**Transparency – Intermediate**

**Module 1: Philosophical concepts of Transparency (1 lessons):**

Transparency is a value that can be important to Responsible AI within many different domains and for different reasons. This module allows learners to gain a better understanding of why transparency is important, how can ensure

We want to start by explaining what we mean with transparency. Transparency focuses on disclosing information. This information transparency can be captured and distributed through different media. (Turilli & Floridi, 2009) however, to be transparent requires more than just releasing information. Being transparent requires the disclosure of information as well as the ability of the people to whom this information is disclosed to understand and access the information that is being shared. Regarding algorithms, the shared information is often poorly accessible for different and sometimes valid reasons.(Mittelstadt et al., 2016) Companies may keep their proprietary algorithms secret to maintain their competitive advantage, in some cases we need to protect algorithms to protect other interests such as national security or privacy. Transparency can conflict with other ethical principles and values, but we will discuss that further on as well.

Being transparent is contextual and dependent on what information is being shared, how this information can be accessed, to whom this information is being shared and for what purpose and how it will help people make decisions. (Turilli & Floridi, 2009) Transparency is therefore highly dependent on the users of the shared information and the purpose for which that user will need to access the information. Similarly, the providers of information, those that make the algorithms transparent, are the ones that shape these factors, but they are not necessarily the only ones who determine the levels of transparency. Regulators are one groups who can influence what information an organisation needs to share, with whom and how. See for example the EU regulation around data protection law, the General Data Protection Regulation or GDPR which explains transparency requests around the processing of personal data.

a. Explaining the intrinsic or instrumental value of transparency

Transparency is not intrinsically valuable, i.e., we consider transparency valuable because of what we can achieve with it. We do not aim to improve transparency for the sake of transparency. So, if we state that improving transparency is desirable, we will need to clarify the purposes transparency is serving for it to be important. Transparency is not an ethical value or principle but it can be ethically desirable to promote transparency. It is therefore important to determine why we ought to be transparent. A software company should not reveal all the personal data it captures and stores about its users. It should not reveal personal data and protect it from third parties accessing it. However, they probably do need to explain that they protect personal data, what personal data they capture, how they share and distribute this personal data and what personal data they do not capture or share. Companies and people need to continuously decide what information they share, in other words, how transparent they are, and for what purposes they share information, and when they do not share information and for what purposes.

“Information transparency is not an ethical principle per se, seeing that it can be ethically neutral, but it can easily become an ethically ‘‘enabling’’ or ‘‘impairing’’ factor, that is a pro ethical condition, when the disclosed information has an impact on ethical principles.” (Turilli & Floridi, 2009)

The importance of transparency for ethics is two-way: (1) it’s a precondition for ethical principles (dependence); and (2) it is a way to examine how ethical principles are applied (regulation) “One is dependence: some amount of information is required in order to endorse ethical principles. The other is regulation: ethical principles regulate information flow by constraining its access, usage, dissemination and storage (see Fig. 1). Information transparency is ethically enabling when it provides the information necessary for the endorsement of ethical principles (dependence) or (and this might be an inclusive or) when it provides details on how information is constrained (regulation). (Turilli & Floridi, 2009)

Transparency, however, can be an important precondition for other principles. If an ethical principle would guide us into acting in an ethical way without a some form of information some ethical principles could not be followed. Turillia and Floridi give three examples of ethical principles that require some form of disclosure: accountability, safety, welfare and informed consent. We need to have some form of transparency to assign accountability. Without information about who is to blame or praise for what action we would not be able to assign accountability. A meaningful way to substantiate the principle of accountability requires information disclosure. Similarly, they mention the need to inform people about risks, for example, when there is a recall. Take for example, the Firestone Tyres that were recalled by Ford or the batteries by Samsung. Without information, the safety of people could not be protected. Safety is often enabled by informing people about risks.

