FAIR Data Spaces

ELSA Training Curriculum for Data Scientists

Version 1.0

Maria Christoforaki

31.8.2023

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

SPONSORED BY THE

Federal Ministry of Education and Research

Contents

Zusammenfassung

Das Projekt FAIR-Data Spaces konzentriert sich auf den Aufbau von auffindbaren, zugänglichen, interoperablen und wiederverwendbaren cloudbasierten Datenräumen in Zusammenarbeit von Wissenschaft und Wirtschaft, indem es Gaia-X und NFDI zusammenbringt.

Der Aufbau eines solchen Datenraums wäre jedoch unzureichend, wenn sich die Daten Wissenschaftler/innen, die ihn aufbauen, nicht über die ethischen, rechtlichen und gesellschaftlichen Aspekte (Ethical Legal and Societal Aspects-ELSA) ihrer Arbeit im Klaren wären. Daher gibt es neben den Anwendungssoftware-Demonstratoren des FAIR-Data Spaces Projekts noch ein weiteres Vorhaben: die Entwicklung eines ELSA-Lehrplans für Datenwissenschaftler/innen im Rahmen des AP 4.5 ELSA Training für Data Scientists.

In diesem Bericht stellen wir die erste Version des oben genannten Curriculums vor und zeigen, wie die Software Demonstratoren, die im Rahmen des Projekts als FAIR-Data Spaces-Demonstratoren entwickelt werden, als Anwendungsfälle für das Unterrichtsprogramm eingesetzt werden können, das auf dem Curriculum basieren wird.

Um das Curriculum zu entwickeln, folgten wir dem in (Taba 1962) beschriebenen Prozess: Diagnose des Bedarfs, Formulierung der Lernziele, Auswahl der Inhalte, Organisation der Inhalte

Auswahl der Lernerfahrungen, Organisation der Lernerfahrungen, Festlegung der zu bewertenden Inhalte und der Mittel und Wege, dies zu tun.

Der Bericht ist in drei Sektionen gegliedert: die Beschreibung des Curriculums mit Untersektionen, die den oben beschriebenen Schritten entsprechen; die Präsentation der FAIR-Data Spaces-Demonstratoren als Use Cases für das Curriculum; Schlussfolgerungen und indikatives Schulungsmaterial als Anhang.

Wir begannen mit der Erstellung des Curriculums, indem wir eine Bedarfsdiagnose erstellten; dies geschah im ersten Projektjahr durch eine Landschaftsbeschreibung, die auf einer Literaturrecherche und einer Reihe von Workshops in der Gemeinde basierte.

Die wichtigsten Erkenntnisse aus der Bedarfdiagnose waren:

- Das universitäre Angebot an ELSA-Kursen ist nicht ausreichend
- Datenwissenschaftler/innen sind der Meinung, dass die sozialen Auswirkungen von Verzerrungen in Daten und Modellen und die Beeinträchtigung der Privatsphäre des Einzelnen die größten Probleme der Datenwissenschaft sind.
- - Was das Profil der Datenwissenschaftler/innen angeht, so sind die meisten von ihnen männlich, relativ jung und haben einen akademischen Abschluss
- Es gibt jedoch auch eine beträchtliche Anzahl von Personen, die keinen Abschluss haben, und diese Zahl steigt mit den Jahren, da es andere Möglichkeiten gibt, technische Fähigkeiten und Erfahrungen zu erwerben (z. B. Online-Kurse), und ein Abschluss keine Voraussetzung für den Einstieg in die Datenwissenschaft ist.

Anhand dieser Erkenntnisse wurden die Zielfächer für den Lehrplan festgelegt und seine Vision und Lernziele bestimmt, nämlich dass Datenwissenschaftler/innen in der Lage sein sollten:

1. Ethische, rechtliche und gesellschaftliche Aspekte ihrer Arbeit zu erkennen (Awareness).
2. Eine gemeinsame Sprache mit den zuständigen Fachleuten zu sprechen, um gemeinsam nach geeigneten Lösungen zu suchen (Kommunikationsfähigkeit).
3. ELSA in den Data-Science-Workflow einzubinden und nicht als Hindernis oder überflüssiges Artefakt zu betrachten, sondern als integralen Bestandteil des Data-Science-Projektlebenszyklus (Aufbau einer professionellen Mentalität).

Der letzte Punkt führte zur Verwendung des CRISP-DM-Modells als Grundlage für den Aufbau des Lehrplans. CRoss - Industry Standard Process for Data Mining - ist ein nicht-proprietäres, frei verfügbares Data-Mining-Modell, das unabhängig von einem bestimmten Tool oder einer bestimmten Anwendung ist. Sein Ziel ist es, Best Practices zu fördern und Organisationen die Struktur zu bieten, die sie für die Umsetzung von Data-Mining-Projekten benötigen. CRISP-DM gliedert den Data-Mining-Prozess in sechs Phasen: Geschäftsverständnis, Datenverständnis, Datenaufbereitung, Modellierung, Bewertung und Einsatz (business understanding, data understanding, data preparation, modelling, evaluation, and deployment) (Shearer 2000)

Die Gründe, die für diese Wahl sprechen, können wie folgt genannt werden:

1. Es ist bekannt und wird häufig in Data-Science-Projekten verwendet.
2. Es wurde bereits verwendet, um ethische und rechtliche Fragen zu den verschiedenen Phasen des Modells darzustellen und Rahmenwerke zu entwickeln, die die Anwendung ethischer Standards bei der Entwicklung von Data-Science-Projekten gewährleisten.
3. Für ein Projekt wie FAIR Data Spaces, das auf die Zusammenarbeit zwischen Forschung und Industrie abzielt, ist die Tatsache, dass es hauptsächlich von der Industrie entwickelt wurde, ebenfalls relevant

Die Struktur des Curriculums hat eine vertikale und eine horizontale Unterteilung. Die vertikale Gliederung besteht aus Modulen, die den CRISP-DM-Phasen entsprechen, während die horizontale Gliederung aus Wissensteilen besteht, die zu drei phasenübergreifenden Bereichen gehören, nämlich ethische und gesellschaftliche, rechtliche und technische Wissensteilen.

Ethische und gesellschaftliche Wissensteile:

Ethische und gesellschaftliche Aspekte der Datenwissenschaft reichen von der Identifizierung der Stakeholder als jede Gruppe oder Einzelperson, die sich auf das Erreichen der Lernziele auswirkt oder davon betroffen ist, über die Einbeziehung von Gemeinschaftswerten in die eigene Arbeit, die Vermeidung von Diskriminierung von Einzelpersonen oder Gruppen (insbesondere von gefährdeten Gruppen), ethisches Dumping, die Berücksichtigung der Umweltauswirkungen der datenwissenschaftlichen Anwendungen bis hin zur Übernahme von Verantwortung und Rechenschaftspflicht für die eigenen Entscheidungen und Handlungen.

Rechtliche Wissensteile:

Das Lernziel dieser Einheiten ist es, Datenwissenschaftler/innen bei der Bewältigung rechtlicher Probleme zu unterstützen, mit denen sie bei ihrer Arbeit konfrontiert werden könnten. Wir zielen nicht darauf ab, sie in die Lage zu versetzen, diese Probleme selbst zu lösen, eine Aufgabe, die bereits hochspezialisiertes Wissen von Domänenexperten/innen erfordert.

Das Lernziel in Bezug auf Datenwissenschaftler ist ein dreifaches:

1. sie für die rechtlichen Fragen zu sensibilisieren, mit denen sie im Laufe ihrer Arbeit höchstwahrscheinlich konfrontiert werden

2. ihnen einen Überblick über die grundlegenden rechtlichen Konzepte zu geben, die sie zum Verständnis der oben genannten Fragen benötigen

3. ihnen das Vokabular an die Hand zu geben, das sie benötigen, um effektiv mit den Fachleuten (z.B. Rechtswissenschaftlern/innen) zu kommunizieren, mit denen sie bei der Lösung dieser Probleme zusammenarbeiten müssen.

Technischer Wissensteile:

Diese Einheiten befassen sich mit der technischen Umsetzung rechtlicher oder ethischer Desiderate wie dem Datenschutz, der Erkennung von Verzerrungen durch Daten und Algorithmen und Strategien zur Milderung dieser Verzerrungen, der Integration von Fairness, Effektivität und Erklärbarkeit in die Bewertung eines datenwissenschaftlichen Projekts, der Bereitstellung und Überwachung außerhalb von Experimentier- und Testumgebungen. Das Curriculum wird keine spezifische technische Schulung zu diesen Themen anbieten, sondern vielmehr anhand von Anwendungsfällen veranschaulichen, wie technische Lösungen eingesetzt werden können.

Die Gründe dafür sind unter anderem die Tatsache, dass jeder Anwendungsbereich spezifische Techniken erfordert und dass Themen wie Sicherheit, Datenverschlüsselung, Datenintegrität, Authentifizierung, gegnerische Angriffe usw. in der Regel in speziellen Kursen während des Studiums oder in der Praxis der Datenwissenschaft behandelt werden.

Bei der Umsetzung des Curriculums stellt sich die Frage nach der Dauer und der Form des Programms. Insbesondere die Frage, wie viel Zeit für jedes Modul zur Verfügung steht und ob dies entweder über einen längeren Zeitraum (z.B. einige Stunden pro Woche) oder in einer intensiven einwöchigen Sommerschule, in Form einer Reihe von Workshops usw. geschehen kann.

Dies hängt auch mit dem Wissensstand und den Bedürfnissen der Zielgruppe zusammen. Das Curriculum richtet sich in erster Linie an Anfänge unter den Datenwissenschaftlern, die sozusagen aus der Vogelperspektive einen Einblick in einige der ELSA-Herausforderungen erhalten, denen sie bei ihrer Arbeit begegnen könnten. In diesem Fall könnte das Programm die Wissenslücke der fehlenden Universitätskurse ausfüllen.

Es gibt jedoch noch eine andere Möglichkeit, das Curriculum zu implementieren: mit dem Lernziel, es an die berufliche Rolle der Teilnehmer/innen oder an die speziellen Anforderungen des Anwendungsbereichs anzupassen, oder beides.

Dies kann durch die Erweiterung spezifischer Module geschehen, die sich im ersten Fall an Projektmanager oder Entwickler richten, im zweiten Fall durch die Anpassung des Inhalts an bestimmte Anwendungsbereiche, z.B. NLP, Image Processing oder das Gesundheitsbereich.

Was die Unterrichtstechniken betrifft, so können diese von Vorlesungen, eingeladenen Vorträgen und Case Studies bis hin zu kooperativen Lerntechniken wie Studentenarbeitsgruppen und Rollenspielen oder praktischen Übungen reichen. Unsere Literaturrecherche hat jedoch ergeben, dass die effektivsten Mittel zur Vermittlung von Inhalten praktische Übungen und die Verwendung von Fallstudien sind.

Die Teilnehmenden können anhand ihres Engagements in den verschiedenen Modulen bewertet werden, während die Programmbewertung durch Rückmeldungen sowohl der Teilnehmenden als auch der Ausbilder darüber erfolgen kann, ob das Programm ihre Erwartungen erfüllt hat oder nicht. In diesem Sinne ist der beste Weg zur Evaluierung des Curriculums die Durchführung eines Pilotunterrichtsprogramms, gefolgt von einer detaillierten Evaluierung anhand von Bewertungsbögen der Teilnehmenden und TutorInnen oder von Interviews. Dies liegt

jedoch nicht im Rahmen des FAIR Data Spaces Projekts.

Die FAIR-Data Spaces Demonstratoren als Use Cases für das Curriculum

Im Rahmen des FAIR Data Spaces Projekts werden drei Demonstratoren entwickelt: a. für einen Datenraum für Biodiversität, b. Qualitätssicherung von Forschungsdaten und c. plattformübergreifende Datenanalyse. Diese drei Demonstratoren werden als Use Cases für das ELSA Curriculum vorgeschlagen. Zu diesem Zweck haben wir Beiträge aus dem AP2-Arbeitspaket "Rechtliche und ethische Rahmenbedingungen" zur rechtlichen und ethischen Überprüfung der Demonstratoren verwendet.

In Sektion 3 finden Sie eine kurze technische Beschreibung jedes Demonstrators mit einem Anwendungsszenario und den ethischen und rechtlichen Fragen, die in jedem dieser Demonstratoren zum Ausdruck kommen. Dazu gehören: der Schutz persönlicher Daten, der Schutz des Urheberrechts und des geistigen Eigentums, Fragen der Voreingenommenheit und der Diskriminierung sowie Fragen, die sich aus dem Datenaustausch zwischen Industrie und Wissenschaft ergeben, ein Thema, das für das FAIR Data Spaces Projekt von besonderer Bedeutung ist.

