**Fairness-Intermediate:**

**Module 1: Responsible AI (1 lesson)**

*This module will focus on a systematic approach to identify ethical and social impacts. A systematic approach is an important first step to explicate possible ethical concerns around fairness and transparency before attempting to address this concern. The module will explain the ethical impact assessment approach that we use in our company to identify ethical and social concerns (see (Wright, 2011). The method will incorporate conceptualisations of fairness from ‘Milestone 1: Conceptual Introduction to Fairness’.*

Within this module we discuss a method to systematically apply ethics to real-world situations. Applied ethics is about what we find morally right or wrong, good or bad in real life situations.: in other words, in different situations applied ethics should give us guidance around what the right thing to do is. Applied ethics can be useful within different domains. Biomedical ethics helps navigate ethical dilemmas around medical interventions, e.g., end of life decisions. Research ethics helps us ensure we are conducting research responsibly, e.g., treating human participants with the dignity they deserve. Business ethics for example helps us think about the responsibility of corporations towards their employees and We can also apply ethics to

Even technologies can be designed with ideas about right and wrong. And we can think of technologies or objects as having morally right or wrong values embedded. This may sound very vague but let me show some examples which may clarify this a bit. One famous example is that of bridges. This example was provided by Langdon Winner in his text Do Artefacts Have Polics (Winner, 1980) According to him, even bridges can have politics and morally undesirable values embedded. A famous urban developer in New York, Robert Moses, created bridges in the 1930s, which were arguably designed for unethical purposes. He was working on the development of a park in New York, Long Island that gave people access to a beach that was mainly used by rich people. He did not like the fact of poor people and minorities visiting these beaches. So he decided on a low height for bridges, so buses, the main transportation of poor people and minorities would not be able to use the access roads to the beaches as they couldn’t go under these bridges. These bridges are examples that politics can play a role in artefacts. I think most of us would agree that the shapes of these bridges is wrong. We would find the reason why these bridges have this design unfair and unjust.

There are different examples that technology can have values build-in. Sometimes this is inadvertently. The technology might perpetuate existing stereotypes or inequalities. For example, a common US health-care algorithm used for the decision-making software that allocated healthcare appeared to have bias build in. It appeared less likely to refer people who self-identified as black to healthcare programmes to improve care for people with complex medical needs than equally as sick people who self-identified as white.(Ledford, 2019). People that were black were less likely to be referred to personalised healthcare programmes. Although the source of this bias is unclear, scientist speculate that the algorithm might perpetuate existing discrimination within the healthcare system by healthcare provider or revealed existing distrust in the black community towards care. The algorithm assigned risk-scores based on cost with the developers assuming that higher healthcare cost indicates greater medical need. However, black people with the same medical need get on average US$1,800 less than their white counterparts.

The previous two examples showed how technology can be designed purposely or inadvertently to advance in ways that have undesirable ethical outcomes. However, applied ethics can also be more forward looking to see how technology changes the world and present us with new dilemmas and opportunities. ”Technology offers new opportunities that will shape the future of our children, our societies, our species, and others who share our plant, in ways never before possible.” (Vallor, Technology & The Virtues, P3) According to Vallor we suffer from "acute technosocial opacity”. All the technological abilities that we have obtained and will obtain make it increasingly more difficult to increasingly difficult to assess and evaluate what we find important in life. They create new opportunities and possibilities and we struggle with how to apply these in a meaningful way. Many of these technologies seem to encourage undesirable behaviour. We need to learn how to use these technologies in ways that work for us. Vallor for example mentions that modern communication channels, such as the internet, is not something comparable to what we had before. It is something fundamentally new that gives us new opportunities that we need to learn how to use in a responsible and ethical way.

Technology does not only present us with new opportunities but they also change the way we experience the world. They transform our world. Peter Paul Verbeek (Moralising Technology, P1) provides the example of cars, which are, according to him, not just mode of transportation to get us to where we need to be. They allow us to bridge distances more easily, so with a car we can visit people further away from us than before. We can live further away from our work because now we can travel more easily and they. They have changed our professional and social lives. Even our environment changed with the introduction of the car. Our cities became car friendly with the infrastructure. These transformation bring ethical challenges and sometimes pose ethical dilemmas. He also gives the example of an ultrasound test. With the introduction of the ultrasound, we can identify birth defects, such as down syndrome or spina bfida. In these case, is abortion allowed? If not, when and under what circumstance is abortion allowed. The ultrasound test changed or transformed the way we look at Similarly, using data we can create insights we have never had before.