“Accountability, safety, welfare and informed consent are examples of ethical principles that depend on the disclosure of some information in order to be endorsed. … the necessity of implementing information transparency into corporate governance because otherwise ‘‘accountability would not exist in any meaningful way’’ (Mallin 2002). Analogously, the recall of 13 million Firestone tyres by Ford in 2001 that were prone to spontaneous ‘blow-out’ (Moll 2003) and the recall of 10 million laptops by several makers because their batteries posed a fire hazard (Ahrens 2006) show how safety and welfare ‘‘enablement’’ may depend on the disclosure of information to the public” (Turilli & Floridi, 2009)

Transparency is also important to explain how ethical principles are adhered to. An example are privacy policies and informed consent forms. They disclose how information is protected. Further on, we will explain that this information can be important to establish and maintain trust, for example, when government bodies interact with the public. Information transparency is also important when explaining when and why you are not disclosing information for privacy, anonymity and copyrights, so the processes and limitations around information transparency.

“Privacy, anonymity and copyright are typical examples of ethical principles that regulate the flow of information. Information transparency enables such principles by providing the details of their regulative constraints to the public. Consent forms for the treatment of personal information clarify the extent to which privacy and anonymity will be granted,”

A false sense of information transparency can therefore be also ethically undesirable. The importance of adapting information disclosure to suit the context should therefore also be apparent. Distributing incorrect information or confusing people would be ethically undesirable.

“Conversely, ethical principles can be impaired if false details (misinformation) or inadequate or excessive amounts of information are disclosed.” (Turilli & Floridi, 2009)

“Information transparency may impair all of the above ethical principles when false, misleading, partial or inappropriate details are released to the public.” (Herzog, 2019)

b. Philosophical concepts of transparency (e.g., transparency as democratic value, trust, communicating trustworthiness)

The instrumental value of transparency also shows in its usefulness as a democratic value. Transparency is an enabling factor for democracy. It helps with public decision-making. Transparency is necessary for public debate and will formation, i.e., it helps people to decide what they want as a society, what they think is important, and how they want this achieved. It is one of the preconditions for participation in the political arena. It helps determine accountability when judging government actions and can be used to reward or punish politicians for their actions.(Buijze, 2013)

“The availability of information can fuel the public debate, and helps with the process of will formation.13 In other words, it facilitates public decision making. It enables citizens to determine what they, as a society, want to do. In addition, transparency is considered a necessary, albeit not sufficient, condition for participation,14 which allows people to exert influence on different types of governmental activities,15 and for accountability,16 which ultimately allows people to judge government actions and attach consequences to them.17” (Buijze, 2013)

Transparency has another function. It helps control organisations. Transparency helps us observe organisations, and see if they are doing what they should be doing and if they are following democratic processes and observe the rule of law.

The second function of transparency is that it makes it possible to observe from the outside what organisations are doing. This enables us to see what the government is doing on our behalf. This can be considered valuable in itself,60 but it is also the first step to ensure that public authorities are actually doing what they should be doing, that is, executing democratically agreed upon policies while observing the rule of law. (Buijze, 2013)

Transparency can contribute to the observance of the rule of law and promote integrity among public officials. The mere fact that officials know they are being watched, and that the quality of their work can be checked, is thought to improve their performance.3 (Buijze, 2013)

Additionally, transparency and democracy are somewhat self-supporting: a democracy strives towards more transparency while transparency helps to sustain democracy. This is partly explained by economic voting, i.e., transparency around economic conditions determines how people vote within a constituency. (Hollyer et al., 2011) Similarly, it helps economic decision-making, which is dependent on the availability of good information.

“Democracy and transparency are in a way self supporting, i.e., democracy have a tendency to strive towards more transparency and transparency sustains democracy: “particularly the phenomenon of economic voting—drive democracies towards greater openness.”(Hollyer et al., 2011)

transparency is argued to increase economic performance and market efficiency.36 As economic decision making is dependent on the availability of information, transparency facilitates good decisions.37 (Buijze, 2013)

Transparency and how we shape transparency can be biased. Governments can choose what information they make available, its quality, how accessible this information is, and what type of information they not share. The information they provide is often a function of their interests. The information is not by accident provided to people, but serves a specific purposes. The availability and quality of information available is often a political choice. Democratic government in contrast with autocracies are more likely to have a higher quality information available.