Die zweite und finale Version des ELSA-Curriculums für Data Scientist wird als Bericht vorgelegt, nachdem wir das Feedback unserer Gaia-X-Industriepartner (und der gesamten Community) zur Nachhaltigkeit des Curriculums eingeholt haben. In dieser Version wird der Einsatz der FAIR Data Spaces-Demonstratoren als Anwendungsfälle aktualisiert werden, da mehr Input von den AP2-Projektteilnehmern in Form von Projektleistungen zur Verfügung gestellt wird.

1 Introduction

The FAIR-Data Spaces project focuses on building findable, accessible, interoperable, reusable cloud-based Data Spaces with the cooperation of science and business by bringing together two initiatives, namely Gaia-X and NFDI.

The construction of such a data space could not be adequate, if the data scientists who are actually building it would not be aware of the Ethical, Legal and Societal Aspects (ELSA) of their work. Thus, in addition to the FAIR-Data Spaces project application software demonstrators, there is also another one, namely the development of an ELSA Curriculum for data scientists.

In this report we present the first version of the above curriculum and how the software applications that are developed as FAIR-Data Spaces demonstrators within the project can be deployed as use cases for the instruction program that will be based on the Curriculum.

In order to develop the curriculum development, we followed the process described in (Taba 1962):

Step 1: Diagnosis of need

Step 2: Formulation of objectives

Step 3: Selection of content

Step 4: Organization of content

Step 5: Selection of learning experiences

Step 6: Organization of learning experiences

Step 7: Determination of what to evaluate and of the ways and means of doing it.

The rest of the report is structured as follows: in section 2 we describe the need diagnosis which was performed during the first year of the project via landscape description based on literature review (described in the project deliverable "E4.5.1: Bewertung der bestehenden Ansätze"), and a series of community Workshops (cf. "E4.5.2: Workshop-Bericht");then we go on to describe the curriculum vision and objectives, the selection and organisation of the contents, means of content delivery, and, finally, an evaluation approach. In section 3 we describe the FAIR-Data Spaces as Use Cases for the curriculum: first we offer a technical description for each demonstrator and then the ELSA issues that can be underlined in each case.

Section 4 closes the report with conclusions and we also supply some indicative training material as an Appendix.

2 ELSA Curriculum

2.1 Need Diagnosis

In the context of the FAIRData Spaces project we have conducted a review of the existing landscape as described in the project deliverable "E4.5.1: Bewertung der bestehenden Ansätze", also published as a report in the FAIR-DS Zenodo Community (Christoforaki 2021). There we reviewed the existing data science ethics courses in tertiary education and identified three major topic categories: computer science, professional and business skills, and ethical and legal topics.

Additionally, in the Workshop Series reported in the project deliverable "E4.5.2: Workshop-Bericht", we broadly sketched the profile of the data scientist based on the ideal one as described in the bibliography and the realistic one as painted by surveys (Christoforaki 2022) .

The main takeaways from both deliverables with respect to the need diagnosis were:

- The review of the university study offers in the US, Europe as well as all over the world, revealed that few tertiary education institutes offer these kinds of courses (numbers vary between 15% and 34%, depending on the survey).
- When the data science professionals were asked to point out which is the biggest problem in Data science/machine learning, the respondents identified firstly the social impacts caused by bias in data and models, followed by impacts to individual privacy.
- Regarding the demographics, the surveys indicate that:
    - most of the data scientists are men (around 80%) and relatively young (aged 24-40)
    - over 2/3 have a higher degree
    - there is a significant number that has no degree and this number is increasing through the years, since there are other ways to build technical skills and experience (eg., online courses), and having a degree is not a prerequisite for getting started in data science.
- The industry does not require that all data scientists have a relevant degree, a trend that is manifested by the influx of people transitioning from other disciplines to data science, as documented in the World Economic Forum of 2020 Future of Jobs report.

As a result, regarding the profile of the data scientist we have to take into consideration that:

1. There are two kinds of data scientists: one with experience and/or background studies in maths, statistics and visualisation techniques, and a second one, with computer science and programming skills, acquired either by working experience and or by studies in computer science/engineering. However, a more holistic approach may include people with social science research background knowledgeable in methods and possessing skills used in raising the appropriate questions and hypotheses, as well as ones with soft skills associated with communication and teamwork or having some domain knowledge and business and strategy competences.
2. In the future, data scientists will not form a homogeneous group, such as being university graduates, which is the target group of today's ethics curricula. They will also come from a variety of educational, social and national backgrounds, and quite a lot of them will land in data science following quasi-academic routes, especially when being employed in job positions with less demanding high-end knowledge or education.

These two points guided the target subjects for the Curriculum, since it will have to face challenges that emanate from different educational needs.

The Workshops also provided us with feedback regarding the curriculum content which was used in the preparation of this document.

2.2 Curriculum Objectives and Vision

The basic objectives of the ELSA Curriculum are that data scientists should be able to:

1. Recognise 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).

2.3 General topics outline

In order to achieve the above, 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. Specifically, we propose three main area topics:

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 (especially vulnerable ones), taking into consideration the environmental impact of the data science applications, and assuming responsibility and being accountable for one's decisions and actions.

Business ethics, namely Professional ethics codes and codes of conduct, responsibility and accountability, leadership.

We devise the content within the framework of the four + 1 ethical principles introduced by (Floridi et al. 2018): the four bioethics principles, namely, Autonomy, Non-maleficence, Beneficence and Justice, and Transparency and their manifestation in data science projects.

The bioethics principles as presented in (Beauchamp and Childress 2019) are rephrased and augmented by transparency in (Floridi et al. 2018) as follows:

Autonomy in bioethics is connected liberty as independence from controlling influences, as agency as capacity for intentional action, so that an individual can act freely according to their own plan. Regarding Data Science application the focus is on the balance between decision making automation (what kind of decision can and should be made by either artificial agents or humans).

Nonmaleficence in bioethics is the principle of not inflicting, preventing or removing harm, while in data science can be seen as the prevention of accidental or deliberate harms.

Beneficence requires that not only we refrain from harming people but that we contribute to their welfare and regarding Data Science application that might also include promoting their well-being, preserving their dignity, and sustaining the planet.

Justice can be interpreted in bioethics as fair, equitable, and appropriate treatment. Analogously, Data Science applications should promote prosperity and preserving solidarity by trying to correct past wrongs such as eliminating unfair discrimination; ensuring that the created benefits are shared or at least shareable and preventing the creation of new harms, such as the undermining of existing social structures

Finally, transparency via explicability can make sure that the decision-making processing can be understood and held accountable.

The original Bioethics principles and their Data Science counterparts are presented in Table 1

Principle	Bioethics (Beauchamp and Childress 2019)	AI/DS (Floridi et al. 2018)
Autonomy	• act freely in accordance with a self-chosen plan • liberty • agency	• contain the risk of delegating too much to machines • humans should always retain the power to decide which decisions to take
Non-maleficence	• not inflict evil or harm • prevent evil or harm • remove evil or harm	• prevent harms arising, whether from the intent of humans or the unpredicted behaviour of machines
Beneficence	• actively promote good	• promote well-being • preserve dignity • sustain the planet
Justice	• fair, equitable, and appropriate treatment	• Prevent/eliminate discrimination • equally shared benefits
Explicability		• understand and hold accountable the decision-making processes of AI

Table 1. The principles of Bioethics and their AI renditions

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. The objective is not to make them able to solve these problems, which already demands highly specialised knowledge by domain experts, but to:

a. Make them aware of the legal issues they will most probably face during the course of their work

b. Give them an overview of the basic legal concepts they need to understand the above-mentioned issues and

c. provide them with the vocabulary that they need in order to communicate effectively with the domain experts (such as legal scholars) who will be the ones that will provide the solution in each specific situation.

The main areas that will be covered are Data Protection (especially as demanded by GDPR) and Intellectual Property issues that relate to Data Science (e.g. the use of training data) as well as Basic legal terminology and concepts.

Technical Renderings Knowledge Units: While technical solutions can be employed to comply with a wide variety of either ethical or legal demands, these units are not aiming to teach technical skills to data scientists. These skills may be covered in other specific courses during their studies or their practice and concern wider domains, for example, security issues, like encryption, data integrity, authentification, adversarial attacks, etc.

They will rather 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. The Curriculum will not provide specific technical training regarding these issues, since each application domain might require specific techniques, it would rather illustrate the way technical solutions can be employed via use cases.

2.4 Selection and Organisation of Content

The basic premise of the curriculum proposal is that we base the whole training on the CRISP-DM Model as described in (Shearer 2000).

The reasons that support this choice can be stated as follows:

- It is well known and often used in data science projects.
- It has been developed mainly by industry. This is an advantage for a project like FAIR Data Spaces, which aims at collaboration between research and industry
- 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. Specifically, (Saltz and Dewar 2019) conducted a systematic literature review and mapped the main ethical themes that were identified to the various phases of the CRISP-DM model. This paper was used as a basis for this proposal.

In the following subsections we present the CRISP-DM model and approaches to use it for mapping ELSA subjects as presented in the bibliography and were used as an inspiration for the formation of this curriculum

2.4.1 The CRISP-DM Model

The CRISP-DM (CRoss-Industry Standard Process for Data Mining), is a non-proprietary, documented, and freely available industry-, tool-, and application-neutral model data mining model developed by industry leaders with input from more than 200 data mining users and data mining tool and service providers. It encourages best practices and offers organisations the structure needed to realise better and faster results. 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 (Shearer 2000).

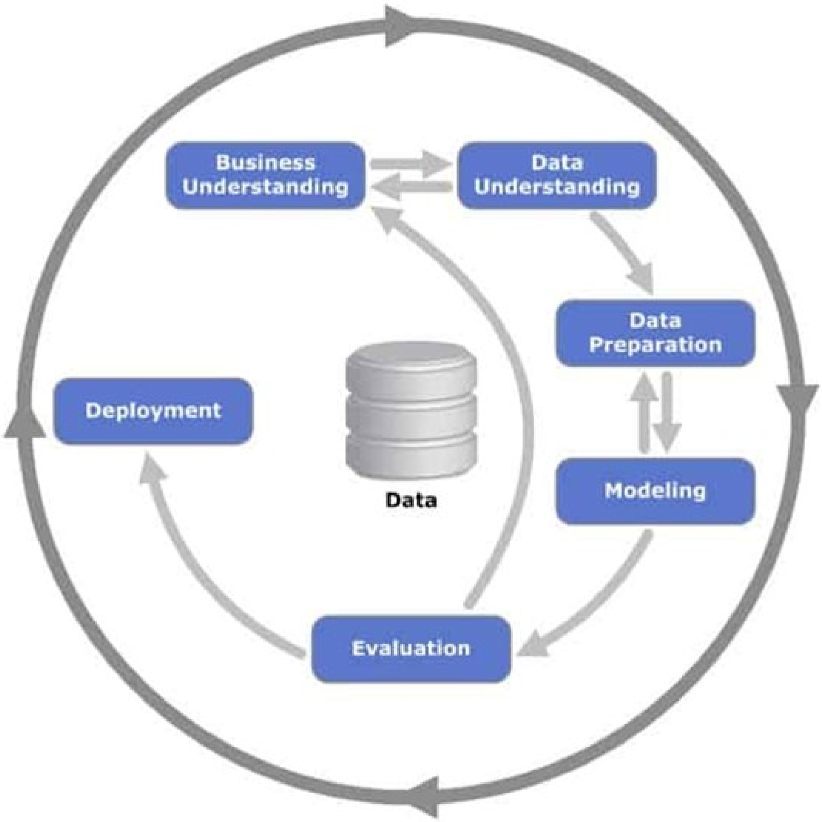

Figure 1. The CRISP-DM Model. Image by Kenneth Jensen CC BY-SA 3.0 https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png

Figure 1 shows the phases of a data mining process. The arrows indicate the most important and frequent interactions and dependencies between the phases, while the outer circle symbolises the cyclical nature of the process itself. Figure 2 shows the steps included in each phase.

