a study based on 330 training programs on ethics education focused in Science and Engineering, but also included studies from the business domain (in order to add stability to the cluster procedure they employed). Specifically, they classified the ethical education programmes in 8 clusters according to four instructional categories, namely, instructional content, processes, delivery methods, and activities, and measured the effectiveness of each cluster using Cohen's d effect size (Cohen 1992).

The classroom models of the courses surveyed in the above reviews, were either standalone or incorporated into the already existing curriculum into specific subjects such as data science, machine learning or computer security. They covered a wide variety regarding the level (graduate or undergraduate), while the form (workshops, ordinary courses, summer schools, online courses, etc.) and the teaching hours devoted to ELSA training differed accordingly.

Among the most common topics taught were ethics issues such as privacy, security, algorithms, and inequality/justice/fairness, bias, transparency, accountability, and responsibility, as well as more specific cases such as courses on fairness and data validity, and philosophical frameworks and moral theories. Important part was also the teaching of Code of Ethics from a professional body, Codes of Conduct and Professional Decision Processes Training, professional and leadership skills for computer science students, specifically in senior project classes.

Additionally, legal and societal issues that were addressed were:

- legal frameworks (e.g., GDPR),
- human rights, risk and liabilities,
- privacy and civil liberties,
- work and labour,
- social responsibility,
- social context and stakeholders,
- ethics washing and environmental impact.

In this context, (Bielefeldt et al. 2017) note that privacy, ethical theories, war (and "other") were taught more commonly to computing than other engineering students while environmental protection, sustainability, engineering decisions under uncertainty, safety, engineering codes of ethics, engineering & poverty, and bioethics are more commonly taught in courses for engineering students than computing students.

Table 2 summarises the kinds of courses where ELSA issues are taught, while Table 3. Their teaching scope, as presented in the above surveys.

4.1.2 Teaching methods

Regarding the teaching methods, in addition to the above reviews, we examined a less recent study by (Dexter et al. 2013) who conducted an investigation into practises in computer ethics curricula at the graduate level at four US institutions[6]. The survey was addressed to both students and faculty members and focused specifically on the instruction methods used in already existing courses and offered suggestions on additional ones.

In all the papers reviewed, commonly referred teaching methods included lectures by faculty staff and guest speakers (either from the academia or the industry), and case or problem based studies as the most popular ones, followed by debates, reflection, discussion and role-playing.

Additionally, (Mulhearn et al. 2017) present as a special case, online courses and their respective methods of online and self directed instruction as well as web based discussion. Furthermore (Dexter et al. 2013) propose methods such as real world examples, mixed methods, hands on activities while they also advocate on integrating ethics training into the existing curriculum as opposed to having it as a standalone course.
(Fiesler, Garrett, and Beard 2020) concentrate on the topics offered and the goals of ethics training and do not elaborate on the training methods. They generally refer to pedagogy, with respect to argumentative writing in order to achieve the goal of improving communication skills, or employing gamification, immersive theatre, incorporation of science fiction, integration into HCI, and machine learning classes, as referred in the bibliography.
(Saltz et al. 2019) focus on ML courses; while they identify topics dealt in ethics education, they do not elaborate on the methods. However, they present a pilot

[6] Brooklyn College, University of Texas at Austin, University of Wisconsin-Stout and University of Illinois Springfield.The survey was administered to 924 students at the four institutions with an overall response rate of 27%.The faculty survey was administered to 167 faculty members at computer and information science departments of the four institutions with an overall response rate of 33%.

program (described in the next subsection) that was inspired by a Systematic Literature Review (SLR) on teaching ethics and provide example course assignments that embed ethics within core ML topics.

Table 4 summarises the teaching methods described in the above papers

4.1.3 Evaluation

In the reviewed papers, the authors did not conduct their own experimental evaluations of the the respective courses but based their concluding remarks either on the reaction level of Kirkpatrick's evaluation framework (i.e., measuring, affective reactions toward training, realism of training, utility of training, difficulty obtaining materials, etc.) or the learning level, regarding the knowledge, skills and behavioural impact of the courses surveyed.

While reaction level assessment does not indicate the effectiveness of a training program, since it is highly subjective, it can provide useful insights and is widely used since it requires the fewest resources in comparison to others; additionally, the diversity in the learning criteria employed make the measurement of the learning outcomes of the various courses not comparable, since the notion of 'learning' may have different meanings in each of them (Steele et al. 2016).