“Despite electoral incentives towards obfuscation, democratic governments are more likely to release policy- relevant data than are autocracies. … Whether information about a country is available is no accident. The availability, precision, and quality of data is driven, in part, by political institutions.” (Hollyer et al., 2011).

Another relation is the one between trust and transparency. One assumption is that transparency increasing legitimacy and trust. Because citizens have information they are more likely to accept decisions and enforcement costs will be lowered. (Buijze, 2013) However, this relation is often debated. Evidence that establishes this link is weak in the case of EU institutions.

Transparency is often thought to increase the legitimacy of the EU institutions as well as the trust that EU citizens have in them.22 This in turn improves the efficacy of the institutions, as people are more inclined to accept their decisions, and enforcement costs will be lowered.23 Whether transparency actually has these effects is debatable. At the very least the conditions under which it does so are ill-understood. (Buijze, 2013)

Transparency alone and out of context is not enough to establish trustworthiness or trust. Transparency is one thing required to establish trust but it is by itself not enough to establish trustworthiness or trust. Out of context, transparency is a way one communication between an entity, developers, public body or corporation. They share information and release information. But to establish trust and trustworthiness we also need other characteristics, such as honesty, competence and reliability. Things that are only established over time. Transparency is one part needed to make it easier to trust someone or something. That is not to say that transparency is not useful. Onora O’Neill calls it an antidote against secrecy and an incentive to meet legal or required standards, a way to force or nudge people to develop or use things in a social, legal or maybe even ethical way.

“Transparency requirements may fail to improve either trustworthiness or trust because they set a one-sided standard for public, corporate, or other communication. Although transparency demands too little for effective communication, it is an effective antidote to secrecy.” (O’Neill Transparency and the Ethics of Communication)

“To judge trustworthiness, we need to judge honesty, competence, and reliability. Meeting honesty, competence and reliability standards cannot be achieved merely by relying on individual choice combined with legal and regulatory constraints, nor is transparency enough for judging trust- worthiness: transparency is only a matter of putting content in the public domain. Transparency has the advantage of creating certain incentives to meet required or expected standards, but that is not enough—all too often the content is not accessible, intelligible, or assessable to those with less time or knowledge.” [O’Neill](https://www.thebritishacademy.ac.uk/documents/2563/Future-of-the-corporation-Trust-trustworthiness-transparency.pdf)

1. Discussion of the importance of transparency within different domains of application (e.g., healthcare, democracy, financial sector)

Information transparency in health care is a pressing legislative and regulatory issue. Transparency is a precondition for trust (to examine how ethical principles are applied by the party that needs to be trusted) and informed consent within healthcare (to allow for autonomous decision-making): (1) for patients to decide whether to undergo a treatment, (2) to participate in medical research, (3) to establish a relationship of trust between providers of treatment/diagnosis and patient and (4) to have their data used to improve policy and the health system (evidenced-based healthcare and learning health systems).

Transparency in healthcare and in other domains is not self-evident and different types of barriers exists:

1. Opacity from intentional secrecy

Secrecy can used to gain economically gain. Proprietary algorithms in healthcare can be intentionally kept opaque for companies to protect their informational advantage. Their concern would be that when they become transparent , they lose their competitive advantage.

Without regulatory intervention, medical technology providers can consolidate their informational power by applying algorithm secrecy as a means to intentionally generate opacity. (Technological Opacity of Machine Learning in Healthcare Herzog, Christian)

1. Opacity from technical illiteracy

“While patient compliance in medicine is an on- going issue, it stands to reason that patients with varying degrees of technological illiteracy may show ‘consent fatigue’ (Royakkers *et al.*, 2018), when confronted with algorithmic suggestions. Further, it may not be meaningful to require disabled and impaired people (e.g., when suffering from dementia) to consent to technology that, e.g., monitors them (Royakkers *et al.*, 2018) for lack of an understanding about the implications. (Technological Opacity of Machine Learning in Healthcare Herzog, Christian)”

1. Opacity from technological complexity

Sometimes technology might be too complex for physicians to understand the consequences, benefits and limitations of applying the technology to specific cases. In that case, a lack of transparency means that they cannot take full responsibility for the application of the technology.