Another way technology can have ethical implications is that it can be designed to influence the behaviour of people affected. Verbeek gives the example of speedbumps that are intentionally designed to influence people’s behaviour by slowing them down whereas the ethical dilemmas of an ultrasound tests were not intentional. AI examples of these types of technology are for example technology that automatically adjust temperature or the verification of traveler’s identities on-the-move through 3D face, iris, and somatotype recognition (see D4FLY).

We have to think how the design of technology can have ethical and social impacts. Addressing ethical concerns is important. First of all, to mitigate potential negative impacts on society, see the example with the discriminatory healthcare algorithm, and maximise positive impacts. Second, to make sure we use technology and the new opportunities we have created by developing the technology appropriately and ethically. Third, a failure to address ethical issues can have negative economic consequences for the developers and vendors of the technology. An example is the failed NHS England ‘care.data’ project. Nuffield Council on Bioethics wrote a report using care.data as a use-case.(Nuffield Council on, 2015) Care.data care.data collected data from GP practices and hospitals and combined it with existing data within the NHS. By doing so, I wanted to make health data available for research into new treatments and to assess the performance of NHS services. Care.data provides an example of what can happen when social and ethical considerations are overlooked. Then Prime Minister, David Cameron, said the plan would “make the UK the best place in the world to carry out cutting edge research”. The care.data programme was announced in 2013 and was abandoned in 2016. Nearly £8 million was lost. It was plagued with unaddressed ethical and social concerns. Privacy concerns were not sufficiently protected: they stored the data in a central database and involved private parties with a risk that they could access potentially personal medical data. The people whose medical data were being used were not properly informed, public consultation was lacking, and the opt-out method were questioned. These are concerns around people’s autonomy. They did not allow people the ability to choose for themselves.” ([Jacquemard 2020](https://data.oireachtas.ie/ie/oireachtas/libraryResearch/2021/2021-03-31_spotlight-ehealth-in-ireland-social-and-ethical-values-in-irish-policy-on-ehealth_en.pdf)).

However, it is not always straightforward to identify and address ethical concerns. This difficulty is captured by the Collingridge dilemma: Collingridge (1980) formulated a dilemma discussing the difficulties of addressing ethical concerns at an early stage when you do not exactly know what the consequences of these technologies are. He stated that “when change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time consuming” (Collingridge 1980). Consequently, potential issues either need to be addressed at an early stage of development when information is limited, or the technology becomes embedded and the ability to address challenges has decreased. (Worthington, 1982)

This issue can be somewhat alleviate by addressing ethical concerns at different stage of the technology’s lifecyle.

Ethical concerns can be identified and addressed at different stages of the technology’s lifecyle and there are systematic approaches to address these concern at different stage. Reijers et al identified systematic approaches to identify and address ethical concerns at three different stage: before research and innovation, during research and innovation, and after (Reijers et al., 2018).

We will discuss these here:

The first ones are before development. This is at the start of research and innovation practices. These methods are about emerging technologies and are aimed to deal with the uncertainty from new and emerging technologies. According to Reijers, often they do not only discuss ethical impacts, but also how they affect our thinking about ethical values. Emerging technologies have ethical impacts but can also change the way we think about ethics. We can see, for example, how digital technologies have affected our concept and norms around privacy. These methods have a few steps in common.

* + It often includes steps to identify potential emerging technologies and their characteristics and defining features. Because these are emerging, these features might not be evident yet.
  + It also often includes steps to construct scenarios about future ethical impacts. It describes how the technology is used. Scenarios can be seen as a useful instrument to think about the future.
  + They also involve steps to evaluate potential ethical impacts. These impacts have not occurred yet, because they are emerging. For example, potential compromises of privacy. They might look at the likelihood and impact of ethical impacts from the technology.
  + It takes steps to assess and address the status of uncertain normative claim. The claims we do around emerging technology are less ‘certain’ than those for technology that is current. This epistemic uncertainty needs to be addressed.

Reijers also identified a second stage, the intra stage, which discusses ethical impacts during the research and innovation stage to enable, organise and ensure ethical technology design.