Taking the above into consideration, ELSA courses in Computer Science and the related disciplines are assessed very important (Stavrakakis et al. 2021; Fiesler, Garrett, and Beard 2020). However, only two third of the universities surveyed in (Stavrakakis et al. 2021) offer them**,** and those who do, devote a small number of hours in teaching ethics.Where ethics is not taught, the main reason given is the lack of skilled staff and expertise. This result is echoed in the perception of both students and faculty surveyed in (Bielefeldt et al. 2017) that "the majority of both computing and engineering programs are seen as providing insufficient ESI education".

Regarding the learning objectives, computer ethics is often taught as a standalone subject and is considered more important for some computer science topics, such as AI/Data Science and Computer Security (Stavrakakis et al. 2021), while (Fiesler, Garrett, and Beard 2020) note that their results reveal a great variety regarding the content taught in the surveyed courses. The lack of consistency on the topics covered is attributed by the authors to the lack of standards and the wide variety of the disciplines covered. Additionally, they codify the learning objectives mentioned in their review data, as (i) critique, (ii) spotting ethical issues, (iii) make arguments, (iv) improve communication, (v) see multiple perspectives, (vi) create solutions, (vii) consider consequences and (viii) apply rules (in descending mention-frequency order) while in parallel they underline the challenge of assessing them.

(Mulhearn et al. 2017) assessed the effectiveness of training approaches to ethics education by using a typology they created based on combinations of instructional content, processes, delivery, and course activities. Their typology consists of 8 clusters : Field-specific compliance, Online, Professional decision processes, General discussion, Targeted experimental interventions, Norm adherence, Exemplar based, and Philosophical self-reflection. They evaluated the clusters according to nine commonly used ethics criteria (corresponding to learning objectives):moral reasoning, knowledge, ethical awareness, ethical decision making, perceptions of self, moral judgement, meta-cognitive strategies, conceptual development, and perceptions of others using Cohen's *d* effect size (Cohen 1992)[7].

They conclude that the surveyed types demonstrate small to moderate effectiveness in terms of above mentioned criteria.They observe that the two most commonly employed instructional approaches, General Discussion Training and Norm Adherence Training, proved to be the least effective, a fact that can likely be attributed to the lack of a specific emphasis on content and processes. On the other hand, the most effective training type was Targeted Experimental Interventions (case studies, active participation, etc.), while Exemplar Based Training also scored very high in effectiveness, which is something to be expected since it is based on examples relating to human rights and welfare concerns.
However, they support that it is the moderately effective Professional Decision Processes Training and Field-Specific Compliance Training, present the most viable ethics training approaches. Professional Decision Processes Training focuses on processes for navigating through complex, ambiguous ethical issues, whereas Field-Specific Compliance Training emphasises on regulatory aspects commonly accepted in a given field (guidelines, legality and human rights) as opposed to ambiguous ethical decision making. Concluding, they remark that while some intervention types appear to be more effective than the others, they would hardly suggest that any of them should be abandoned and that further research is needed on determining the effectiveness of a training process.

Generally, there is a consensus that infused ethics in all of computer science curriculum is deemed as more effective than standalone courses (Stavrakakis et al. 2021; Bielefeldt et al. 2017; Fiesler, Garrett, and Beard 2020; Saltz et al. 2019; Dexter et al. 2013)**,** an observation that is corroborated by the higher effectiveness of Targeted Experimental Intervention and Exemplar Based Training, which are more hands on training approaches and include case studies and real world examples (Mulhearn et al. 2017)**.**

[7] Effect sizes of .20 indicate a small effect, .50 indicate a medium-sized effect, and .80 or greater indicate a large effect

There is also support in employing instructors not only from a computer science background but also humanities and social sciences, as well as guest lecturers form academia and industry, which underlines the interdisciplinary nature of ELSA courses (Stavrakakis et al. 2021; Fiesler, Garrett, and Beard 2020; Bielefeldt et al. 2017).

4.2 Pilot programs and exemplary courses

In this section we will present two recent pilot ELSA courses, an example course module for integrating ethics in ML courses and one in Human-Centered-Computing(HCC) as presented in recent bibliography. The pilots were selected because unlike a standalone ethics class they implemented an embedded approach of including ELSA issues in data science courses, an approach that is supported by the findings presented in the previous section, and which we consider would be most appropriate to a curriculum addressed to data science practitioners.

We examine them as prototypes regarding their syllabus, material, teaching methods and evaluation.