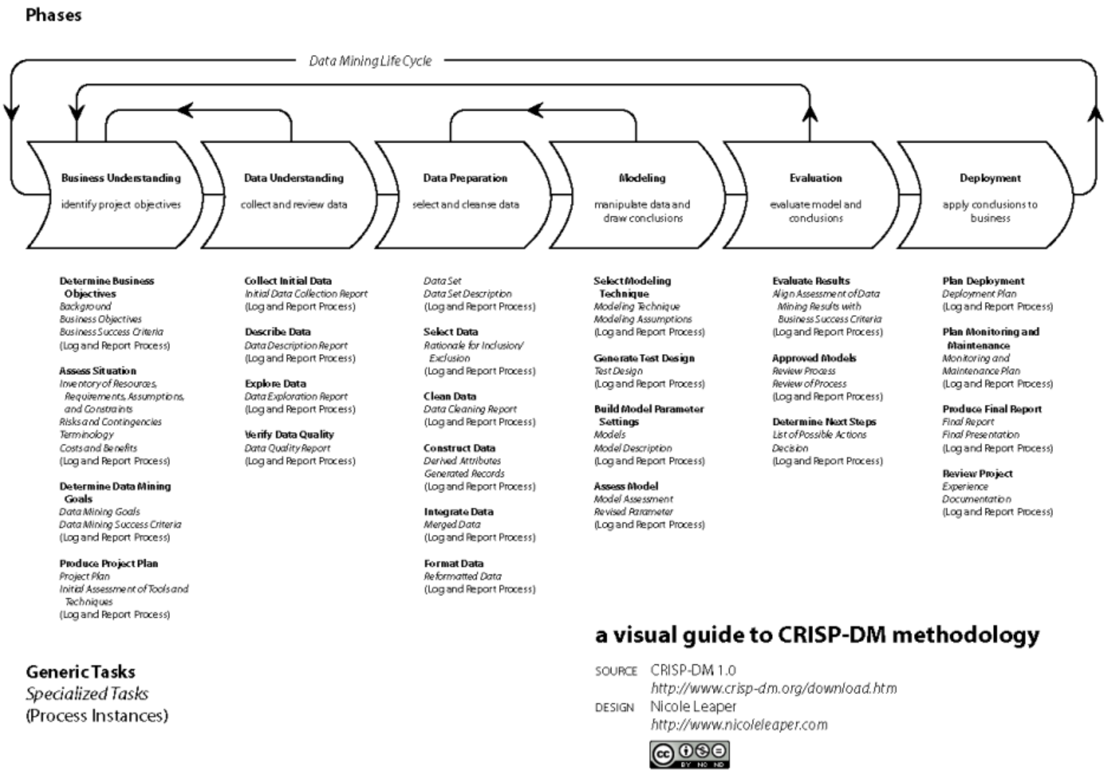

Figure 2. Detailed description of CRISP-DM phases. Image by Nicole Leaper, [CC BY-NC-ND 3.0 US](), https://exde.wordpress.com/2009/03/13/a-visual-guide-to-crisp-dm-methodology/

2.4.2 Mapping data science ethical considerations to the CRISP-DM model

As mentioned above, the basic inspiration for the curriculum concept is (Saltz and Dewar 2019) where the authors perform a systematic literature review in order to map and describe the main ethical themes permeating data science. They go on to identify a possible structure to integrate these themes within a data science project, in order to help data scientists, cope more efficiently with the respective issues.

Similar approaches were followed (among others) by (Morley et al. 2020) who mapped to CRISP-DM existing Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, i.e., moving from *What* to *How,* and the Turing Institute's Guide for the responsible design and implementation of AI systems in the public sector (Leslie 2019). The latter proposes an ethical platform for the responsible delivery of an AI project, focused primarily on CRISP-DM but it can also be applied on other related workflow models. These two papers were also taken into consideration in the Curriculum formation.

(Saltz and Dewar 2019) codified their results in a framework to explore the key ethical considerations by phase of project as presented in Table 2.

While their research identified four key themes, i.e., the need for an ethics framework, the newness of the field, data related challenges and model related challenges, for our curriculum only two of them are important, namely Data and Model challenges.

Data related challenges lead to the following ethical considerations: Privacy and anonymity, Data misuse, Data accuracy and validity.

Model challenges lead to consideration about personal and group harm (e.g., discrimination resulting from algorithmic bias) and subjective model design (e.g. The use of data points as proxies for missing facts) and model misuse and misinterpretation stemming from the statistical nature of predictive models and various degrees of transparency.

Regarding the mapping of the ethical considerations to the CRISP-DM model the authors found no ethical theme related to the business understanding phase, since this phase (focused mainly on accountability), presents no opportune subjects for paper exploring new ethical issues relating to data science. Hence, they propose two ethical new considerations, i.e., two ethical new considerations and team accountability.

Data related challenges map to the data understanding and data preparation phases, and the model related challenges map to the modelling, evaluation and deployment phases.

Based on this initial mapping we are going to structure the ELSA Curriculum along the above defined lines.

Project phase	**Key ethical themes**	**Ethical considerations**
Business understanding	Project initiation/ management challenges	Personal and group harm Team accountability
Data understanding/ data preparation	Data challenges	Data misuse Data privacy & anonymity Data accuracy
Modelling	Model challenges	Personal and group harm
Evaluation		Subjective model design
Deployment		Misuse/misinterpretation

Table 2. Framework to explore the key ethical considerations by phase of project (Saltz and Dewar 2019)

2.4.3 Organization of Content: ELSA Curriculum for Data Scientists Modules

The curriculum structure has a vertical and a horizontal partition.

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 that correspond to the three topics areas detailed in 2.3 General topics outline.

Figure 3. illustrates this vertical and horizontal partitioning.

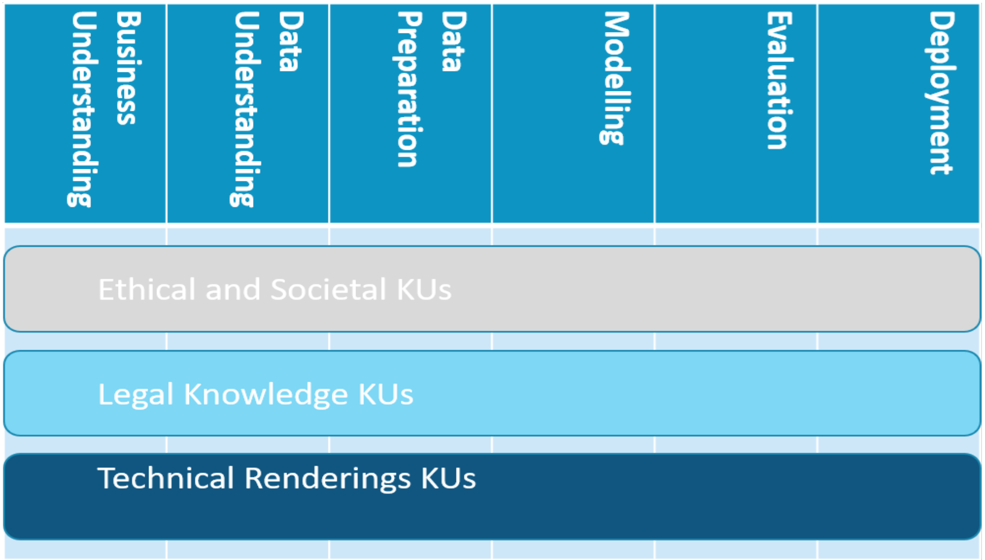

Figure 3. Ethical and Societal, Legal and Technical Renderings Strands. With grey we denote the ethical/societal strand, with light blue the legal strand, with and deep blue the technical renderings of the ELSA challenges.

The format followed for the description of each module is the following

Module number and title: This always corresponds to a CRISP-DM model phase.

Phase description: Detailed description of the CRISP-DM phase as presented in (Shearer 2000) (the CRISP-DM model refers to *Data Mining* we will replace this term with *Data Science)*

Knowledge Unit (KU) number and title

Description: Detailed description of content.

Suggested activities: description of activities.

Module I: Business understanding

Phase description: 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. (Saltz and Dewar 2019) identify the main ethical consideration for this phase as personal and group harm, i.e., adverse impact of the application on individuals and/or groups, especially ones that are already disadvantaged and/or marginalised.

KU I.1: Stakeholder identification

Stakeholders are defined as any group or individual that affects or is affected by the achievement of the business objectives and their interactions with a corporation(Freeman 2010); these, depending on the specific application might include from shareholders and customers to the data scientist employed and the legal and HR departments or an organisation as well as marginalised groups and the physical environment. Students are presented with a Typology of stakeholders in Data Science/AI and explanation for each proposed type.

Suggested activities: The students are given examples of data science problem formulation and are asked to provide the stakeholders according to the presented typologies and the way the application might cause them harm. They are encouraged to think outside the typology if their assessment does not fit in the provided types

KU I.2: Incorporating community values

Data science as a profession lacks a specific certification that comes with the respective fiduciary duties and accountability, However, a variety of national and international professional associations have issued professional codes that foster responsibility and advocate principles and values that can enable data scientists to see themselves as a part of a community with specific expectations. In Germany, these include the Gesellschaft Für Informatik and the KI Bundesverband. Additionally, there is the development of the so-called “-by design” guidelines and frameworks: Ethics, Privacy, Data Protection, etc. Students are introduced to these frameworks and guidelines and the respective duties they impose.

Suggested activities: The students are given problems with ethical challenges and asked what their actions do before they are familiarised with the professional codes and then asked the same questions afterwards. They discuss whether being aware of the codes made any difference in their decisions and whether voluntary commitment to them is enough or other measures must be taken

KU I.3: Organisational culture

Organisational culture is an organisation’s systems, procedures, and practices for guiding and supporting ethical behaviour that clearly communicates the range of acceptable and unacceptable behaviours through leader role-modelling, reward systems, and rigorous code enforcement, and is associated with fewer unethical decisions. Organisations that foster a climate that encourages employees following rules that protect the company and others when coping with ethical challenges, are shown to be associated with fewer unethical decisions.

Suggested activities: Case studies of companies where poor (or debatable) organisation culture has provoked controversial ethical decisions.

KU I.4: Basic concepts

In order to establish a communication channel with other domain experts, such as legal scholars, the data scientists should be familiar with a number of basic concepts used by the aforementioned experts. These include 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, etc.

Suggested activities: Case studies of data science applications where the manifestations of the above concepts are displayed.

Module II: Data understanding

This phase concerns data collection, exploration and quality verification, gaining insights and identifying possible issues. (Saltz and Dewar 2019) identify the same issues for this and the following phase (i.e., Data preprocessing), which they collectively name as *data challenges* and specify as Data misuse, privacy & anonymity, and accuracy. While this might be adequate for identifying the general issues, we follow a more detailed approach by creating two different modules. The reason for that is that there are a number of challenges that are to be faced before any preprocessing takes place and that has to be underlined in the curriculum. Additionally, this phase can alter the previous one of business understanding, since data understanding can alter the initial problem definition. A third reason is that 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. This is relevant in the implementation of the curriculum, as it is stated in the section Implementation Strategies, as modules can be adapted to the target audience.

That said, a variety of issues are indeed common in both phases and can be addressed in either or both of them, as time and purpose constraints vary.

KU II.1: Data Protection

Data protection content is framed in the GDPR context; however, the goal is not to create a GDPR tutorial but to create familiarisation with the basic concepts and processes that relate to it. Additionally, data protection issues span through all the CRISP-DM phases and the content may be overlapping or shifting from one phase into another depending on the implementation requirements.

In this phase are presented basic concepts and issues that have to do with data collection, such as what is personal and sensitive data, data subject and their rights, data controller and data processor, the role of the DPO and legal grounds for processing of personal data. Additionally, data protection goals based on the GDPR requirements are introduced-a suggested list linked to the SDM framework (cf. below in *Suggested activities*): Data minimisation, Availability, Integrity, Confidentiality, Unlinkability, Transparency and Intervenability. International data transfer is also covered, especially in conjunction with KU II.4: Ethics dumping in data collection and KU III.4 ethics dumping in data preprocessing

Suggested activities: Illustration of the application of Data Protection in use cases via a framework, e.g. The Standard Data Protection Model(Unabhängiges Landeszentrum für Datenschutz 2020b; 2020a). The suggested framework can be used throughout the modules illustrating appropriately chosen use cases. Reflection on how the data protection challenges can change the problem definition and/or objectives as formed in the previous (*Business Understanding*) phase

KU II.2: Data bias during collection and ways to mitigate it

Bias is well documented in computer systems and specifically in Data Science applications has been in the spotlight from the very beginning, bringing up issues of justice, fairness and discrimination, especially of sensitive and marginalised groups. Bias can be found in all the CRISP-DM phases and in this KU special focus is given to historical and representation bias that are relevant in the data understanding phase, and more specifically in data collection. Ways of detecting and mitigating data bias in this phase are also presented.