* + These applied ehics approaches incorporate steps to disclose ethical issues in design (e.g. identify how the technology affects privacy or the potential of hidden values)
  + The also contain steps to explain how values are embedded in technology design. For example, how ideas around accountability, responsibility and transparency shape the technology.
  + It also includes steps to help organise the process of “practising ethics” in the technology-development pipeline (see the process of Ethical Impact Assessment we will discuss below)

The last stage for applied ethics is ex-post. It identifies ethical impacts of existing technologies. This can help end-users of the technology identify and address ethical issues within their work. The use of checklist is an example.

* These methods include steps to identify ethical issues raised by existing technologies.
* These methods include steps to analyse ethical issues raised by existing technologies: for example by using a matrix which shows what ethical issues is promoted or hindered by a research and innovation outcome .
* Additionally, it contains steps to support ethical decision-making with existing technology. In relation to algorithms, for example, it could be used to identify parameters for when an algorithm leads to unfair outcomes.

**Interview (10 min)**

* What is an ethical impact assessment?
* Why is an ethical impact assessment important?
* Can you give an example where an ethical impact assessment made an important difference to the technology or maybe society at large?
* How common are ethical impact assessment these days?

Within this module, we want to focus on ethical impact assessments which is a method that can be used during the research and innovation stage. They are an intra method to ensure that ethical impacts can be considered during design so negative impacts – for example on privacy or fairness – can be addressed while positive impacts can be maximized – for example, by making the technology more accessible. This method is especially relevant when we want to introduce fairness and transparency considerations into the design of a technology we want to develop. It supports developers with thinking about the values and outcomes they want to achieve and address potential conflicts between values.

An ethical impact assessment is a systematic method to ensure ethical considerations are adequately examined. There are different benefits behind an EIA.

* To identify ethical issues early, Wright calls it “an early warning system”, so they can addressed timely. For example, in relation to privacy, it can help detect privacy problems, allow the developer to tak eprecuations and develop safeguards before investments into technologies that will be ethically undesirable and may trigger public backlash.
* The methodology might unearth valuable insights and hidden assumptions about the technology. By doing so, it can help make informed decisions and expose communication gaps.
* A PIA can help an organisation to gain the public’s trust and confidence that privacy has been built into the design of a project, technology or service….
* An organisation that undertakes a PIA demonstrates to its employees and contractors that it takes privacy seriously and expects them to do so too….
* A proper PIA also demonstrates to an organisation’s customers and/or citizens that it respects their privacy and is responsive to their concerns.” [Wright et al 2011](https://piafproject.files.wordpress.com/2018/03/piaf_d1_21_sept2011revlogo.pdf)

There are certain steps that can be taken to conduct adequate ethical impact assessments. These were based on previous literature and good practices. We will briefly explain them below.

Steps of an EIA ([Wright 2013](https://www.dhi.ac.uk/san/waysofbeing/data/data-crone-wright-2013.pdf))

* “Determine whether an EIA is necessary (threshold analysis): the first on is if an EIA is necessary at all. One criterium is to assess if the project potentially has strong negative impacts. Another is if the EIA can have any impact itself. If the technology cannot be changed to address issues found within the EIA or abandoned if the negative impacts appear decisive then the scope for an EIA decreases.
* Identify the EIA team and set the team's terms of reference, resources and time frame: to conduct an EIA, expertise on ethics is required as well as enough resources and time to conduct the EIA.
* Prepare an EIA plan: this lays out what work will need to be done to complete the EIA successfully including roles within the project and consultation with external stakeholders.
* Describe the proposed project to be assessed: this step captures the technology and its use context (why the technology is being developed, who is impacted by the technology, who is responsible for the technology)
* Identify stakeholders: the following step is to identify those affected by or interested in the technology. These stakeholders can be considered widely and include internal and external stakeholders.
* Analyse the information flows and other privacy and ethical impacts: the information flows and other ethical impacts need to be identified. This includes information about how the organization uses information, what personal information is stored, if this information is sensitive, how this information is processed, for what purposes this information is used, etcetera.
* Consult with stakeholders: Consultation with stakeholders helps us identify ethical issues previously not considered and could help reconsider previously identified impacts. Additionally, including stakeholders within this process helps include new voices into the design, who can therefore not complain about not being heard. It shows that people’s concerns are being heard and might unearth hidden assumptions mentioned previously.
* Check the project complies with legislation: although compliance is not the primary aim of the technology, the technology should conform to legal standards.
* Identify risks and possible solutions: with this information the assessor can identify risks and their impact and likelihood and identify solutions to address these risks, if there are any.
* Formulate recommendations: Based on the previous steps, the assessor can formulate recommendations. One point to remember is to clearly identify who to address the recommendation to and who should act upon these recommendations.
* Prepare and publish the report: ideally, the report will be published publicly. Transparency has the additional advantage of showing the public that the company takes ethical concerns serious and intends to act upon it.
* Implement the recommendations: a company should act upon the recommendations and explain why they do not follow some or all of the recommendations.
* Third-party review and/or audit of the P+EIA: EIA vary greatly in quality. An external auditor who check the quality of the EIA is recommended.
* Update the P+EIA if there are changes in the project: the EIA should often be an iterative process. Projects and use contexts often change, and so do the findings of the EIA. To ensure the ethics are sufficiently embedded into the project, an EIA might need updating.
* Embed privacy and ethical awareness throughout the organisation and ensure accountability: Employees should be sensitive to ethical issues and often that comes top-down. Good ethical practice can be worthy for companies who want to prevent reputational damage.