(Grosz et al. 2019) present the Harvard Pilot program of "Embedded EthiCS which integrated ethics in the standard computer science curriculum by modifying already existing courses. It lasted three semesters (Spring 2017, Fall 2017, and Spring 2018) and included 14 separate courses in a variety of areas, such as introductory and theory courses, computer science and economics, programming languages and data systems, HCI and AI.
The program embedded teaching assistants, either Ph.D. students or postdoctoral fellows in philosophy with a strong background in ethics and considerable teaching experience to design the ethics modules.
The issues covered were privacy and intellectual property, fairness and discrimination, fake news and social media platforms, diversity and equal opportunity, verifyability and interpretability, social responsibility and moral decision-making.
The embedded EthiCS module for each course introduced the relevant philosophical or legal framework, the students were engaged in a group based discussion or simulation, and the module's assignment was incorporated into the course's final project.
The aim of the pilot was that students acquire the ability to identify and anticipate ethical problems in the development and use of computing technologies; the ability to reason, both alone and in collaboration with others, about those problems and potential solutions to them, using concepts and principles from moral philosophy; and the ability to communicate effectively their understanding of how to address those problems. The modules emphasised *active learning* activities and assignments that teach students to apply the philosophical ideas they have learned to concrete, real-world ethical problems as recommended by recent studies of ethics education.

(Reich et al. 2020) report on a curricular experiment at Stanford University focused on teaching computer ethics. The course took a multidisciplinary approach by including three faculty instructors, from philosophy, political science, and CS. It involved four course modules: algorithmic decision-making, data privacy and civil liberties, AI and autonomous systems, and the power of platform companies.

Each unit follows the following structure: Philosophical framework, Technical perspective, Policy implications, Case study, Making choices.

Teaching methods included: lectures, small-group discussions, panels of guest speakers drawn from academia, industry, civil society, and government to provide a practitioner's view of the topics addressed, and custom case studies[8].

The students are given assignments which include: writing essays or memos relevant to specific case studies, e.g., on a public policy issue or justifying the decisions they would make; manipulate given code, e.g., for measuring a model's fairness, or simulation of information bubble formation; interviewing stakeholders to assess how they are impacted by certain design decisions.

The course is cross-listed in CS as well as the Philosophy, Political Science, and Public Policy departments (among others) and draws students from across the university having as a prerequisite only an introductory CS course.

The following two courses are integrating ethics in the specific domains of Human-Centered-Computing (HCC) and Machine Learning (ML) respectively.

(Skirpan et al. 2018) report on an upper-level undergraduate Human-Centered-Computing (HCC) which expanded to address activities that require ethical judgement in order to evaluate students' reactions to a combined ethics and engineering practice course.
The class taught methods for prototyping and evaluating computing systems from a user-centred perspective, further emphasising ways in which the design, development, and deployment of technologies have human consequences, helping students assess and plan for those consequences and develop the skills necessary to be socially-conscious and responsible engineers.

Teaching methods and assignments: In-Class Activities Class time was split up into two parts, namely lectures and workshops.

[8] The materials of (videos of the speaker panels and the case studies themselves) are freely available for use by the broader community
http://ai.stanford.edu/users/sahami/ethicscasestudies/ last accessed 22.11.21

Lectures were given by the primary instructor or an invited guest lecturer which combined the lecture content with practical exercises and/or a course related group project. The guest lecturers included a researcher focused on co-operative ownership of data and software, a privacy lawyer with practical knowledge of the laws and regulations relevant to computer scientists, an artist specialised in typography and layout, a researcher focused on terms of service agreements and online harassment, and an emeritus professor who detailed his experience watching a 50-year transition of technology.

Workshops combined the lecture content with hands-on assignments. Assignments comprised individual assignments (weekly reading reflections, participation submissions, and an individual applied exercise regarding a topic from the week) and a course-long project which included reflecting on the social impacts of their system if it were to become a commercialised product.

(Saltz et al. 2019) whose review is presented in the previous subsections, went on supplementing their analysis by conducting a systematic literature review (SLR) in order to identify the concepts that will enable them to establish a robust set of materials for teaching ethics. Their SLR identified three key areas of focus: (i) Oversight Challenges (accountability and responsibility-both in a legal and ethical sense), (ii) Data Challenges (privacy and anonymity, availability and validity-both legal and ethical aspects) and (iii) Model Related Challenges(model and modeller bias, model transparency and interpretation-both regarding ethical issues and technical solutions).

Consequently, they propose a framework for integrating ethics in a Machine Learning course. Their purpose is to imbue ethics into assignments that are already common to ML classes thus eliminating the need for extra work. Their proposed example modules consist of the common teaching subjects of logistic regression, random forest classifier, and deep convolutional neural models[9].

Each module beside the objectives includes questions regarding ethics, such as the appropriateness of the dataset for the given task, possible kinds of data bias, imbalances in the training set, issues form data aggregation, which laws might be applicable in the specific case, model suitability, multiple model application and assessment of each model for a given problem, and whether there is room for results misinterpretation as well as real-world, job-relevant considerations bordering with business ethics, such as what will be their attitude to ethically dubious demands in a working environment .