The use of synthetic data can also be investigated as a means to mitigate bias. Synthetic data creation and use can also be addressed in KUs about Ethics Dumping

Suggested activities: Use Cases of historical and representation bias in data collection are presented, for example in well known, publicly available image datasets. Practical exercises can be included where publicly available datasets and open source software that detects and mitigates bias in them are given to the students to experiment. The students are encouraged to reflect on how bias in data collection and bias mitigation might affect the other phases of application development, e.g., evaluation (regarding accuracy of results), and deployment, as well as business understanding (problem definition and objectives). Also, they might examine as an alternative use cases of successful and unsuccessful use of synthetic data

KU II.3: Intellectual property issues and licences

This KU concentrates on the explanation of intellectual property challenges that govern data collection and understanding, specifically the licences that data come with as well as the permissible ways to collect material, e.g. by scraping the internet.

Suggested activities: Students are presented with a use case and asked to identify possible IP problems; students are asked to scrape data from the internet taking into consideration robots.txt files.

KU II.4: Ethics dumping in data collection

Ethics dumping is an issue found not only in data science but in other scientific domains as well. It can be defined as “the malpractice of (a) exporting research activities about digital processes, products, services, or other solutions, in other contexts or places (e.g. by European organisations outside the EU) in ways that would be ethically unacceptable in the context or place of origin and (b) importing the outcomes of such unethical research activities”(Floridi 2019). In our case, collecting data from countries that do not have the same safeguards regarding, for example, privacy or legally acquiring already existing datasets from such countries.

Synthetic Data can be seen as a method to overcome ethics dumping as well as challenges examined in the previous KUs (privacy, bias, IP). However, they come with their own issues of how appropriate they are for the specific problem and how they will be used in real world applications.

Suggested activities: Students are asked to reflect on the use of datasets obtained via ethics dumping and whether the practice, even if it is legal, means the continuing exploitation of vulnerable people not only in the Global South, vs jeopardising their project because the inability to obtain similar data from their own more data protecting region (country or the EU).

KU II.5: Dataset documentation

Dataset documentation is very important, not only for identifying data provenance, data quality and validity, IP licences but also whether there is detected bias in the dataset and if some measure is taken in order to mitigate it. The students are presented with a number of data documentation suggestions, e.g. Datasheets for data sets, dataset nutrition labels, etc., and how to use them. Dataset documentation expands in the next phase as well.

Suggested activities: Apply some selected data documentation framework on a dataset and students are asked to reflect on the merits of the process vs the overhead.

Module III: Data preparation

This phase aims to the construction of the final dataset that will be used by the model. It consists of selection of the actual data that will be used for the analysis task depending on the relevance of the project goals, quality and technical constraints; cleaning; construction integration and formatting of data. As noted in Module II: Data understanding (Saltz and Dewar 2019) identify the same *data challenges* for both phases, namely, data misuse, privacy & anonymity, and accuracy. This is reflected also in the Curriculum, since some of the KUs have the same titles; however, we try to differentiate between the relevant issues in either phase. The way that the continuity of the two phases could be implemented is discussed in 2.5.1 Implementation Strategies

KU III.1: Data Protection

Further elaboration on the practical methods that the data protection goals described in KU II.1: Data Protection. These two KUs can be covered as one continuous unit starting with the basic definitions and concept introduction in Module II: Data understanding and moving to more concrete examples in the present KU.

Suggested activities: Illustration of the application of Data Protection in use cases via a framework, e.g. The Standard Data Protection Model(Unabhängiges Landeszentrum für Datenschutz 2020b; 2020a)

KU III.2: Data challenges in preprocessing

Depending on the project, data preprocessing may include a variety of challenges, in 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 scholars annotation, error induced by automatic machine learning annotation, in the creation of synthetic data depending on the problem Outliers are sometimes relevant for data analysis). Solutions like bias identification and mitigation tools as well as documentation practices might be presented here as a continuation of the KU II.2: Data bias during collection and ways to mitigate it Ethics dumping is also an issue in this phase (annotation)

Suggested activities: Use cases of relevant bias can be presented and studied here, as well as Practical exercises can be included where publicly available datasets and open source software that detects and mitigates bias in them are given to the students to experiment

KU III.3: Intellectual property issues of training data

Continuation of KU II.3: Intellectual property issues and licences. These two KUs can be presented in both modules or in condensed form in either one of them.

This KU focuses on the different kinds of training data according to their licences when specifically considering generative AI issues: using copyrighted works as training dataset, legally storing them for training process, copyright on the generated data (maybe later on deployment), especially if we are talking about code. Provision of EU law on text and data mining (TDM) exception.

Suggested activities: Students are presented with a use case and asked to identify possible IP problems

KU III.4: Ethics dumping in data preprocessing

Ethics dumping in data preprocessing may comprise annotation and preprocessing activities outsourced to countries with lower standards in workers protections laws. This may include countries in the Global South but not exclusively there. Ethics dumping can be considered in a broader sense here, as for example using unqualified people to annotate data instead of experts because of lower costs and thus also related to the societal issues of worker exploitation and gig workers.

This unit could be taught in conjunction with the respective one in the previous module. The issue of synthetic data can also be addressed here, since in their capacity to be produced from existing data, can be regarded as a form of data preprocessing.

Suggested activities: Continue the previous module activity in the preprocessing phase. Reflect how the data quality is affected by ethics dumping practices.

KU III.5: Dataset documentation

This KU is an extension of *KU II.5* for data preprocessing. These two KUs can be presented in both modules or in condensed form in either one of them. The students are presented with a number of data documentation suggestions, e.g. Datasheets for data sets, dataset nutrition labels, etc., and how to use them in order to document the preprocessing that has be done to the dataset, regarding annotation, cleaning, bias detection and mitigation, etc

Suggested activities: Continue the previous module activity in the preprocessing phase

Module IV: Modelling

This phase concerns the selection and parameter calibration of models according to the specific problem definition and specific data requirements. (Saltz and Dewar 2019) identify the ethical issues mainly around personal and group harm, specifically bias and discrimination. The curriculum addresses other issues as well in the following KUs:

KU IV.1: Model bias and mitigation techniques

Model bias can be identified in various forms, such as learning and aggregation bias as well as bias that can be found in pertained models and then transferred to the final trained model, which can lead to unfair outcomes and discrimination. The identification and mitigation of such kinds of bias are presented in this KU.

Suggested activities: Use cases of model bias can be presented and studied here, as well as practical exercises can be included where publicly available datasets and open source software that detects and mitigates bias in them are given to the students to experiment.

KU IV.2: Model transparency and explicability

Model transparency and explicability are covered in this KU not as technical issues, although a brief presentation of how well-known models are positioned regarding these criteria might be advisable. The main focus is given on the ethical and legal (e.g. “right to explanation”) requirements of using transparent models. Special focus is put in the cases of automated decision-making applications, where factors like automation bias are crucial in discrimination of sensitive groups.

Suggested activities: The students are presented with use cases, where opaque algorithms caused disparate negative impact. Practical exercises can be given when the same problem can be addressed with a number of different algorithms of various transparency.

KU IV.3: Environmental impact of model training

Model training may demand large amounts of energy. Depending the way that this energy is produced this may have grave environmental implications (e.g. increased carbon footprint). A variety of methods that calculate the environmental impact of a model as well as the factors that determine their energy consumption.

Environmental impact issues can also be examined under the ethics dumping umbrella, since the model training can be done in countries with lower environmental standards and then import the outcomes (trained model). In that sense this KU can be taught in conjunction with the respective ones in modules II and III.

Suggested activities: The students are presented with examples of models and their energy consumption. Students are encouraged to reflect on the trade-offs of environmental impact and application goals.

KU IV.4: Intellectual property issues

The use of pretrained models, commercial secrets, property of the trained model and the outcome of the algorithm. This KU concerns itself with intellectual properties issues that pertain to the algorithm itself and its products. In that case it can be linked to the deployment phase, corresponding to the respective module. The content encompasses the latest advances in LLM and generative AI models, an area that is (at the time of writing of the present document), still not well developed. However, it contains important legal (and ethical) issues and therefore is included in the curriculum.

Suggested activities: The students are given case studies and they are asked to reflect about the limitations of IP regulation on product development and in broader subjects as in artistic creation.

KU IV.5: Model documentation

In correspondence with the dataset documentation, there are model documentation frameworks as well. Documenting the model is extremely important, especially when issues like bias and transparency are concerned. Model documentation frameworks (e.g. model cards) are presented

Suggested activities: Students are asked to employ a model documentation framework for a made-up use case.

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. (Saltz and Dewar 2019) identify as ethical considerations subjective model design, i.e., various kinds of bias in the AI/ML pipeline and the inherent biases of the data scientists themselves resulting in personal and group harm issues are to be considered due to algorithmic discrimination. In this Module are addressed the issues of the trade-off between algorithmic accuracy and a variety of other important factors, such as resource efficiency and trustworthiness. Specific focus is given to the concept of fairness which is addressed in a KU of its own.

KU V.1: Evaluation beyond accuracy

Data science algorithms are usually evaluated with respect to their accuracy only. However, there are a number of other factors that have to be taken into consideration when we evaluate a model: fairness, efficiency (in terms of resource allocation, e.g. energy consumption as covered in KU IV.3: Environmental impact of model training, but also the size of the model and the dataset), explainability (choosing a more transparent model as demonstrated in KU IV.2: Model transparency and explicability at the cost of less accurate results), trustworthiness (as the qualitative measure of confidence one can objectively assign to the output of an AI system, which includes apart from explainability robustness outside the experimental settings), whether the selected features and proxies actually solve the initial problem, although they provide accurate results. As per the CRISP-DM model, the evaluation phase might lead to reconsidering the problem definition (phase 1: Business understanding). This is not only meant as failing to accomplish the project goal but also as redefining the accuracy notion in order to incorporate the above-mentioned issues.

Suggested activities: Students are given use cases and are encouraged to reconsider the project goal with respect to the above described issues. Practical exercises are also possible in conjunction with the next KU dealing with fairness.

KU V.2: Fairness

The bias and discrimination that may result in a data science application are dealt with in the previous modules, where they are examined at specific data science circle phases. However, it is often that the evaluation phase indicates plainly whether there are such issues present, since in the previous phases we act proactively and while checking in advance is needed, it is not always possible to grasp all the implications. The evaluation phase gives us a first tangible result we can use to assess whether there are such issues present (the other being the deployment phase of course).
In this phase are examined the various definitions of fairness and how these are mathematically formulated and executed in code. There is also a description of the legal provisions in EU regarding fairness and the human perceptions of it and how these impacts the trustworthiness of a product, as well as insights from other disciplines (e.g., ethical philosophy) that go beyond technical solutions.

Suggested activities: Practical exercises where the students are asked to apply various fairness definitions in coding sessions and assess the results.

KU V.3: Model documentation

This is a continuation of the previous phase KU with the same name. All evaluation metrics that are used to assess the model and the respective results are documented according to a specific documentation model.

Suggested activities: Students are asked to employ a model documentation framework for a made-up use case.

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 (Saltz and Dewar 2019) 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.

KU VI.1: System deployment limitations

The data scientist must be able to explain to the clients, managers and the project stakeholders (as identified in KU I.1: Stakeholder identification), the system limitations, both with respect to known issues and with what the system is actually able to do or not, and also explain the chosen level of automation and the possible impact that these might have, especially in the case of adverse outcomes.

Suggested activities: The students are given use cases that illustrate individual and group harms that specific systems caused, because they were deployed without the proper understanding of their abilities and limitations. They are encouraged to provide suggestions of how a proper communication to the stakeholders of the application limitations would have averted these harmful consequences.

KU VI.2: Visualisation bias

This specific kind of bias is addressed in this phase, since most of the visualisation is used to convey results to the final system users. Since most data science applications provide results of statistical nature, there are a number of well-known pitfalls that exist when trying to visualise them; however, data scientists are not especially familiar with them as they are usually thought of as UX issues. An introduction in the subject will help data scientists to fruitfully cooperate with UX designers in order to avoid as much as possible misinterpretation of results.

Suggested activities: The students are given use cases where visualisation bias is present and they are asked to reflect on possible ways to avoid it.