Unfortunately undertaking an EIA can be challenging ([Wright 2013](https://www.dhi.ac.uk/san/waysofbeing/data/data-crone-wright-2013.pdf)) (Wright & Friedewald, 2013)

* An EIA requires expertise in ethics: “Finding the right people to undertake the P+EIA is probably the principal challenge
* Identifying and operationalising criteria against which to assess the privacy and ethical impacts may be a challenge and may require inputs from others, perhaps from both internal and external stakeholders. We need good criteria to ensure the validity and credibility of the assessment itself. Ideally the criteria are measurable;
* Identifying the privacy and ethical risks is also challenging. Identifying risks should be done systematically, taking future threats and vulnerabilities into account. Again, the collaboration of stakeholders will be helpful in this regard. Similarly Considering the privacy and ethical impacts of new and emerging technologies is a difficult challenge, because technologies may have intended as well as unintended con- sequences.
* Finding and encouraging stakeholders to participate in consultation exercises is a challenge. Consultation fatigue may exist

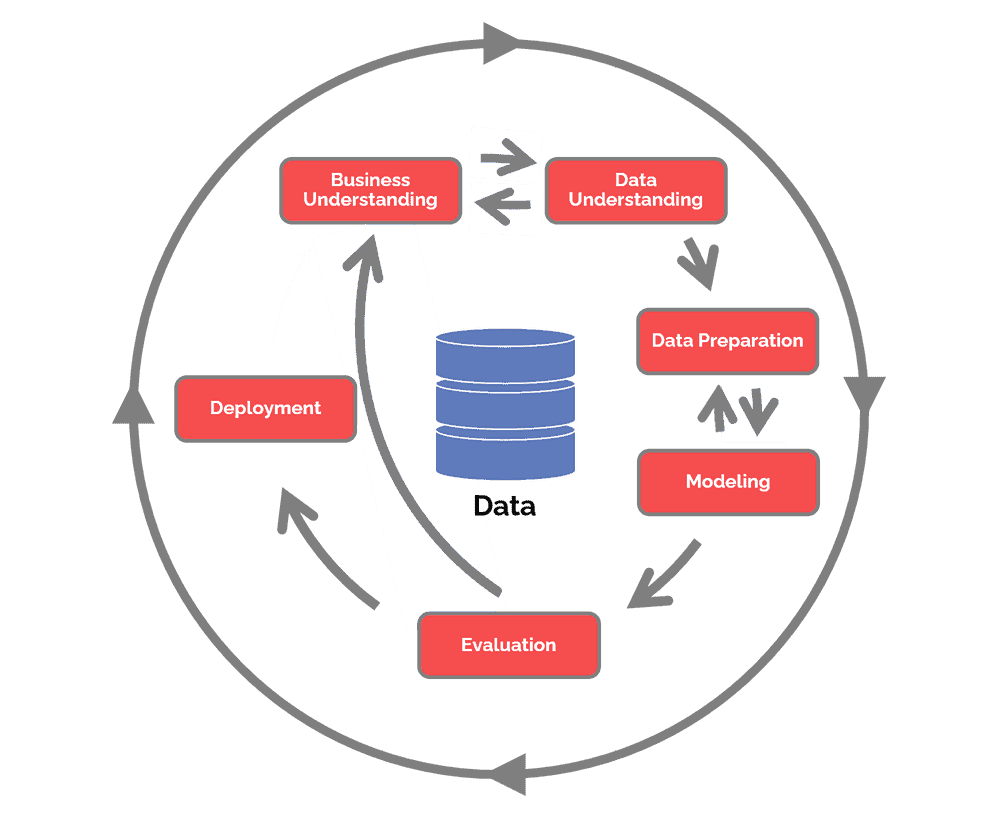
This module will provide an example of a use case – for example about National Data Analytics Solution (NDAS)