“For a globalised, automated inference-based diagnosis, “physicians would have to work cooperatively with medical diagnosis devices and would need to understand the principles of the learning algorithms, be provided with transparent interfaces to be able to enter data with the appropriate quality and, hence, make use of automated diagnosis tools responsibly” (Char, Shah and Magnus, 2018). These tools could rely on data that can be biased (most of the medical data is acquired in intensive care situations, which is often not representative), incomplete or out-of-date, facts that could not be ap- parent to the medical practitioner, the development team or service provider and, least of all, the patient. Furthermore, the apparent superiority of some AI systems in clinical trials (Hannun et al., 2019; Topol, 2019) may lead doctors to refrain from questioning the validity of the computer models altogether and develop over-confidence in machine intelligence (Burrell, 2016). Practi tioners might not be able to take responsible action, in particular, when the data presentation is overwhelming and poorly interpretable. The process of knowledge generation and preserva- tion might become largely commercialized and potentially opaque to science or unavailable as educational resources. The distributed nature of medical data fusion and aggregation may further result in epistemic opacity.”

“Clearly, errors can never be completely avoided, but it appears as though popular algorithms in machine learning, such as deep learning, currently are susceptible to incorporating large epistemic gaps, i.e., a lack of scientific foundation in the modeling approach, that incurs opacity. (Technological Opacity of Machine Learning in Healthcare Herzog, Christian)”

Meaningful oversight and human intervention in algorithmic decision-making ‘‘is impossible when the machine has an informational advantage over the operator . . . [or] when the machine cannot be controlled by a human in real-time due to its processing speed and the multitude of operational variables’ (Mittelstadt et al., 2016)

iv. Opacity from techno-social interdependence

The idea that algorithms are commercially developed and that we may become dependent on these commercial algorithms might also mean that vendors of these algorithms will dictate secondary terms around access and distribution of the personal, medical data required to run the algorithm. The use of personal data by third parties may lead to a loss of control and oversight over medical data.

“It becomes apparent, however, that an ever more wide-spread use of automated analysis of medical data could even- tually force patients to consent to data sharing and opting-out of their rights to privacy, as oth- erwise they cannot be sure to receive the same quality of treatment (Char, Shah and Magnus, 2018). The fiduciary relationship between physician and patient may break entirely, potentially leaving patients without an adequate notion about the whereabouts of their data. (Herzog, 2019)

1. Opacity from application transdisciplinarity

Another contextual issue is that transparency is complicated by the fact that different professionals and people with different backgrounds may have different vocabularies, understand information different and attached different meanings to words.

“It is illustrative to compare the model semantics in papers written for technologists and physicians, which gives a hint on the difficulty to express mathematical expressions in, e.g., prose. This friction in trans- disciplinary research cannot be avoided, but it is important to be sensitive to the mutual opacities of the partners in a development team.” (Technological Opacity of Machine Learning in Healthcare Herzog, Christian)

vi. Opacity from technology-driven processes

Technology may make it more difficult to establish what the underlying motives are behind the use of the technology. The technology may promote certain interest and may fail to incorporate other.

“On behalf of the patient, this incurs procedural opacity, be- cause expectations might be to receive an all-encompassing treatment or diagnosis.. …. To counter this and at the same time provide an actual perspective in saving time for medical specialists, automated diagnosis technology should be designed as holistically as possible, which is more challenging to achieve than advertised. Char, Shah and Magnus, 2018, further warn that machine learning designers could be tempted to optimize for reimbursement rather than quality of care—a vision which lies at the intersection of opacity in processes and from complexity. (Technological Opacity of Machine Learning in Healthcare Herzog, Christian)