KU VI.3: Accountability and processes to ensure it

This is a mirror image of some of the subjects dealt in Module I: Business understanding where are addressed issues like 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.

Suggested activities: The students are exposed to auditing frameworks and are asked to apply them to a use case.

Figure 4. 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-we merely present existing options for each phase, like model cards for example.

A summary table for all KUs can be found in Appendix II

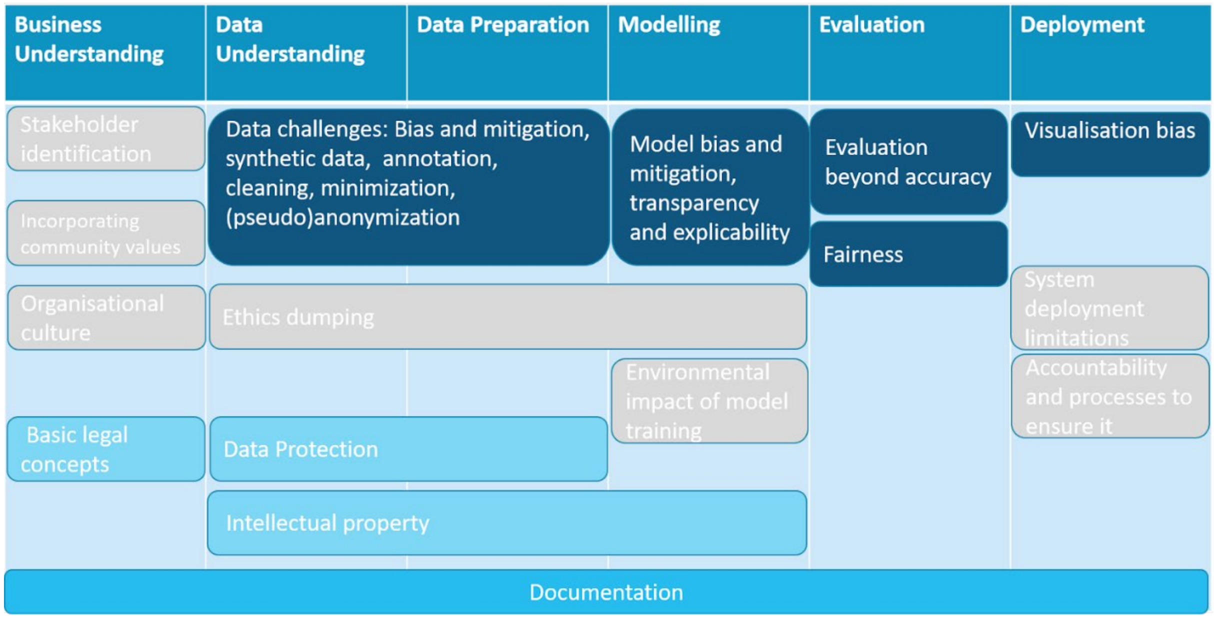

Figure 4. A detailed presentation of the Curriculum. Not all KUs are represented, rather the mail subjects that are addressed. A in Figure 3., with grey we denote the ethical/societal strand, with light blue the legal strand, with and deep blue the technical renderings of the ELSA challenges.

2.5 Means of Content Delivery: Selection and Organization of the Learning Experience

2.5.1 Implementation Strategies

The implementation strategy followed to apply the Curriculum depends on the resources and the time given as well as the target audience.

The ELSA Curriculum is intended for data science practitioners, i.e., people who are already employed in various data science projects. This is a varied and not homogenous group, as it may consist of people having various levels of experience both in data science and exposure in ELSA challenges. Additionally, they may be fulfilling a variety of roles, ranging from programmers who only implement specific pieces of code to system architects and project managers.

The application field is also very important to take into consideration: there are different challenges when implementing a data science project for marketing compared to one in health care.

In its conception the Curriculum tries to assume as generic an audience with as little knowledge both in data science and in ELSA issues as possible. However, we recognize that this is seldom the case.

We propose the following implementation strategies:

1. Introductory course assuming little or no knowledge of ELSA: This course is targeted to new professionals, or to professionals who never had any education or experience regarding ELSA issues. These may be either people who have just finished their studies or moving into data science from another discipline and have no background in ELSA issues. As we have seen in the landscape description and the profile of data scientist(Christoforaki 2022) in the first case of fresh university graduates, very few academic institutions offer ELSA courses and the

latter case comprises people who may have a tertiary education degree or not and they have arrived in data science via non formal education (like online courses or self-study). To those, we may include data scientists who originate from countries outside the one where they presently work in (Germany in our case) and are not knowledgeable about the current working environment and its ELSA requirements (for example, EU Data protection laws). The purpose of the instruction program in this case, is to raise awareness of ELSA issues and enhance sensibility to them, as well as offering general knowledge to be used as a roadmap when encountering such issues.

2. Selection of modules that fit the roles of the target audience: As shown in the Landscape description (Christoforaki 2022), we can distinguish between two kinds of data scientists, one with mastery of maths, statistics and visualisation techniques, including some social science research methods skills –such as the ability to raise the appropriate questions and hypothesis--, as well as soft skills associated with communication and team work and a second one with computer science and programming skills. The first category might benefit more from an in-depth study of some modules (for example, Business understanding and Deployment) or some knowledge units within modules (for example legal KUs), while the second category from deep dives into technical renderings of ELSA issues within modules. This approach does not exclude any content but provides more advanced content with respect to the target audience needs and preferences.

3. Adaptation of the content according to the application domain that interests the target audience: While a general awareness about ELSA challenges can be applied to all application domains, there are some applications that necessitate a special approach. For example, the development of Health Care data science applications has specific and strong demands regarding the Data Protection issues. Additionally, there is variance regarding the technical renderings that make sense, for example in NLP vs, Image processing applications. Whenever possible, it might be desirable to focus on the application domain the target audience works in, for example proposing relevant use cases or practical exercises.

4. A combination of points 2. and 3. This means that the modules will be highly specialised regarding both the application domain and the role of the audience members. That might call for specific Curriculum implementations which could include only some of the Modules or Knowledge units within each module and can either contain more specialised knowledge, have shorter duration than in the general Curriculum coverage or both. These can lead to the development of advanced programs, aimed at specific professionals.

2.5.2 Classroom Techniques

Who should teach

The nature of the curriculum is interdisciplinary, so there is a need for a variety of domain experts (some highly specialised) to cooperate and teach the respective KUs. However, they should always keep in mind that the purpose of the program is to raise ELSA awareness and provide the participants with the means to communicate effectively with the domain experts, not to turn them into ones. Tutors must always have in mind that while the material is multifaceted, containing ethical, legal, societal and technical strands, the focus is always on how a data scientist would be able to cope with it. While we admit that this provides a challenging environment for both instructors and program participants, we hope and believe that the mere coexistence and cooperation between different domains of expertise will benefit both of them. An example of a multidisciplinary course that can be used as a guide can be found in (Reich et al. 2020).

Classroom techniques

For the curriculum implementation we can employ a variety of classroom techniques that are fitting to the specific KUs. These can range from lectures, invited talks and case studies to cooperative learning techniques like student workgroups, role-playing to practical exercises where students are asked to write code, for example detecting data bias into given datasets. At the description of the landscape deliverable (Christoforaki 2022), our review lead us to the conclusion that the students evaluated more positively practical exercises and use case studies (an example illustrating how the FAIR Data Spaces demonstrators can be employed as use cases is given in section 3 .Case Studies: The FAIR-DataSpaces Demonstrators of the present document), as well as engagement activities like discussion and debates based on real world cases provided via public media (newspaper articles, videos, news, etc). While lectures can impart a big amount of knowledge in a small amount of time with not very demanding resources, they also have lowest evaluation assessment with respect to their effectiveness.

2.5.3 Resources for Implementing the Knowledge Units

The material used in the Curriculum implementation is to be decided by the instructors of each KUs. It can be adapted according to the implementation strategy chosen as described in 2.5.1 Implementation Strategies and must be updated since new issues develop constantly as well as methods to tackle them, most notably, in the legal and technical rendering strands.

However, an indicative list of material is provided in Appendix I : Proposed material. It must be underlined that it represents a specific (and in many cases non-expert) person's view at a given point in time and may be used for illustrative purposes only. It mainly consists of what the author of the present deliverable based her review on, that led to the formation of the current Curriculum structure; personal bias should also be taken into consideration.

2.6 Evaluation

The evaluation issue is twofold: how to evaluate the participants and how to evaluate the curriculum and the program implementation.

The first one depends on the implementation selected strategy. The participants could be evaluated by their engagement in the various modules, for example by participation and application of the knowledge acquired in Use Cases, via the organisation of Hands-On Workshops sessions during the study program that implements the Curriculum.

This can be differentiated according to the objective of the program implementation, e.g. awareness or creating competencies in specific domains (e.g. bias detection and mitigation).

Regarding the program, the usual way is to get feedback both from the participants and the instructors on whether the program fulfilled their expectations or not. In this sense, the best way to evaluate the curriculum is the implementation of a pilot instruction program, followed up with a detailed evaluation via participants' and tutors' evaluation sheets, or interviews. However, this is not in the scope of the FAIR Data Spaces project.

3. Case Studies: The FAIR-Data Spaces Demonstrators

In the context of the FAIR Data Spaces project there is the development of three demonstrators: a. for a biodiversity data space, b. research data quality assurance, and c. cross-platform data analytics. These three demonstrators are proposed as *Use Cases for the ELSA Curriculum*. For this purpose, we used input from the *AP2 "Rechtliche und ethische Rahmenbedingungen"* work package regarding the Legal and Ethical Review of the Demonstrators.

In the following sections we will present: a. A brief description of each demonstrator and b. How the demonstrator can be considered as a Use Case for the ELSA data science curriculum.

3.1 Biodiversity Demonstrator

The first demonstrator on biodiversity has as the main software component Geo Engine. It is a cloud-based research environment for spatio-temporal data processing, supporting interactive data analyses for geodata (such as vector and raster data), that allows data scientists to focus on the actual data analyses rather than data preparation.

In the context of the FAIR-Data Spaces project, Geo Engine adds new data connectors and cloud features to demonstrate important aspects of FAIR-DS, including novel use cases which combine data from a variety of sources as well as authentication according to GAIA-X specifications('WP3 WP4 Technical Foundations and Demonstrators' n.d.).

Technical description

Geo Engine consists of a backend and two frontends: geoengine-ui for Web and geoengine-python library (Beilschmidt, et al. 2023).

The backend handles data access, data management and query execution and provides APIs for the frontends and third-party applications and consists of three modules:

- data types: contains the primitives for vector and raster data as well as basic operations and spatial projections
- operators: contains the spatio-temporal query execution engine and the implementation of operators
- Services: contains the data management, e.g., adding, updating and removing datasets, workflows and projects, and Web APIs on top of this functionality. It can't also handle user management, authentication via OpenID Connect single sign-on (SSO) providers, and authorization, which allows restricting access to resources such as data and workflows to certain users and groups.

The main output of Geo Engine are layers of spatio-temporal data: either feature collections or raster images.

Geo Engine can access internal and external data. Loadable data is identified by an ID used by an input operator to resolve the necessary loading information (e.g., name, the location, and the used spatial projection) using a metadata provider. Internal data are stored as datasets in a database. Users can create their own datasets and share them with other users. External data is provided by Data Providers who have to allow browsing and to access data that is not managed by Geo Engine. In contrast to local datasets, external data cannot be edited or deleted and the available data may change over time.

The Web frontend geoengine-ui consists of three parts:

- a core library: provides a client implementation for the backend API services, e.g. for managing layers, and building block components like the map, plots, and operator dialogs
- a GIS application: offers the full functionality of Geo Engine, which is targeted at expert users who can work with multiple layers, apply operators and review workflow graphs
- multiple apps and dashboards: simpler applications that focus on a concrete use case and only require access to a few selected inputs

While the demonstrator is offered in the cloud, in contrast to previous systems, which are operated by cloud providers such as Amazon or Google and prevent the exchange of data between individual providers, this system is intended to explicitly promote such an exchange and prevent a so-called vendor lock-in. The system implemented for this purpose is to be open source and expandable in order to be able to take into account the specific needs of individual users with regard to data protection and data security.

However, data sovereignty over the data used is maintained. Access is not universal, but can be restricted for certain user groups to consider legal frameworks of the data

Application example

As a Geo Engine application, we can examine the use case of a forester that wants to survey the health status of the forest trees. In order to do that, they have to follow a temporal development, for example, to compare the condition of the vegetation in the same month over several years to detect changes that might entail possible measures.