[9] The full assignment documents, which further link to Python notebooks containing walk-through solutions http://github.com/ProbableModels/acm_paper_2018 last accessed on 22.11.21

Table 5. Summarises the key points regarding the pilot and exemplary courses

4.2.1 Evaluation of pilot programs

In assessing the Harvard pilot embedded EthiCS program on the reaction level (Grosz et al. 2019) present a student survey which shows that more than 80% of the students and 90% of the faculty were positive regarding the program, while the modules are most effective when the technical material is connected with ethical issues already salient to students. On the negative side, the problems encountered had to do with the cross-disciplinary nature of the program. Specifically, insecurities of the faculty towards the other disciplines, the use of different methodologies and vocabularies in each discipline, and the lack of philosophers willing to teach and develop modules for CS classes.

The authors mention also that the effectiveness of the program should also be measured by its long-term impact, but this is left as an open question, They also underline the institutional challenges of such a program, such as funding and administrational support, made more difficult due to the cross-disciplinary nature of the course which demands the cooperation of several university departments.

(Reich et al. 2020) in their course evaluation student survey regarding the content, quality of instruction and course organisation achieved scores comparable and above the university media and whatever problems existed were stemming from its multidisciplinary nature. The students' perception was also that they learned the most from the non-technical assignments and that they found the coding assignments not too challenging. However, the assignments were made having also in mind students from other disciplines with a minimal prerequisite, who incidentally, the course failed to attract (at least as many as they expected). Regarding other dimensions, the effectiveness of the use cases was more or less uniformly distributed, while on the negative/improvement suggestion side, the main issue was on the evaluation of the assignments, which was perceived by the students as too vague. This is attributed by the authors both to the newness of the course as well as its multidisciplinary nature, since many of the students were not familiar with types of assignments such as a policy memo. As an improvement they suggest a diversification of student assignments according to their respective discipline background.

(Skirpan et al. 2018) in their assessment of the expanded HCC course, did not observe any differences in the quality of the work delivered by the students or their grades distribution when compared with similar courses which did not include embedded ethics modules(which was one of the targets of the course. The post-survey they conducted affirmed that the students learned the core contents effectively, so whatever ethics components injected, did not affect the core competence quality. Regarding the

ethics part, they conducted a pre- and post-class survey of the students, containing short-answers, open-ended questions related to the ethical implications of technology and takeaways from the course. Findings included that the number of students regarding ethics as important to their careers increased after the course. The students reported that they developed a higher level of sensitivity to how technology impacts society, while many of them asserted that this was the most thought-provoking course they had taken. On the negative side the authors mention that some of the assignments were too technically complex for an undergraduate course, specifically that the biggest limitation of the course was that some of the technical counterparts of the social issues mentioned (e.g., diving deep into machine learning or cryptography) could not be fully addressed due to time and content scoping constraints. The main takeaway of the experiment was, according to the authors, that ethical thinking should be introduced early in CS programs, be continuous and embedded into the regular courses, employing relative case studies and real world problems.

The results are summarised in Table 6.

5. Conclusions

In the context of the FAIR-Ds Project, as part of WP4, we are going to develop a curriculum for ELSA training for Data Scientists. In this first deliverable we present the existing landscape as determined by two factors: the guidelines issued by a variety of international organisations and the existing ELSA training approaches, as manifested in tertiary education courses.

Identifying the underlying principles of the existing guidelines provided us with the framework of such a course, while in parallel taking into account the existing criticism helped us spot the challenges in our endeavour.

The survey of the existing training approaches helped us to position our course and the examination of specific ELSA courses provided us with prototypes we could follow, as well as a list of issues that resulted from the assessment of said courses and we have to take into consideration.

This report will be the input for the following steps as defined in the Project Proposal, namely defining the first version of the curriculum in a series of Workshops.

References

Abiteboul, Serge, and Julia Stoyanovich. 2015. 'Data, Responsibly'. *ACM SIGMOD Blog* (blog). 20 November 2015. http://wp.sigmod.org/?p=1900.

'AI Ethics Guidelines Global Inventory by AlgorithmWatch'. 2020. AI Ethics Guidelines Global Inventory. 2020. https://inventory.algorithmwatch.org.

Avci, Ercan. 2017. 'Learning from Experiences to Determine Quality in Ethics Education'. *International Journal of Ethics Education* 2 (1): 3–16. https://doi.org/10.1007/s40889-016-0027-6.