The National Data Analytics Solution (NDAS) is a national project, funded by the Home Office, developing the capability for advanced data analytics for policing. The use case utilises artificial intelligence (AI) to aggregate police information concerning events suspected to be modern slavery, events identified as potentially modern slavery and people associated with these events and potential events. It is planned that the insights generated by the AI interactive dashboard will be provided to a select group of end users, to inform strategic decision making through trend analysis as well as inform operational decision making through network creation and visualisation. A key component of the modern slavery use case is to bring already known modern slavery events and potential events together to provide a fuller picture of the modern slavery landscape and its victims.

Trilateral Research conducted a 3-month assessment to review the NDAS modern slavery dashboard for bias, fairness and transparency, data science accuracy and integrity and potential operational benefits to the end user.

The following method provided guidelines for the interviews and the assessment. The scope of the assessment was established by the assessment questions provided by the NDAS partners and included below.

The assessors employed a method following the **CR**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (CRISP-DM) while implementing ethical values into the assessment at every stage. This approach adopts a data analysis framework that is already familiar to many data analysts so that it can be more easily understood and followed by participants in a data analysis context. It involves 6 phases that naturally describe the data science life cycle. The details of the steps represented in the image are described in detail below.



(Image: Data Science Project Management. <https://www.datascience-pm.com/crisp-dm-2/> )

Bias and discrimination in AI can occur in four dimensions: data collection, training data, bias in algorithms, and bias in the interpretation of the results. In order to for the assessors to identify the existence or potential for bias at each phase, they considered the relevant data collection, data use, data interpretation and data ethics issues at the outset and throughout the data analytics process.

In addition, the assessment incorporated the principles identified by the EU high-level expert group on artificial intelligence (AI HLEG) to create human-centric and trustworthy AI, The CEPEJ European Ethical Charter on the Use of AI in Judicial Systems and Their Environment, adopted in 2018 and the considerations of the ALGO-CARE principle developed for the policing sector.

The AI HLEG’s seven key requirements are:

* Human agency and oversight
* Technical robustness and safety
* Privacy and data governance
* Transparency
* Diversity, non-discrimination and fairness
* Societal and environmental well-being
* Accountability

CEPEJ’s core principles to be respected in the field of AI and justice:

* Principle of respect of fundamental rights: ensuring that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights;
* Principle of non-discrimination: specifically preventing the development or intensification of any discrimination between individuals or groups of individuals;
* Principle of quality and security: with regard to the processing of judicial decisions and data, using certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment;
* Principle of transparency, impartiality and fairness: making data processing methods accessible and understandable, authorising external audits;
* Principle “under user control”: precluding a prescriptive approach and ensuring that users are informed actors and in control of their choices.

ALGO-CARE’s core principles:

* **Advisory**: Is the assessment made by the algorithm used in an advisory capacity? Does a human officer retain decision-making discretion?
* **Lawful**: On a case-by-case basis, what is the policing purpose justifying the use of algorithm, both its means and ends? Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes? In what way will the tool improve the current system and is this demonstrable? Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim? Is the operation of the tool compliant with national guidance?
* **Granularity**: Does the algorithm make suggestions at the right level of detail/granularity, given the purpose of the algorithm and the nature of the data processed? Do the benefits outweigh any technological or data quality uncertainties or gaps? Is the provenance and quality of the data sufficiently sound? Consider how often the data should be refreshed. If the tool takes a precautionary approach towards false negatives, consider the justifications for this.
* **Ownership**: Who owns the algorithm and the data analysed? Does the force need rights to access, use and amend the source code and data analysed? How will the tool be maintained and updated? Are there any contractual or other restrictions which might limit accountability or evaluation? How is the operation of the algorithm kept secure?
* **Challengeable**: What are the post-implementation oversight and audit mechanisms e.g. to identify any bias?
* **Accuracy**: Does the specification match the policing aim and decision policy? Can the stated accuracy of the algorithm be validated reasonably periodically? Can the percentage of false positives/negatives be justified? How was this method chosen as opposed to other available methods? What are the consequences of inaccurate forecasts? Does this represent an acceptable risk (in terms of both likelihood and impact)? Is the algorithmic tool deployed by those with appropriate expertise?
* **Responsible**: Would the operation of the algorithm be considered fair? Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review alongside other IT developments in policing? Would it be considered to be for the public interest and ethical?
* **Explainable**: Is appropriate information available about the decision-making rule(s) and the impact that each factor has on the final score or outcome? Is the force able to access and deploy a data science expert to explain and justify the algorithmic tool (in a similar way to an expert forensic pathologist)?