Geo Engine can be used to display the location data of trees on a map in the web browser to which satellite images can be added for analysis, which can lead to measures initiated by experts. For example, the felling of trees due to a beetle infestation.

Ethical and legal concerns exemplified

The Geo Engine demonstrator can be used to exemplify concerns regarding the collection and processing of personal data. Although the demonstrator is not aimed to be used with personal data, it uses a wide range of data that are aggregated and combined with geolocation. A possible subject that can be tackled is the consideration of how exactly the processing of personal data is prevented and how any liability for misuse of the demonstrator via illegitimate processing of personal data might be avoided. One possible approach is to demonstrate how to communicate clearly to potential users of the restrictions on the functionality in relation to personal data, as well as, in the terms of use, exclusion of liability on the part of the service provider - whoever this is - relating to any illegitimate choice to process personal data.

While the use case presented has to do with forest surveillance, a possible application might be surveillance of a disease, for example malaria, or HIV and its spread in different parts of the world. This might introduce considerations regarding bias and stigmatisation not of individuals only (if we accept that no personal data is held by the system), but of whole countries and regions that may be negatively characterised. This fact can reflect badly on the individual inhabitants of these areas that can lead to discrimination, when, for example, these people travel to other countries by border authorities. This can be extrapolated to generalisations used by extremists for political purposes, especially when coupled with race or vulnerable status (e.g. refugees).

3.2 Data Validation and Quality Assurance

The purpose of this demonstrator is to exhibit the use of decentralised task runners to perform automated quality control and data assurance within a commonly available or easily provided environment.

Within the FAIR Data Spaces project, it is aimed to develop a GAIA-X compatible demonstrator that leverages the hardware sources provided with the Open Telecom Cloud and allows a theoretical research group to provide simple information about the data they intend to collect. When data is added to local or cloud-based storage, this information is used to ensure that each new file matches the specified data schemas and to monitor those files for unusual or unexpected values. The demonstrator also ingests these data files to create three types of simple web-based reports that allow users to view simple summary data without having to access the individual data files

Technical description

The demonstrator consists primarily of a python library, which contains all the code necessary to run the analysis. Also provided are a publicly available GitLab repository, which contains a CI/CD script that automatically calls the library and can be customised to the user's use case, and a Docker Container that comes preinstalled with the library and all dependencies needed to execute it.(‘WP3 WP4 Technical Foundations and Demonstrators’ n.d.)

The demonstrator is not intended for use with restricted access or sensitive data sets, but has an authentication process that does provide some security with respect to data access.

The repository is available under an open source licence.

Application example

A use case can be designed as follows:

Researchers aim to monitor crop yields, in order to determine what particular growing conditions and what crop varieties do best. For that reason, they:

- recruit a large group of farmers willing to participate in a semi-annual survey conducted in person via a web form
- Ingest fine-grained weather data provided through a publicly available API.

It is important that this data is properly cleaned and cleared of collection anomalies for later analysis.

Users must provide the repository (run within the GitLab installation at RWTH Aachen University), with access credentials and, at the beginning of the project, schema files in a frictionless data format that describe the table format of their survey data and weather data, including value types, expected value ranges, and field validation patterns. In addition, they can use these schema files to provide metadata about each field in the form of brief explanations.

As surveys are collected, the demonstrator is scheduled to run every night and process any new or updated files. If, for example, it discovers that the surveyor accidentally recorded the date in a format that does not match the one specified in the schema file, this is flagged in the error report and can be corrected immediately before it is included in any further analysis.

The same procedure is followed for automatically created or downloaded data.

In both cases, users are quickly alerted to a quality problem in their data, as when the report is generated, it triggers an email to members of the repository indicating those problems while the report is posted on a website so that all members of the project can see and fix those problems. The report also includes a collection of summary statistics so that users can perform a cursory analysis of each data file.

Ethical and legal concerns exemplified

This demonstrator can be used to exemplify intellectual property issues, specifically secrecy rights (trade secrets) and copyright.

Regarding the first case, it is not the data itself that is protected, but the information behind it. A trade secret is defined in German business secret law as "information (a) which is not generally known or readily accessible, either as a whole or in the precise arrangement and composition of its components, to persons in the circles which normally handle this type of information, and is therefore of economic value, and (b) which is the subject of secrecy measures by its rightful owner which are reasonable under the circumstances, and (c) for which there is a legitimate interest in maintaining secrecy."(§ 2 No. 1 GeschGehG)

Whether data in Gaia-X and NFDI is protected under the above law therefore depends largely on whether it qualifies as a trade secret in the individual case.

It therefore depends on whether the information that the algorithm checks for quality is generally accessible or can only be obtained specifically, for example by the respective members of a company. However, this is very likely to occur when checking a corresponding amount of data, as the FAIR Data Spaces Project. One solution here could be to licence the relevant data as a kind of consent for quality control.

Regarding copyright, according to the relevant German Copyright law (§ 2 UrhG) only works are eligible for copyright protection (definition of work is presented in § 2 II UrhG- what is to be protected must always reach a certain level of creation).

In the case of the specific demonstrator, a copyright issue can be identified when the data are duplicated for quality control by the algorithm, at least in the buffer. According to § 16 UrhG, this right belongs in principle to the creator alone.

However, the author can also grant certain rights to third parties, if necessary comprehensively (§ 31 I, III UrhG), or actions that infringe copyright may be permitted by other justifications, for example for scientific research (§ 60c UrhG), albeit presumably to a rather limited extent.

Automated quality control could, however, fall in particular under text and data mining (§§ 44b, 60d UrhG), which could also permit reproduction (§ 16 UrhG).

3.3 Cross-Platform FAIR data analysis

The third demonstrator uses a federated learning-based platform called PHT (Personal Health Train), which operates on Health care data (e.g. hospital data) and its basic characteristic is, that since it operates on sensitive data, ensures privacy and data governance by keeping the cloud-based data spaces separate from hospital data silos, which are only accessed by local clients.

Technical description

Since the technical implementation of PHT is quite complex, here we describe the basic architecture and the workflow. A more technical description can be found in (Jaberansary et al. 2023)

The PHT concept originates from an analogy from the real world, specifically from the railway system. The basic elements are trains, stations, and train depots:

- The train uses the network to visit different stations and contains specific analysis tasks, which are executed at distributed data nodes (the Stations); they move from Station to Station to consume data as an outcome of the executed analytical task.

  The results are incrementally generated and can be anything based on the Train code. For example, the result set can contain data on an aggregated level, for example, a number showing a cohort size, which has no relationship to individual patient data of the input level, or updated parameters of a statistical model, such as a regression model that is incrementally fitted from Station to Station.

- Stations hold confidential data and execute analytic tasks; hence they have two main components:
  - The data source: A Station can hold the data itself or provides an access point to the sensitive data.
  - Station software: The main task of the Stations is the execution of the containerised analytic algorithms, while they can also reject requests if the Station admin has doubts about the data usage or a lack of capacity and after the task is completed the results are inspected and can be rejected if the result set contains confidential information.

- The depot is represented by our Central Service (CS) including procedures for Train orchestration, operational logic, business logic, and data management. (Welten et al. 2022). The Central Services are responsible for coordinating the execution of the analysis on the distributed data sources within the data space.

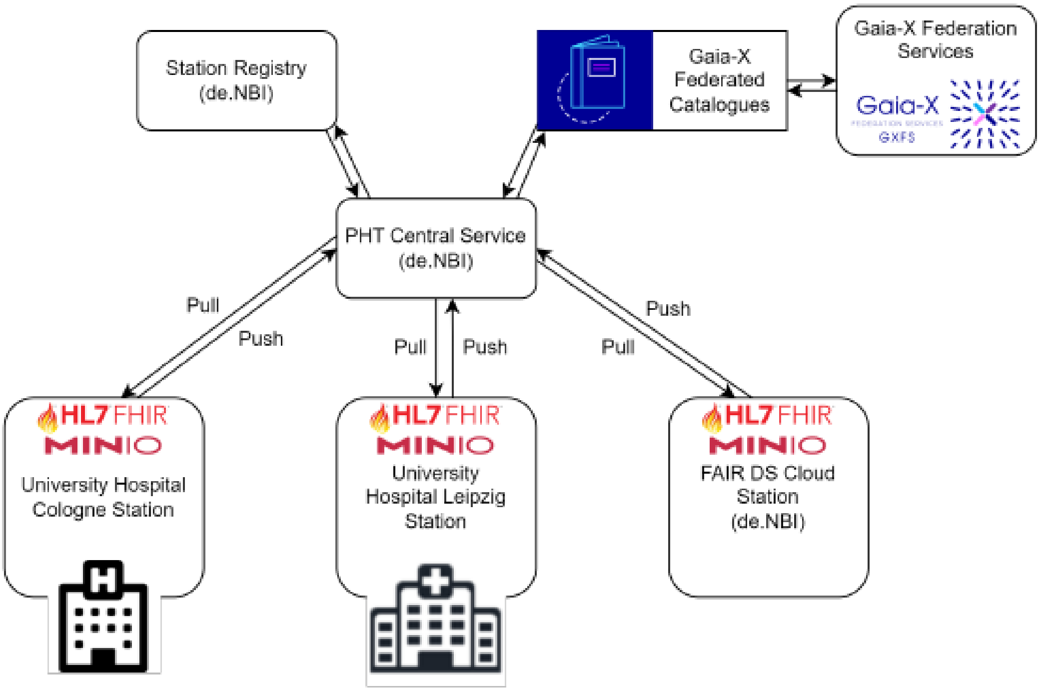

Figure 3. High-level Architecture of the PHT Infrastructure for the Demonstrator (Jaberansary et al. 2023)

A workflow in PHT as illustrated in Fig 3. Goes like this:

• First, the Train (analytical task, in this example a classification of patient record images) is pulled and decrypted by the Station admin.

• The Station admin provides connection information to the data source.

• In the initial phase of the analytical task, existing patient resources in the training set are queried to obtain clinical data and related image references (URLs). The dataset is then divided into training and test datasets. Typically, the ratio is 80:20.

• Based on the referenced URLs, the analysis task then downloads the corresponding images from a server. A classification model is trained using both data types. The existing derived model weights directly update the previous model trained in earlier stations. The test dataset is used for evaluating the accuracy of the model. Such monitor data is saved and used later for inspection and feedback.

• Finally, the Train image is saved with the new results (updated classification model) and then returned to the Train repository in the central service (Jaberansary et al. 2023).

Application example

The application use case concerns Skin Lesion Detection. The dataset is accessible via three different PHT stations and consists of structured clinical data (including age and gender) and, mainly, image data showing different kinds of skin lesions; each image shows one snapshot of one type of (labelled) skin lesion. Apart from the cloud-based station, the two other stations use on-premise infrastructures, which allows the data providers (hospitals in this case) full control of their data.

The analysis aims to train a classification model to predict the skin lesion type of new images of potentially unknown persons. In order to achieve this aim, the process follows the workflow described in the previous section where the algorithm that creates the trained model is loaded on a train that stops at each station, is executed and moves on to the next station, In the end, the trained model arrives to the original researcher who asked it and can be used to identify skin lesions in new images. In the process, no personal or sensitive data is moved around and the data providers exercise full control over their data, revealing as much or as little of it as they wish.

Ethical and legal concerns exemplified

PHT aims *a*t providing a system which is suitable for scientific analyses of sensitive data stored in multiple locations, without the data itself ever leaving the storage locations. This stems from its application domain (health care) and the increased sensitivity regarding the treatment of sensitive personal data. The system can be used to exemplify the challenges regarding personal data in general and specifically in the health care domain, and how the approach of PHT (federated learning) can be used to avoid adverse outcomes.

However, it can also be used to exemplify the issues posed by opaqueness regarding the data, since the end user does not have any access on them, so they are not aware of any issues that might exists, for example underlying biases, either in the original data or the preprocessed data that are made available from each provider. The system is practically a black box where the only result that the end user sees is the trained model but does not have any inkling about the data that lead to that outcome. A possible solution is to apply proper documentation regarding the characteristics of the data set, so, even if the data itself is not available for scrutiny, the metadata regarding the dataset characteristics are. This will enable the end user (researcher) to be able to understand the nature of the dataset and anticipate and mitigate potential issues (e.g. bias), as well as correctly interpret the results.