Bates, Jo, David Cameron, Alessandro Checco, Paul Clough, Frank Hopfgartner, Suvodeep Mazumdar, Laura Sbaffi, and Peter Stordy. 2020. 'Integrating FATE/Critical Data Studies into Data Science Curricula: Where Are We Going and How Do We Get There?', 11.

Belfield, Haydn. 2020. 'Activism by the AI Community: Analysing Recent Achievements and Future Prospects'. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 15–21. AIES '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375814.

Bielefeldt, A., M. Polmear, Chris Swan, D. Knight, and N. Canney. 2017. 'An Overview of the Microethics and Macroethics Education of Computing Students in the United States'. *2017 IEEE Frontiers in Education Conference (FIE)*. https://doi.org/10.1109/FIE.2017.8190445.

Borenstein, Jason, and Ayanna Howard. 2021. 'Emerging Challenges in AI and the Need for AI Ethics Education'. *AI and Ethics* 1 (1): 61–65. https://doi.org/10.1007/s43681-020-00002-7.

Burton, Emanuelle, Judy Goldsmith, Sven Koenig, Benjamin Kuipers, Nicholas Mattei, and Toby Walsh. 2017. 'Ethical Considerations in Artificial Intelligence Courses'. *AI Magazine* 38 (2): 22–34. https://doi.org/10.1609/aimag.v38i2.2731.

Cohen, Jacob. 1992. 'A Power Primer'. *Psychological Bulletin* 112 (1): 155–59. https://doi.org/10.1037/0033-2909.112.1.155.

Danyluk, Andrea and Paul Leidig. 2021. 'Computing Competencies for Undergraduate Data Science Curricula-ACM Data Science Task Force'. ACM. https://dstf.acm.org/DSTF_Final_Report.pdf.

Dexter, Scott, Elizabeth Buchanan, Kellen Dins, Kenneth R. Fleischmann, and Keith Miller. 2013. 'Characterizing the Need for Graduate Ethics Education'. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, 153–58. SIGCSE '13. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2445196.2445245.

Eaton, Eric, Sven Koenig, Claudia Schulz, Francesco Maurelli, John Lee, Joshua Eckroth, Mark Crowley, et al. 2017. 'Blue Sky Ideas in Artificial Intelligence Education from the EAAI 2017 New and Future AI Educator Program'. *AI Matters* 3 (February). https://doi.org/10.1145/3175502.3175509.

Ethics4EU. 2021. 'Existing Competencies in the Teaching of Ethics in Computer Science Faculties – Research Report –'. Erasmus+ Project. http://ethics4eu.eu/outcomes/existing-competencies-in-the-teaching-of-ethics-in-computer-science-faculties-research-report/.

European Commission. 2021. 'Proposal for a Regulation on a European Approach for

Artificial Intelligence'. Proposal COM(2021) 206 final2021/0106 (COD). Brussels. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence.

Fiesler, Casey, Natalie Garrett, and Nathan Beard. 2020. 'What Do We Teach When We Teach Tech Ethics?: A Syllabi Analysis'. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 289–95. Portland OR USA: ACM. https://doi.org/10.1145/3328778.3366825.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI'. SSRN Scholarly Paper ID 3518482. Rochester, NY: Berkman Klein Center for Internet & Society. https://doi.org/10.2139/ssrn.3518482.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, et al. 2018. 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations'. *Minds and Machines* 28 (4): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Garzcarek, Ursula, and Detlef Steuer. 2019. 'Approaching Ethical Guidelines for Data Scientists'. *ArXiv:1901.04824 [Cs, Stat]*, January. http://arxiv.org/abs/1901.04824.

Grosz, Barbara J., David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2019. 'Embedded EthiCS: Integrating Ethics across CS Education'. *Communications of the ACM* 62 (8): 54–61. https://doi.org/10.1145/3330794.

Hagendorff, Thilo. 2020. 'The Ethics of AI Ethics: An Evaluation of Guidelines'. *Minds and Machines* 30 (1): 99–120. https://doi.org/10.1007/s11023-020-09517-8.

High-Level Expert Group on AI (AI HLEG). 2019. 'Ethics Guidelines for Trustworthy AI'. Brussels: European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. 'The Global Landscape of AI Ethics Guidelines'. *Nature Machine Intelligence* 1 (9): 389–99.

Lacoste, Alexandre, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. 2019. 'Quantifying the Carbon Emissions of Machine Learning'. *ArXiv:1910.09700 [Cs]*, November. http://arxiv.org/abs/1910.09700.

Leonelli, Sabina. 2016. 'Locating Ethics in Data Science: Responsibility and Accountability in Global and Distributed Knowledge Production Systems'. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 374 (2083). https://doi.org/10.1098/rsta.2016.0122.