The assessors operationalised these principles via the following steps:

**Step 1.** *Inclusion of ethics requirements and ethical assessment of business objectives.*

To integrate ethics into the business understanding phase,the assessors tested the business objectives against the ethics requirements listed above. This included an assessment of whether any special issues (e.g., vulnerable populations, sensitive data such as medical data or biometrics, etc.) were likely to be involved, and if so, whether guidelines for these special issues are included. This step is needed to identify any possible tensions between the business objectives and ethics requirements.

**Step 2.** *Ethical assessment of data objectives*

In step 2, the assessors evaluated the data objectives against the ethics principles. This step is important because even if business objectives are compatible with ethics requirements, data objectives may be formulated in a way that is not compatible (e.g., data objectives could propose a segmentation of people into social categories that was not referred to in the business objectives and that does not fit well with principles of fairness and equality).

**Step 3.** *Stakeholder analysis (a) or involvement (b) in the business understanding phase*

Inclusion of ethical criteria in the development process benefits from a stakeholder analysis, in which direct and indirect stakeholders to the project are identified and their values and interests are assessed. This step makes it easier to identify more specific ethical requirements, make ethical assessments, and assess possible tensions between objectives and requirements and ethical criteria. Going further, stakeholders were interviewed about how they were involved in the development of the tool. During the duration of the project, the assessors engaged with users (virtually owing to Covid-19 travel restrictions) to validate the findings.

**Step 4.** *Ethical data collection and assessment*

To integrate ethical requirements into this phase of the assessment process, the assessors evaluated the data collection choice. At this stage, bias, discrimination, fairness and diversity, privacy, and data quality were particularly important. Data selection and fair algorithmic design are coupled with an ongoing ethical need to understand the historical and social contexts into which the modern slavery dashboard is being deployed. While definitions and statistical measures of fairness are certainly helpful, they cannot consider the nuances of the social contexts into which an AI system is deployed, nor the potential issues surrounding how the data were collected. As a result, the assessment process examined the role of human judgement in the decision-making loop to continuously assess whether the AI system has sufficiently minimised unfair bias so that it can be safely released for use. The assessment drew on many disciplines including data science, law, ethics and modern slavery domain expertise.

**Step 5.** *Ethical data description, exploration, and verification*

To integrate ethical requirements into the rest of the assessment process, the assessors evaluated the ethical consequences of describing, exploring and verifying the data. At this stage, issues relating to privacy, data quality, precision, accuracy, transparency, explainability, bias, discrimination, fairness and diversity were particularly important. The assessors examined the “gross” or “surface” properties of the acquired data (such as format and quantity) and evaluated whether the data satisfies the relevant requirements. Special attention was paid to data mining questions that concern patterns in the data (e.g., distribution of key attributes, relationships between pairs of attributes, properties of significant sub-populations, simple statistical analyses), through queries, visualisation and reporting techniques. Finally, the assessors examined the quality of the data, including completeness, correctness and missing variables (e.g., were more affluent neighbourhoods more easily accessible for data gathering?). Data description, exploration and verification methods taken by the police are subject to their standardised work and administrative processes.

**Step 6.** *Ethical assessment of data science techniques*

To integrate ethical requirements into this phase,the assessors examined whether ethical criteria are considered in the data science techniques used by the developer, and that the selection of these techniques are evaluated relative to these ethical criteria. Issues that were particularly relevant are those relating to false positives, transparency, safety and robustness, accuracy and lack of bias. For the NDAS modern slavery dashboard, the data science techniques used were text analysis using word embeddings and TfIdf vectorisor including SME input. The result of this process generated a list of keywords for the primary rules to be encoded.