4. Conclusions

In this report it is presented the first version of an ELSA Curriculum for data scientists. The basic premise of the curriculum structure is that it must follow the CRISP-DM model and employ material organised in Knowledge Units (KUs) belonging to three strands, namely ethical and societal, legal and technical rendering subjects.

While this report is focused on the subjects taught, issues such as means of content delivery and evaluation is also addressed.

Finally, the three demonstrators of the FAIR Data Spaces project are presented as use cases that can be employed in an instruction program, each one exemplifying a variety of ethical, legal and societal aspects.

The second and final version will be presented as a report after gathering feedback from our Gaia-X industry partners (and the overall community) regarding the sustainability of the curriculum. In that version the demonstrators as use cases will be updated as more input from the AP2 project participants will be made available in the form of project deliverables.

References

Beauchamp, Tom L, and James F Childress. 2019. *Principles of Biomedical Ethics*. 8th ed.

Beilschmidt, Christian, Drönner, Johannes, Mattig, Michael, Schweitzer, Philip, and Seeger, Bernhard. 2023. 'Geo Engine: Workflow-Backed Geo Data Portals'. https://doi.org/10.18420/BTW2023-55.

Christoforaki, Maria. 2021. 'ELSA Training for Data Scientists-Describing the Landscape'. Zenodo. https://doi.org/10.5281/zenodo.7233569.

———. 2022. 'Data Scientists and ELSA:Describing the Landscape'. June 1. https://doi.org/10.5281/zenodo.6637759.

'Ethical Guidelines - Gesellschaft Für Informatik e.V.' 2018. 29 June 2018. https://gi.de/ethicalguidelines.

Floridi, Luciano. 2019. 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical'. *Philosophy & Technology* 32 (2): 185–93. https://doi.org/10.1007/s13347-019-00354-x.

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.

Freeman, R. Edward. 2010. *Strategic Management: A Stakeholder Approach*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139192675.

Jaberansary, Mehrshad, Macedo Maia, Yeliz Ucer Yediel, Oya Beyan, and Toralf Kirsten. 2023. 'Analyzing Distributed Medical Data in FAIR Data Spaces'. In *WWW '23 Companion*. Austin, TX, USA.

Leslie, David. 2019. 'Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector'. Zenodo. https://doi.org/10.5281/ZENODO.3240529.

Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. 'From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices'. *Science and Engineering Ethics* 26 (4): 2141–68. https://doi.org/10.1007/s11948-019-00165-5.

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.

Saltz, Jeffrey S., and Neil Dewar. 2019. 'Data Science Ethical Considerations: A Systematic Literature Review and Proposed Project Framework'. *Ethics and Information Technology* 21 (3): 197–208. https://doi.org/10.1007/s10676-019-09502-5.

Shearer, Colin. 2000. 'The CRISP-DM Model: The New Blueprint for Data Mining'. *Journal of Data Warehousing* 5 (4): 13–22.

Taba, Hilda. 1962. *Curriculum Development: Theory and Practice*.

Unabhängiges Landeszentrum für Datenschutz, ed. 2020a. 'Das Standard-Datenschutzmodell (SDM) - ULD'. AK Technik der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder. https://www.datenschutzzentrum.de/sdm/.

———, ed. 2020b. ‘The Standard Data Protection Model’. AK Technik of the Independent Data Protection Supervisory Authorities of the Federation and the Länder. https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology_V2.0b.pdf.

Welten, Sascha, Yongli Mou, Laurenz Neumann, Mehrshad Jaberansary, Yeliz Yediel Ucer, Toralf Kirsten, Stefan Decker, and Oya Beyan. 2022. ‘A Privacy-Preserving Distributed Analytics Platform for Health Care Data’. *Methods of Information in Medicine* 61 (S 01): e1–11. https://doi.org/10.1055/s-0041-1740564.

‘WP3 WP4 Technical Foundations and Demonstrators’. n.d. GitHub. Accessed 14 March 2023. https://github.com/FAIR-DS4NFDI/FAIR-DSWiki/wiki/WP3-WP4-Technical-Foundations-and-Demonstrators.

Appendix I: Proposed material

In this Appendix, we present a (non exhaustive) list of resources that can be used in order to compose the teaching material. The inspiration of compiling and organising this material is drawn from (Morley et al. 2020), where the the authors attempt to provide a list of available tools (via bibliographic references) in order to help Machine Learning experts to move from *What* (ethics principles) to *How* (practices that apply ethics principles). For that reason they create a typology based on the five principles presented in (Floridi et al. 2018) where the tools are assigned to each phase of the ML project development (the phases used correspond to the CRISP-DM model phases, with some name variations).

In a similar fashion, here we use the CRISP-DM phases to propose material that corresponds to the KUs assigned to each phase. Some of the publications appear more than once, since they contain material that can be used to illustrate more than one issue.

As mentioned in 2.5.3 Resources for Implementing the Knowledge Units, the following material is indicative and reflects the author's preferences and choices.

Business understanding

Stakeholder definition and identification

Donaldson, Thomas, and Lee E Preston. 'The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications', 2023.

Freeman, R. Edward. *Strategic Management: A Stakeholder Approach*. Cambridge: Cambridge University Press, 2010. https://doi.org/10.1017/CBO9781139192675.

Typology of stakeholders in Data Science/AI

Ayling, Jacqui, and Adriane Chapman. 'Putting AI Ethics to Work: Are the Tools Fit for Purpose?' *AI and Ethics* 2, no. 3 (1 August 2022): 405–29. https://doi.org/10.1007/s43681-021-00084-x.

High-Level Expert Group on AI (AI HLEG). 'Ethics Guidelines for Trustworthy AI'. Brussels: European Commission, 8 April 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Professional codes of ethics-professional responsibility

Computer science and engineering societies' professional codes

ACM. 'ACM Code of Ethics and Professional Conduct', 22 June 2018. https://ethics.acm.org/.

Chatila, Raja, and John C. Havens. 'The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems'. In *Robotics and Well-Being*, edited by Maria Isabel Aldinhas Ferreira, João Silva Sequeira, Gurvinder Singh Virk, Mohammad Osman Tokhi, and Endre E. Kadar, 95:11–16. Intelligent Systems, Control and Automation: Science and Engineering. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-12524-0_2.

'Ethical Guidelines - Gesellschaft Für Informatik e.V.', 29 June 2018. https://gi.de/ethicalguidelines.

KI Bundesverband e.V. 'KI Gütesiegel', 22 February 2019. https://ki-verband.de/wp-content/uploads/2019/02/KIBV_Guetesiegel.pdf.

The philosophy and use of professional codes

Frankel, Mark S. 'Professional Codes: Why, How, and with What Impact?' *Journal of Business Ethics* 8, no. 2–3 (1989): 109–15. https://doi.org/10.1007/BF00382575.

Metcalf, Jacob. 'Ethics Codes: History, Context, and Challenges'. *Council for Big Data, Ethics and Society* 9 (9 November 2014): 15.

Mittelstadt, Brent. 'Principles Alone Cannot Guarantee Ethical AI'. *Nature Machine Intelligence* 1, no. 11 (November 2019): 501–7. https://doi.org/10.1038/s42256-019-0114-4.

Rochel, Johan, and Florian Evéquoz. 'Getting into the Engine Room: A Blueprint to Investigate the Shadowy Steps of AI Ethics'. *AI & SOCIETY*, 17 September 2020. https://doi.org/10.1007/s00146-020-01069-w.

- by design approaches

Ethics

Aquin, Mathieu d', Pinelopi Troullinou, Noel E. O'Connor, Aindrias Cullen, Gráinne Faller, and Louise Holden. 'Towards an "Ethics by Design" Methodology for AI Research Projects'. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 54–59. AIES '18. New York, NY, USA: Association for Computing Machinery, 2018. https://doi.org/10.1145/3278721.3278765.

Dainow, Brandt, and P. Brey. 'Ethics by Design and Ethics of Use Approaches for Artificial Intelligence', 25 November 2021. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf.

Privacy

Cavoukian, Ann. 'Privacy by Design :The 7 Foundational Principles', n.d., 2.

Danezis, George, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Métayer, Rodica Tirtea, and Stefan Schiffner. *Privacy and Data Protection by Design - from Policy to Engineering*, 2014. https://doi.org/10.2824/38623.

Hildebrandt, Mireille. 'Saved by Design? The Case of Legal Protection by Design'. *NanoEthics* 11, no. 3 (1 December 2017): 307–11. https://doi.org/10.1007/s11569-017-0299-0.

Hildebrandt, Mireille. *Law for Computer Scientists and Other Folk*. First edition. Oxford, United Kingdom: Oxford University Press, 2020.

Basic concepts

Legal concepts

Hildebrandt, Mireille. *Law for Computer Scientists and Other Folk*. First edition. Oxford, United Kingdom: Oxford University Press, 2020.

Leslie, David, Christopher Burr, Mhairi Aitken, Josh Cowls, Michael Katell, and Morgan Briggs. ‘Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: A Primer’. *SSRN Electronic Journal*, 2021. https://doi.org/10.2139/ssrn.3817999.

Rodrigues, Rowena. ‘Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities’. *Journal of Responsible Technology* 4 (1 December 2020): 100005. https://doi.org/10.1016/j.jrt.2020.100005.

Viljanen, Mika, and Henni Parviainen. ‘AI Applications and Regulation: Mapping the Regulatory Strata’. *Frontiers in Computer Science*, 2022, 141.

Audit, impact assessment, compliance, risk assurance

Ada Lovelace Institute and DataKind UK. ‘Examining the Balck Box-Tools for Assessing Algorithmic Systems’, 2020. https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf.

Ayling, Jacqui, and Adriane Chapman. ‘Putting AI Ethics to Work: Are the Tools Fit for Purpose?’ *AI and Ethics* 2, no. 3 (1 August 2022): 405–29. https://doi.org/10.1007/s43681-021-00084-x.

Rochel, Johan, and Florian Evéquoz. ‘Getting into the Engine Room: A Blueprint to Investigate the Shadowy Steps of AI Ethics’. *AI & SOCIETY*, 17 September 2020. https://doi.org/10.1007/s00146-020-01069-w.

Data understanding and Preparation

Data protection

Hildebrandt, Mireille. *Law for Computer Scientists and Other Folk*. First edition. Oxford, United Kingdom: Oxford University Press, 2020.

Unabhängiges Landeszentrum für Datenschutz, ed. ‘Das Standard-Datenschutzmodell (SDM) - ULD’. AK Technik der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder, 17 April 2020. https://www.datenschutzzentrum.de/sdm/.

Data bias during data collection and ways to mitigate it

Naden, Jesse Russell, Zixiao (Ken) Wang, and Brent Hecht. ‘Turkers, Scholars, “Arafat” and “Peace”: Cultural Communities and Algorithmic Gold Standards’. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 826–38. Vancouver BC Canada: ACM, 2015. https://doi.org/10.1145/2675133.2675285.

Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. ‘Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries’. *Frontiers in Big Data* 2 (2019). https://doi.org/10.3389/fdata.2019.00013.

Suresh, Harini, and John Guttag. ‘A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle’. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9. EAAMO ’21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3465416.3483305.

Sen, Shilad, Margaret E. Giesel, Rebecca Gold, Benjamin Hillmann, Matt Lesicko, Samuel Bellamy, Rachel K. E., Kuntal Dey, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, et al. ‘AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and

Mitigating Unwanted Algorithmic Bias'. *ArXiv:1810.01943 [Cs]*, 3 October 2018. http://arxiv.org/abs/1810.01943.

Image Datasets and bias

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 'Imagenet: A Large-Scale Hierarchical Image Database'. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. Ieee, 2009.

Kuznetsova, Alina, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, et al. 'The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale'. *International Journal of Computer Vision* 128, no. 7 (July 2020): 1956–81. https://doi.org/10.1007/s11263-020-01316-z.

Yang, Kaiyu, Klint Qinami, Li Fei-Fei, Jia Deng, and Olga Russakovsky. 'Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy'. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 547–58. Barcelona Spain: ACM, 2020. https://doi.org/10.1145/3351095.3375709.

Intellectual property

General Concepts

Cohen, Julie E. 'What Kind of Property Is Intellectual Property?' *Houston Law Review* 52, no. 2 (12 December 2014). https://houstonlawreview.org/article/3995-what-kind-of-property-is-intellectual-property.