Martin, Kirsten. 2015. 'Ethical Issues in Big Data Industry'. *MIS Quarterly Executive*, June.

———. 2019. 'Ethical Implications and Accountability of Algorithms'. *Journal of Business Ethics* 160 (December). https://doi.org/10.1007/s10551-018-3921-3.

Mittelstadt, Brent. 2019. 'Principles Alone Cannot Guarantee Ethical AI'. *Nature Machine Intelligence* 1 (11): 501–7. https://doi.org/10.1038/s42256-019-0114-4.

Moore, Jared. 2020. 'Towards a More Representative Politics in the Ethics of Computer Science', 11.

Mulhearn, Tyler J., Logan M. Steele, Logan L. Watts, Kelsey E. Medeiros, Michael D. Mumford, and Shane Connelly. 2017. 'Review of Instructional Approaches in Ethics Education'. *Science and Engineering Ethics* 23 (3): 883–912.

https://doi.org/10.1007/s11948-016-9803-0.
O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 1st edition. New York: Crown.
Reich, Rob, Mehran Sahami, Jeremy M. Weinstein, and Hilary Cohen. 2020. 'Teaching Computer Ethics: A Deeply Multidisciplinary Approach'. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 296–302. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3328778.3366951.
Richards, Neil M., and Jonathan H. King. 2014. 'Big Data Ethics'. *Wake Forest Law Review* 49 (2): 393–432. https://heinonline.org/HOL/P?h=hein.journals/wflr49&i=405.
Saltz, Jeffrey, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. 'Integrating Ethics within Machine Learning Courses'. *ACM Transactions on Computing Education* 19 (4): 32:1-32:26. https://doi.org/10.1145/3341164.
Skirpan, Michael, Nathan Beard, Srinjita Bhaduri, Casey Fiesler, and Tom Yeh. 2018. 'Ethics Education in Context: A Case Study of Novel Ethics Activities for the CS Classroom'. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 940–45. SIGCSE '18. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3159450.3159573.
Stavrakakis, Ioannis, Damian Gordon, Brendan Tierney, Anna Becevel, Emma Murphy, Gordana Dodig-Crnkovic, Radu Dobrin, et al. 2021. 'The Teaching of Computer Ethics on Computer Science and Related Degree Programmes. a European Survey'. *International Journal of Ethics Education*, October. https://doi.org/10.1007/s40889-021-00135-1.
Steele, Logan M., Tyler J. Mulhearn, Kelsey E. Medeiros, Logan L. Watts, Shane Connelly, and Michael D. Mumford. 2016. 'How Do We Know What Works? A Review and Critique of Current Practices in Ethics Training Evaluation'. *Accountability in Research* 23 (6): 319–50. https://doi.org/10.1080/08989621.2016.1186547.
Taguma, Miho, Eva Feron, and Meow Hwee. 2018. *Future of Education and Skills 2030: Conceptual Learning Framework*. OECD. https://www.oecd.org/education/2030/Education-and-AI-preparing-for-the-future-AI-Attitudes-and-Values.pdf.
Zeng, Yi, Enmeng Lu, and Cunqing Huangfu. 2018. 'Linking Artificial Intelligence Principles'. *ArXiv Preprint ArXiv:1812.04814*.