**Step 7.** *Ethical assessment of outcomes*

An ethical assessment was performed on the results. To assess explainability of the full AI system, the assessors examined:

* To what degree the decisions and outcomes made by the system can be understood, including whether end users have access to the internal workflow of the model.
* Whether explainability is facilitated through technical aspects and if there are trade-offs between explainability and performance.

***Step 8.*** *End user analysis (a) or involvement (b) in the evaluation phase*

An end user analysis was performed. This guaranteed that their interests and values have been continuously taken into account and provided the information concerning potential operational benefits of the dashboard.

Conclusion: The partners we worked with on this project were very satisfied that with their work and the Trilateral recommendations for transparent AI.

**Reflection Sheet**

* At what stages of the technology’s lifecycle can we aim to address ethical concerns?]
* What are the benefits of an EIA?
* Why would an EIA assessment be important when we are designing technology to be transparent and fair?

**Task:**

* Write a three paragraph pitch to persuade a project manager that an ethical impact assessment is important. Show that you understand the benefits of an EIA, how technology can impact ethical values, and what an EIA entails.

* Write another three paragraphs about your own use case:
  + Identify a technology who needs a EIA;
  + Explain why they need this assessment ;
  + Explain if an EIA is still useful;
  + Potentially identify some ethical issues.

**Prereading:**

Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms. *Nature*.

Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O’Sullivan, D., & Gordijn, B. (2018). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. *Science and Engineering Ethics*, *24*(5), 1437-1481. <https://doi.org/10.1007/s11948-017-9961-8>

Wright, D., & Friedewald, M. (2013). Integrating privacy and ethical impact assessments. *Science and Public Policy*, *40*(6), 755-766. <https://doi.org/10.1093/scipol/sct083>

Royakkers, L., Timmer, J., Kool, L. et al. Societal and ethical issues of digitization. *Ethics Inf Technol* 20, 127–142 (2018). https://doi.org/10.1007/s10676-018-9452-x

<https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/>

What was Collingridge contribution to technology assessment?

1. He stated that technology is inherently political;
2. He stated that technology change how we view society;
3. He created a method to explain how values are embedded within the technology;
4. **He formulated a dilemma about when to address ethical concerns in emerging technology**

What is acute technosocial opacity?

1. The inability to identify ethical impacts from new and emerging technology
2. **The difficulty new technologies pose for people to shape their lives in meaningful ways**
3. The inability to find technological fixes for global societal issues
4. The want of developers and tech companies to hide the effects of technology on society

Scenarios to identify and address ethical impacts of technology are predominantly associated with the following methods:

1. **Ex-ante methods**
2. Intra methods
3. Ex-post

An ethical impact assessment is which of the following methods?

1. Ex-ante methods
2. **Intra methods**
3. Ex-post

If you want to embed ethical values in the design of the technology, which type of method is most suitable?

1. Ex-ante methods
2. **Intra methods**
3. Ex-post

What is not a step of an ethical impact assessment?

1. Consider if an ethical impact assessment is relevant
2. Identify stakeholder
3. Publish the report
4. **All are relevant**

What is not mentioned as a challenge of an ethical impact assessment?

1. **Publishing a risk assessment**
2. Finding experts
3. Identifying operational criteria
4. Participation of stakeholders

References

Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms. *Nature*. <https://www.nature.com/articles/d41586-019-03228-6#ref-CR1>

Nuffield Council on, B. (2015). The collection, linking and use of data in biomedical research and health care: ethical issues. <https://apo.org.au/node/249651>

Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O’Sullivan, D., & Gordijn, B. (2018). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. *Science and Engineering Ethics*, *24*(5), 1437-1481. <https://doi.org/10.1007/s11948-017-9961-8>

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121-136. <http://www.jstor.org/stable/20024652>

Worthington, R. (1982). The Social Control of Technology. By David Collingridge. (New York: St. Martin's Press, 1980. Pp. i + 200. $22.50.). *American Political Science Review*, *76*(1), 134-135. <https://doi.org/10.2307/1960465>

Wright, D., & Friedewald, M. (2013). Integrating privacy and ethical impact assessments. *Science and Public Policy*, *40*(6), 755-766. <https://doi.org/10.1093/scipol/sct083>