“In both respects, algorithmic processing contrasts with traditional decision-making, where human decision-makers can in principle articulate their ration- ale when queried, limited only by their desire and capacity to give an explanation, and the questioner’s capacity to understand it. The rationale of an algorithm can in contrast be incomprehensible to humans, rendering the legitimacy of decisions difficult to challenge.” (Mittelstadt The ethics of algorithms: Mapping the debate)

Another example of that transparency is not just sharing information but also examining how this information is interpreted and use is the financial sector

“disclosing only ethical or professional codes, recordings or summaries of activities, minutes or meeting reports diminishes when compared to the effects of making transparent also the details about how such information has been produced, elaborated and interpreted. In this way, organisations, companies or public institutions cannot limit their ethical involvement to public declarations of intent but have to show how the ethical principles, to which they are committed, are prioritised and translated into practise and governance.” (Turilli & Floridi, 2009)

1. Discussing the value of transparency in Responsible AI

In relation to responsible AI, transparency plays a role at different stage and in different ways balancing the interests of different stakeholders. During the development stage, transparency depends can play a role when deciding what algorithmic models use and with what datasets to train an algorithms. How can the AI be designed in a way to provide sufficient contextual information, so that its output is comprehensible, accessible and assessable to the people who need to be able to comprehend and weigh the AI’s recommendation. During the implementation and use, documentation about the algorithm – the purposes, its accuracy, its intended uses, its benefits and its risks - and how it ought to be used should be transparent. As the technology might still being developed not all relevant information of the technology will be available. Developers may have intended effects of the technology in mind but

1. Examining Responsible AI frameworks including the High-Level Expert Group on AI ‘Ethics Guidelines for Trustworthy Artificial Intelligence that discuss transparency.

Transparency has been a focus of EU guidelines for ethical AI. In 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence. The guidelines were formulated after public consultations.The guidelines identified three aspect for AI to trustworthy. It needs to be:

“(1) lawful - respecting all applicable laws and regulations

(2) ethical - respecting ethical principles and values

(3) robust - both from a technical perspective while taking into account its social environment”

One of the ethical principles relevant to algorithms is the principle of explicability. Transparency and this principle are closely related. Algorithms need to be transparent, including its purposes and outcomes, and explainable to people directly and indirectly affected by its outputs. According to the guidelines, this is important so people can contest its outputs. In case its outputs cannot be made transparent, the algorithm for example is too complex, other measures are required. For example, we need to be able to audit the algorithm or its capabilities need to be clearly explained. The EU guidelines mention that the degree of explainability is dependant on the contextual factors including the potential effects of applying the algorithm. Transparecy

Based on the ethical principles the guidelines identified seven requirements for an algorithm to be trustworthy. Transparency is one of the requirements. Transparency includes the data, the algorithm and the business models. The transparency requirement is divided into three concepts: traceability, explainability and communication. Traceability is about documenting the dataset used to train the algorithm, and the processes behind algorithmic decision-making including data labelling and data gathering. Documentation allows for the identification of errors and the auditing of the system.

Explainability concerns both the technology as well as how it is used to make decisions. The decisions by the system need to be understood and traced by human beings. If necessary, it might be desirable to increase explainability even if it might negatively impact accuracy. Sometimes the ability to audit and check the output of the algorithms is important, for example, when it concerns decisions that can have significant impacts on someone or society. The explanation should be adapted to the relevant person. Besides the algorithm, also how the algorithm influences decision-making should be explained. This is what the EU guidelines call the business model. A third aspect of transparency is communication. People should be informed when they are dealing with an algorithm, for example, when communicating with a chatbot. Users of algorithms should be informed about its limitations including its level of accuracy.

The UK government (Central Digital and Data Office) published in 2021 government a standard for algorithmic transparency, which was the first in its kind. The standard provides a template and a space for public sector organisations to share information about the algorithms they are using and why they are using them. One of the main aims of using this standard is showing how algorithms support decision-making by these public institutions in a way accessible to the wider public.

“The Algorithmic Transparency Standard helps public sector organisations provide clear information about the algorithmic tools they use, and why they’re using them.