Appendix A

Hagendorf 2020	Jobin et al. 2019		Floridi et al. 2018		Fjeld et al. 2020	Zeng, Lu, and Huangfu 2018		Richards and King 2014 (not a survey)	
	Principle	Encompasing concets	Principle	Encompasing concets		Principle	Encompasing concets	Principle	Encompasing concets
privacy protection	Transparency	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing	Beneficence:	Promoting Well-Being, Preserving Dignity, and Sustaining the Planet	privacy	Humanity	humanity, beneficial, well-being, human value, human right, dignity, freedom, edu-cation, common good, human-centered, human-friendly	Privacy	rules to regulate the flows of data
fairness, non-discrimination, justice	Justice & fairness	Justice, fairness, consistency, inclusion, equality, equity, (non-)bias, (non-) discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution	Non-maleficence:	Privacy, Security and "Capability Caution"	accountability	Collaboration	collaboration, partnership, cooperation, dialogue	Confidentiality	private digital information that we share with third parties we trust can still be regulated by privacy law
accountability	Non-maleficence	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion	Autonomy:	The Power to Decide (Whether to Decide)	safety and security	Share	share, equal, equity, inequality, inequality	transparency	fosters trust by being able to hold others accountable; Transparency Paradox: inadvertent disclosures will cause unexpected outcomes that harm privacy and breach confidentiality
transparency, openness	Responsibility	Responsibility, accountability, liability, acting with integrity	Justice:	Promoting Prosperity and Preserving Solidarity	transparency and explainability	Fairness	fairness, justice, bias, discrimination, prej-udice	Identity	the ability of individuals to define who they are
safety, cybersecurity	Privacy	Privacy, personal or private information	Explicability:	Enabling the Other Principles Through Intelligibility and Accountability	human control of technology	Transparency	transparency, explainable, predictable, in-telligible, audit, trace, opaque		
common good, sustainability, well-being	Beneficence	Benefits, beneficence, well-being, peace, social good, common good			professional responsibility	Privacy	privacy, personal information, data pro-tection, informed, explicit confirmation, control the data, notice and consent		
human oversight, control, auditing	Freedom & autonomy	Freedom, autonomy, consent, choice, self-determination, liberty, empowerment			promotion of human values	Security	security, cybersecurity, cyberattack, hacks, confidential		
solidarity, inclusion, social cohesion	Trust	Trust				Safety	safety, validation, verification, test, con-trollability, under control, control the risks, human control		
explainability, interpretability	Sustainability	Sustainability, environment (nature), energy, resources (energy)				Accountability	accountability, responsibility		
science-policy link	Dignity	Dignity				AGI/ASI	AGI, superintelligence, super intelligence		
legislative framework, legal status of AI systems	Solidarity	Solidarity, social security, cohesion							
future of employment/worker rights									
responsible/intensified research funding									
public awareness, education about AI and its risks									
dual-use problem, military, AI arms race									
field-specific deliberations (health, military, mobility etc.)									
human autonomy									
diversity in the field of AI									
certification for AI products									
protection of whistleblowers									
cultural differences in the ethically aligned design of AI systems									
hidden costs (labeling, clickwork, content moderation, energy, resources)									
notes on technical implementations									

Table 1. Principles according to bibliography

Table 2. Kinds of courses surveyed				
(Stavrakakis et al. 2021)*	**(Fiesler, Garrett, and Beard 2020)**	**(Saltz et al. 2019)**	**(Bielefeldt et al. 2017)**	**(Mulhearn et al. 2017)**
Stabdalone ethics classes	Stabdalone ethics classes	Machine learning	Full course on ESI	Ethics programs in science, engineering and business
1. PE6_7: Artificial Intelligence, Intelligent Systems, Multi Agent Systems			ESI in professional issues courses	
2. PE6_5: Cryptology, Security, Privacy, Quantum Crypto				
3. PE6_9: Human Computer Interaction and Interface, Visualization and Natural Language Processing				
4. PE6_11: Machine Learning, Statistical Data Processing and Applications using Signal Processing (e.g. Speech, Image, Video)				
5. PE6_3: Software Engineering, Operating Systems, Computer Languages				
6. PE6_10: Web and Information Systems, Database Systems, Information Retrieval and Digital Libraries, Data Fusion.				
*classifications are based on criteria by ACM – Association for Computing Machinery				

Table 3.Teaching Scope				
(Stavrakakis et al. 2021)	**(Fiesler, Garrett, and Beard 2020)**	**(Saltz et al. 2019)**	**(Bielefeldt et al. 2017)**	**(Mulhearn et al. 2017)**
Ethics specific to subjects (e.g. Data Science).	Law & policy	Ethics	Professional practice issues	Moral philosophy
Code of Ethics from a professional body	Privacy & surveillance	privacy	Societal impacts of technology	Guidelines and values
Ethical Theory	Philosophy	responsibility	Privacy and civil liberties	Historical development, and contemporary ethical issues
Responsibility	Inequality, justice & human rights	fairness	Ethical failures / disasters	personal responsibility
Legal Issues	AI & algorithms	bias	Risk and liabilities	human rights
Ethics Washing	Social & environmental impact	transparency	Engineering codes of ethics	examples evidencing salient ethical situations
Epistemic Issues	Civic responsibility & misinformation	accountability	Ethics in design projects	codes of conduct
EU Ethics	AI & robots		Responsible conduct of research	privacy
	Business & economics		Ethical theories	field-specific content
	Professional ethics		Safety	stakeholders
	Work & labour		Engineering decisions under uncertainty	protection of human participants
	Design		Social justice	legality
	Cybersecurity		Sustainability	
	Research ethics		War, peace, military applications	
	Medical/health		Environmental protection	
			Engineering and	