Davies, Colin R. 'An Evolutionary Step in Intellectual Property Rights – Artificial Intelligence and Intellectual Property'. *Computer Law & Security Review* 27, no. 6 (1 December 2011): 601–19. https://doi.org/10.1016/j.clsr.2011.09.006.

Hildebrandt, Mireille. *Law for Computer Scientists and Other Folk*. First edition. Oxford, United Kingdom: Oxford University Press, 2020.

Intellectual property of training data

Franceschelli, Giorgio, and Mirco Musolesi. 'Copyright in Generative Deep Learning'. *Data & Policy* 4 (ed 2022): e17. https://doi.org/10.1017/dap.2022.10.

Sobel, Benjamin. 'A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning'. SSRN Scholarly Paper. Rochester, NY, 19 August 2020. https://doi.org/10.2139/ssrn.3677548.

Dataset documentation

Bender, Emily M., and Batya Friedman. 'Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science'. *Transactions of the Association for Computational Linguistics* 6 (December 2018): 587–604. https://doi.org/10.1162/tacl_a_00041.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 'Datasheets for Datasets'. *ArXiv:1803.09010 [Cs]*, 19 March 2020. http://arxiv.org/abs/1803.09010.

Holland, Sarah, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 'The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards'. *ArXiv Preprint ArXiv:1805.03677*, 2018.

Ethics dumping

Floridi, Luciano. 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical'. *Philosophy & Technology* 32, no. 2 (1 June 2019): 185–93. https://doi.org/10.1007/s13347-019-00354-x.

Schroeder, Doris, Julie Cook, François Hirsch, Solveig Fenet, and Vasantha Muthuswamy, eds. *Ethics Dumping: Case Studies from North-South Research Collaborations*. SpringerBriefs in Research and Innovation Governance. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-64731-9.

Synthetic Data

Howe, Bill, Julia Stoyanovich, Haoyue Ping, Bernease Herman, and Matt Gee. 'Synthetic Data for Social Good'. arXiv, 24 October 2017. http://arxiv.org/abs/1710.08874.

Li, Der-Chiang, Susan C. Hu, Liang-Sian Lin, and Chun-Wu Yeh. 'Detecting Representative Data and Generating Synthetic Samples to Improve Learning Accuracy with Imbalanced Data Sets'. *PLOS ONE* 12, no. 8 (3 August 2017): e0181853. https://doi.org/10.1371/journal.pone.0181853.

Savage, Neil. 'Synthetic Data Could Be Better than Real Data'. *Nature*, 27 April 2023. https://doi.org/10.1038/d41586-023-01445-8.

Social impact of digital labour

Casilli, Antonio A. 'Digital Labor Studies Go Global:Toward a Digital Decolonial Turn', 2017, 21.

Couldry, Nick, and Ulises A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. *The Costs of Connection*. Stanford University Press, 2019. https://doi.org/10.1515/9781503609754.

Graham, Mark, Isis Hjorth, and Vili Lehdonvirta. 'Digital Labour and Development: Impacts of Global Digital Labour Platforms and the Gig Economy on Worker Livelihoods'. *Transfer: European Review of Labour and Research* 23, no. 2 (1 May 2017): 135–62. https://doi.org/10.1177/1024258916687250.

Hao, Karen. 'Artificial Intelligence Is Creating a New Colonial World Order'. *MIT Technology Review*, 19 April 2022. https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/.

Irani, Lilly. 'The Hidden Faces of Automation'. *XRDS: Crossroads, The ACM Magazine for Students* 23, no. 2 (15 December 2016): 34–37. https://doi.org/10.1145/3014390.

Reese, Hope. 'Inside Amazon's Clickworker Platform: How Half a Million People Are Being Paid Pennies to Train AI'. *TechRepublic*, 16 December 2016. https://www.techrepublic.com/article/inside-amazons-clickworker-platform-how-half-a-million-people-are-training-ai-for-pennies-per-task/.

Ross, Joel, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 'Who Are the Turkers? Worker Demographics in Amazon Mechanical Turk', 5, 2010.

Silberman, M. S., B. Tomlinson, R. LaPlante, J. Ross, L. Irani, and A. Zaldivar. 'Responsible Research with Crowds: Pay Crowdworkers at Least Minimum Wage'. *Communications of the ACM* 61, no. 3 (21 February 2018): 39–41. https://doi.org/10.1145/3180492.

Modelling

Model bias and mitigation techniques

Bellamy, Rachel K. E., Kuntal Dey, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, et al. 'AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias'. *ArXiv:1810.01943 [Cs]*, 3 October 2018. http://arxiv.org/abs/1810.01943.

Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 'Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries'. *Frontiers in Big Data* 2 (2019). https://doi.org/10.3389/fdata.2019.00013.

Suresh, Harini, and John Guttag. 'A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle'. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9. EAAMO '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3465416.3483305.

Environmental impact

Ananthaswamy, Anil. 'In AI, Is Bigger Always Better?' *Nature* 615, no. 7951 (8 March 2023): 202–5. https://doi.org/10.1038/d41586-023-00641-w.

Kaack, Lynn H., Priya L. Donti, Emma Strubell, George Kamiya, Felix Creutzig, and David Rolnick. 'Aligning Artificial Intelligence with Climate Change Mitigation'. *Nature Climate Change* 12, no. 6 (June 2022): 518–27. https://doi.org/10.1038/s41558-022-01377-7.

Luccioni, Alexandra Sasha, Sylvain Viguier, and Anne-Laure Ligozat. 'Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model'. arXiv, 3 November 2022. http://arxiv.org/abs/2211.02001.

Schwartz, Roy, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 'Green AI'. *ArXiv:1907.10597 [Cs, Stat]*, 13 August 2019. http://arxiv.org/abs/1907.10597.

Vinuesa, Ricardo, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. 'The Role of Artificial Intelligence in Achieving the Sustainable Development Goals'. *Nature Communications* 11, no. 1 (13 January 2020): 233. https://doi.org/10.1038/s41467-019-14108-y.

Transparency

General

Lipton, Zachary Chase. 'The Mythos of Model Interpretability'. In *Proceedings of the 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, edited by Been Kim, Dmitry M.

Malioutov, and Kush R. Varshney, 96–100. New York: eprint arXiv:1607.02531, 2016. https://arxiv.org/pdf/1606.03490.pdf.

Right to Explanation

Goodman, Bryce, and Seth Flaxman. 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"'. *AI Magazine* 38, no. 3 (2 October 2017): 50–57. https://doi.org/10.1609/aimag.v38i3.2741.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation'. *International Data Privacy Law* 7, no. 2 (2017): 76–99.

Documentation

Bender, Emily M., and Batya Friedman. 'Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science'. *Transactions of the Association for Computational Linguistics* 6 (December 2018): 587–604. https://doi.org/10.1162/tacl_a_00041.

Crisan, Anamaria, Margaret Drouhard, Jesse Vig, and Nazneen Rajani. 'Interactive Model Cards: A Human-Centered Approach to Model Documentation'. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 427–39. FAccT '22. New York, NY, USA: Association for Computing Machinery, 2022. https://doi.org/10.1145/3531146.3533108.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 'Model Cards for Model Reporting', 14 January 2019. https://doi.org/10.1145/3287560.3287596.

Open AI. 'GPT-4 System Card', 23 March 2023. https://cdn.openai.com/papers/gpt-4-system-card.pdf

Evaluation

Evaluation beyond accuracy

Ananthaswamy, Anil. 'In AI, Is Bigger Always Better?' *Nature* 615, no. 7951 (8 March 2023): 202–5. https://doi.org/10.1038/d41586-023-00641-w.

Bastani, Hamsa, Kimon Drakopoulos, Vishal Gupta, Ioannis Vlachogiannis, Christos Hadjichristodoulou, Pagona Lagiou, Gkikas Magiorkinis, Dimitrios Paraskevis, and Sotirios Tsiodras. 'Efficient and Targeted COVID-19 Border Testing via Reinforcement Learning'. *Nature* 599, no. 7883 (November 2021): 108–13. https://doi.org/10.1038/s41586-021-04014-z.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations'. *Science* 366, no. 6464 (25 October 2019): 447–53. https://doi.org/10.1126/science.aax2342.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. '"Why Should I Trust You?": Explaining the Predictions of Any Classifier'. *ArXiv:1602.04938 [Cs, Stat]*, 9 August 2016. http://arxiv.org/abs/1602.04938.

Fairness

Introductory

Ekstrand, Michael D., Anubrata Das, Robin Burke, and Fernando Diaz. 'Fairness in Information Access Systems'. *Foundations and Trends® in Information Retrieval* 16, no. 1–2 (2022): 1–177.

Lee, Michelle Seng Ah, Luciano Floridi, and Jatinder Singh. 'Formalising Trade-Offs beyond Algorithmic Fairness: Lessons from Ethical Philosophy and Welfare Economics'. *AI and Ethics* 1, no. 4 (1 November 2021): 529–44. https://doi.org/10.1007/s43681-021-00067-y.

Verma, Sahil, and Julia Rubin. 'Fairness Definitions Explained'. In *Proceedings of the International Workshop on Software Fairness*, 1–7. Gothenburg Sweden: ACM, 2018. https://doi.org/10.1145/3194770.3194776.

Legal Provisions for fairness under EU law

Hacker, Philipp. 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law'. SSRN Scholarly Paper. Rochester, NY, 18 April 2018. https://papers.ssrn.com/abstract=3164973.

Beyond Fairness as a technical solution

Saxena, Nripsuta, Karen Huang, Evan DeFilippis, Goran Radanovic, David Parkes, and Yang Liu. 'How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness'. arXiv, 27 January 2019. http://arxiv.org/abs/1811.03654.

Srivastava, Megha, Hoda Heidari, and Andreas Krause. 'Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning'. arXiv, 8 December 2019. http://arxiv.org/abs/1902.04783.

Woodruff, Allison, Sarah E. Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 'A Qualitative Exploration of Perceptions of Algorithmic Fairness'. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. Montreal QC Canada: ACM, 2018. https://doi.org/10.1145/3173574.3174230.

Deployment

Audit

Ayling, Jacqui, and Adriane Chapman. 'Putting AI Ethics to Work: Are the Tools Fit for Purpose?' *AI and Ethics* 2, no. 3 (1 August 2022): 405–29. https://doi.org/10.1007/s43681-021-00084-x.

Gursoy, Furkan, and Ioannis A. Kakadiaris. 'System Cards for AI-Based Decision-Making for Public Policy'. arXiv, 31 August 2022. http://arxiv.org/abs/2203.04754.

Rochel, Johan, and Florian Evéquoz. 'Getting into the Engine Room: A Blueprint to Investigate the Shadowy Steps of AI Ethics'. *AI & SOCIETY*, 17 September 2020. https://doi.org/10.1007/s00146-020-01069-w.

Visualisation Bias

Ellis, Geoffrey, ed. *Cognitive Biases in Visualizations*. 1st ed. 2018. Cham: Springer International Publishing : Imprint: Springer, 2018. https://doi.org/10.1007/978-3-319-95831-6.

Mansoor, Hamid, and Lane Harrison. 'Data Visualization Literacy and Visualization Biases: Cases for Merging Parallel Threads'. In *Cognitive Biases in Visualizations*, edited by Geoffrey Ellis, 87–96. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-95831-6_7.

Szafir, Danielle Albers. 'The Good, the Bad, and the Biased: Five Ways Visualizations Can Mislead (and How to Fix Them)'. *Interactions* 25, no. 4 (2018): 26–33.

Appendix II: KUs Table

Business Understanding	Data Understanding	Data Preparation	Modelling	Evaluation	Deployment
KU I.1: Stakeholder identification	KU II.1: Data Protection	KU III.1: Data Protection	KU IV.1: Model bias and mitigation techniques	KU V.1: Evaluation beyond accuracy	KU VI.1: System deployment limitations
KU I.2: Incorporating community values	KU II.2: Data bias during collection and ways to mitigate it	KU III.2: Data challenges in preprocessing	KU IV.2: Model transparency and explicability	KU V.2: Fairness	KU VI.2: Visualisation bias
KU I.3: Organisational culture	KU II.3: Intellectual property issues and licences	KU III.3: Intellectual property issues of training data	KU IV.3: Environmental impact of model training		
KU I.4: Basic concepts	KU II.4: Ethics dumping in data collection	KU III.4 :Ethics dumping in data preprocessing	KU IV.4: Intellectual property issues		
	KU II.5: Dataset documentation		KU IV.5: Model documentation		

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

#FAIRDataSpaces @FAIRDataSpaces

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