Algorithmic transparency means being open about how algorithmic tools support decisions. This includes providing information on algorithmic tools and algorithm-assisted decisions in a complete, open, understandable, easily-accessible, and free format.”

One of the benefits of this standard is considered the previously mentioned argument of legitimacy: the transparency or explaining policing decisions helps increase support for policing decisions. The transparency helps improve people’ acceptance of decisions substantially, i.e. the decisions actually made, as well as procedural fairness, i.e., how the decisions that were made came about, even when the decisions are unfavourable to those affected. This transparency, for example, played a role during facial recognition.

“Studies of how people evaluate the legitimacy of the police in the UK5 and internationally6 similarly show that transparency (or explaining policing decisions to demonstrate trustworthy motives) enhances the perceived legitimacy of the police and acceptance of decisions as procedurally fair, even if those decisions are unfavourable. Transparency can improve trust in the motives of the police and enhances legitimacy and normative acceptance of police directives.7 For instance, a recent study in the UK found that trust and legitimacy are essential for public acceptance of live facial recognition technologies used by the police.” (Oswald 2022 The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives)

Transparency relates closely to police accountability for data practices that could affect human rights, particularly rights in the following areas:

* privacy;
* fair trial;
* protection from discrimination;
* liberty and security;
* freedom of assembly;
* and the capacity of the tools to entrench biases experienced by historically disadvantaged communities.(Oswald et al., 2022)

The standard consist of two tiers ([more info here](https://www.gov.uk/government/news/uk-government-publishes-pioneering-standard-for-algorithmic-transparency)):

1. tier 1 information, which contains a broader overview and a short non-technical description of your algorithmic tool, and an overview of what the tool is and why the tool’s being used
2. tier 2 information, provide more detailed technical information, such as specific details on how your tool works and the data the tool uses

Not all algorithms need to conform to the Algorithmic Transparency Standard. Eligibility criteria from [here](https://www.gov.uk/guidance/provide-information-on-how-you-use-algorithmic-tools-to-support-decisions-pilot-version). The tool either needs to be used to interact with the public or have an impact on the public while it is complex and influences the decision-making of the government body using the algorithm.

Included algorithms:

* need to engage directly with the public - for example a chatbot

OR:

Contain complex statistical analysis, complex data analytics or machine learning - for example neural nets or deep learning

AND

has a potential legal, economic, or similar impact on individuals or populations; affects procedural or substantive rights; affects eligibility, receipt or denial of a programme - for example receiving benefits

AND

replaces human decision making

assists or adds to human decision making - for example it provides evidence for decisions

**EUNOMIA USE CASE**

Text from D2.7

EUNOMIA

The COVID-19 pandemic raised afresh the challenge on social media information trustworthiness which threatens democracy, the cohesion of society as well as public health. In fact, the World Health Organisation (WHO) has highlighted the urgent need to address the risks emerging from the spread of misinformation around COVID-19 contextualising it as an Infodemic17. By analogy with the spread of a disease, this concept captures the fast spread of information, both accurate and inaccurate, through a large-scale and far-reaching community of users. Due to the vast volume of information and the mix of messages spread, true, false, rumours, facts, it becomes very difficult for the users to assess which information to trust. The Infodemic is a complex phenomenon, and its impact confirms the importance of relevant projects such as EUNOMIA. The EUNOMIA project offers a decentralised and open-source solution to social media users to raise awareness and promote active participation in the trustworthiness verification process. EUNOMIA contributes to media information literacy aiming to empower users to critically assess online information and in a medical analogy to adopt an information hygiene. Information hygiene here is defined as the systematic evaluation of the information consumed and spread to protect oneself and their network of contacts against misinformation. Similarly, hygiene is not about actively looking for pathogens/false information like fact-checkers do, but about routines that reduce the risk of infection/misinformation on a daily basis.