			poverty	
			Nanotechnology	
			Bioethics	

Table 4. Instruction methods

(Stavrakakis et al. 2021)	**(Dexter et al 2013)**	**(Bielefeldt et al. 2017)**	**(Mulhearn et al. 2017)**
Lectures	Lecture—Instructor	In-class discussion	General Discussion Training -moderate amount of discussion in small and large groups
Case Studies	Case Studies	Case studies	self-reflection and essays
Debates	Reading Current Events	Lecture	online instruction, self-directed instruction, and web based discussion
Project-Based/Problem-Based Learning	Lecture—Guests	Examples of professional scenarios	case based instruction
Guest Lecturers (academic)	Role Playing	Videos, movie clips	active participation
Guest Lecturers (industry)	Simulations	Project based learning	lecture
Guest Lecturers (government-bodies)	Other**	In-class debates/role playing	problem-based learning
Guest Lecturers (professional bodies)	*Curriculum Integration*	Guest lectures	team-based learning
Work-Based Learning	*Real-World Examples*	Reflection	
Role-playing	*Discussion*	Engineering design	
eLearning Blended Learning	*Mixed methods*	Humanist readings	
Other*	*Hands on activities*	Think-pair-share	
		Moral exemplars	
		Problem solving heuristics	
		Service-learning	
	Proposed method in italics	Other***	

*Groupwork, Peer Instruction (using PeerWise), Student Discussions, Seminars and Guest lecturers from the Arts, Interviews with Researchers, Student Presentations, Embedded Videos

**"Design games, a paper assignment" and "Discussion of codes of ethics."
***Current news stories , Student presentations, Position papers, Student research paper , On-line discussions, Online lectures before class, Online module based on textbook, CITI training, Read papers on morality of crypto research, Fiction readings as case studies , Writing analyses of ethical issues, Participate in professional meeting,or conference, Design studios, Develop, describe own moral theory

Table 5. Key points of pilot programs

Bib. reference	(Grosz et al. 2019)	(Reich et al. 2020)	(Skirpan et al. 2018)	(Saltz et al. 2019)
Faculties Departments	CS (introductory and theory courses, computer science and economics, programming languages and data systems, HCI, AI) Harvard University	CS as well as the Philosophy, Political Science, and Public Policy in Stanford University	CS (HCC) University of Colorado	CS (Machine Learning (ML)) course proposal
Level	Graduate and undergraduate	Graduate and undergraduate	intensive five-week, summer implementation of an undergraduate level HCC foundations course	Not defined
Teachers	embedded philosophy lecturers	faculty instructors, from philosophy, political science, and CS	primary instructors and guest lecturers from a variety of disciplines	not addressed
Issues covered	privacy and intellectual property, fairness and discrimination, fake news and social media platforms, diversity and equal opportunity, verifyability and interpretability, social responsibility and moral decision-making	algorithmic decision-making, data privacy and civil liberties, AI and autonomous systems, and the power of platform companies	design, development, and deployment of technologies consequences , assessment and planning, skills development of socially conscious and responsible engineers	ethics in ML cases (dataset appropriateness, data bias, training set imbalances, data aggregation, case specific laws , model suitability, multiple model application, results misinterpretation, business ethics)

Instruction methods	lectures, group discussion, simulation, assignments	lectures, small-group discussions, panels of guest speakers, custom case studies, code manipulation, stakeholder interviews	lectures and workshops, combined lecture content with practical exercises and/or a course related group project	modules adapted to assignments or lecture of common ML subjects (logistic regression, random forest classifier, and deep convolutional neural networks), thought experiments, ethical issues questions identified in the authors' SLR

Table 6. Evaluation of pilot programs

Bibliographic reference	(Grosz et al. 2019)	(Reich et al. 2020)	(Skirpan et al. 2018)
Students' perception	positive	positive	positive
Faculty perception	positive	not applicable	not applicable
Challenges	Cross -disciplinary nature :a) insecurities of the faculty towards the other disciplines, d) the use of different methodologies and vocabularies in each discipline and c) the lack of philosophers willing to teach and develop modules for CS classes d) Institutional (funding, administration support)	Stemming from multidisciplinary nature: technical assignment not too challenging for CS students, not technical one considered vague with respect to their evaluation	Some of the assignments were too technically complex for an undergraduate course, some of the technical counterparts of the social issues mentioned (e.g., diving deep into machine learning or cryptography) could not be fully addressed due to time and content scoping constraints
Open questions/future goals	determine long term effectiveness	provide greater clarity on assignment expectations/divers assignment formats to students from different backgrounds	ethical thinking should be introduced early in CS programs, be continuous and embedded into the regular courses, employing relative case studies and real world problems

https://www.nfdi.de/fair-data-spaces/

#FAIRDataSpaces @FAIRDataSpaces

https://www.nfdi.de/fair-data-spaces-newsletter/