There are plenty of fact-checking guidelines regularly recommended to social media users to assist the online information assessment and limit the spread of misinformation such as "check the source" and "check whether the account is a bot". Even though such guides are indeed useful, they are rather problematic for the users to implement. They are often vague, time-consuming, or just difficult to adopt as part of one's routine. For example, while we are all aware of the risks of bots, the average social media user does not have the toolset to know if an account is a bot or not. Hence, EUNOMIA develops tools to support users to follow information hygiene guidelines.

To start with, when a user assesses a post, EUNOMIA informs whether similar information has been posted before. In this case, EUNOMIA visualises how similar information has been modified between different users' posts in an information cascade. In this cascade, the user can also see whether the context within which the information is provided has also changed. For example, posts including similar information might have different sentiment. Furthermore, EUNOMIA provides a trustworthiness nudge that pulses when the AI to be personalised by the user suggests a closer consideration of the user-driven indicators that can support their assessment. Indicators can signal for example bot activity, such as the ratio of followers to following.

Furthermore, EUNOMIA users are encouraged to act as ‘trust’ sensors in their network, actively slowing down the spread of misinformation. EUNOMIA enables users to vote on content trustworthiness and act as a trust-reference in their network. The number of votes appears as an additional indicator so to assist trustworthiness assessment of online information.

The EUNOMIA social media digital companion is available for both mobile and desktops and has been developed and tested on decentralised social media due to the privacy-first outlook of their communities, their open-source and transparent nature. EUNOMIA's technologies are tested in specifically created new instances of Mastodon with users participating both in the design and evaluation stages.

Method

Foresight scenarios were used in the EUNOMIA activities to identify potential ethical, social and privacy issues. Following an approach inspired by techno-ethical scenario building, scenarios have been developed in the first PIA+ report to raise awareness on privacy, ethical and societal impact mainly among the partners. These scenarios described how EUNOMIA solution could be employed to address an issue related to online trustworthiness assessment with a dystopic twist based on potential misuse. They were effective and playful stimuli to engage participants in workshops and actively discuss related issues and identify appropriate mitigation actions.

During the project’s lifecycle, internal PIA+ workshops have been organised mainly during or after consortia meetings to raise awareness amongst partners around potential privacy, ethical and social challenges emerging from the EUNOMIA project and solution. These workshops enabled a close collaboration between the consortium members and were a means of monitoring the development process so as to better understand the technology and intervene in key stages. Follow-up questionnaires were sent to the technical partners of the project allowing them to reflect on the issues discussed during the workshops and support an effective and clear communication. The recommendations resulting from the ongoing PIA+ assessment were presented. The final workshop discussed recommendations resulting from both user’s requirements and privacy, ethical, and social impact assessment were discussed also considering GDPR principles.

Ethics analysis

The overall impact of the project has been identified as a positive one, aiming to the general well- being of users on social media and improving their awareness of and skills towards trustworthiness assessment of online information.

The research team then identified key ethical principles and societal values related to the domain. These were assessed against the impact to understand when and how potential impacts could both support the benefits or lead to unintentional harms. From this assessment, partners were able to identify how they could intervene in key stages of design and development, suggesting remedial actions and mitigation measures, solving tensions between conflicting principles, and identifying actions that can support beneficial trajectories.

The main ethical principles relevant to the EUNOMIA project and tools identified in the first reporting, and still standing throughout the project, were autonomy, dignity, and privacy.

The notion of autonomy is directly related to liberty and they are protected by Article 6 of the European Charter of Fundamental Rights as well as Article 3 of the United Nations Universal Declaration of Human Rights

Human dignity is also a fundamental value protected by Article 1 of the European Charter of Fundamental Rights. Similarly, Article 1 of the United Nations Universal Declaration of Human Rights refers to equal treatment of humans in terms of dignity and rights.

Privacy is seen as an umbrella term that enables society to flourish24. Related to autonomy, the European Convention of Human Rights relates privacy to “respect for private and family life, home and communications”. Privacy is a key principle to be considered in the EUNOMIA project related to the societal considerations. Indeed, privacy is a prerequisite for the operation of democracy and therefore, ‘group privacy’ (Floridi 2014) is also of high

importance.

Considering these issues and the impacts, the societal considerations relevant to the EUNOMIA project were: discrimination, equality, fairness, and impartiality as well as bias. This preliminary assessment was engaged throughout the early stages of the project to raise awareness amongst the consortium on potential privacy, ethical, and social legal issues

against the positive impact of the project.

As such, early on in the project, design decisions clearly considered these issues. For example, the initially planned user reputation scoring feature was discarded as it was identified as high risk potentially resulting in chilling effect. Instead, design shifted the focus on the trustworthiness scoring of the post itself. The initial PIA+ report (D2.1) provided respective recommendations and contributed to the positioning of the project.

MCQs

1. What is applicable to transparency?:
   1. Transparency needs to be maximised at all times.
   2. Transparency is always valuable.
   3. **Transparency is ethical neutral.**
   4. Transparency is one of the most important ethical principles.
2. What is not applicable to transparency?:
   1. Transparency is a precondition for some ethical principles.
   2. Transparency is important to show how we follow certain principles.
   3. **Transparency cannot be limited by ethical principles.**
   4. Transparency needs to be adjusted to the context in which it is required
3. Which statement is true?:
   1. Transparency is an instrument to decrease government power
   2. Transparency is an instrument for governments to exert political influence
      1. **A and B can both be correct**
      2. A is correct
      3. B is correct
      4. A and C are both incorrect
4. According to the the High-Level Expert Group on AI ‘Ethics Guidelines for Trustworthy Artificial Intelligence’, trustworthy AI requires the following:
   1. **Lawful, ethical and robust AI**
   2. Increased legitimacy of public bodies applying AI
   3. Increased transparency around AI decision-making
   4. A sound technical approach inclusive of its social environment
5. What was one of the main reason behind developing the UK Algorithmic Transparency Standard?
   1. Support data protection legislation
   2. **Increase legitimacy for public bodies**
   3. Promote algorithmic expertise
   4. Educate about the effects of complex machine learning algorithms
6. What is not a condition to determine if the UK Algorithmic Transparency Standard needs to be applied?
   1. **Affects intellectual property rights**
   2. Affects substantive or procedural rights
   3. Support or replaces human decision-making
   4. Contain machine learning algorithms

* **Reading – both pre-reading and course reading**
* Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, *11*(2), 105-112. <https://doi.org/10.1007/s10676-009-9187-9>
* Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. <https://doi.org/10.1177/2053951716679679>

Course reading

* <https://techcrunch.com/2018/05/10/duplex-shows-google-failing-at-ethical-and-creative-ai-design/>
* <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>
* Algorithmic Transparency Standard <https://www.gov.uk/government/collections/algorithmic-transparency-standard>
* Oswald, M., Chambers, L., Goodman, E., Ugwudike, P., & Zilka, M. (2022). The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4155549>
* Reflection worksheet

Reflection Worksheet

Q&A:

1. Why is transparency an instrumental value and why is it important?
2. What do we mean when we say that transparency is contextual?
3. What are three forms of opacity?

Task:

Make notes to give a five-minute talk explaining to developers why their AI for policing needs to be transparent. Describe what transparency considerations are important and why.

Read the Google Duplex case study and answer these questions: How did Google approach transparency? How could the backlash have been avoided?

Buijze, A. (2013). The Six Faces of Transparency. *Utrecht Law Review*, *9*, 3. <https://doi.org/10.18352/ulr.233>

Herzog, C. (2019). *Technological Opacity of Machine Learning in Healthcare* Proceedings of the Weizenbaum Conference 2019 "Challenges of Digital Inequality - Digital Education, Digital Work, Digital Life", Berlin.

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2011). Democracy and Transparency. *The Journal of Politics*, *73*(4), 1191-1205. <https://doi.org/10.1017/s0022381611000880>

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. <https://doi.org/10.1177/2053951716679679>

Oswald, M., Chambers, L., Goodman, E., Ugwudike, P., & Zilka, M. (2022). The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4155549>

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, *11*(2), 105-112. <https://doi.org/10.1007/s10676-009-9187-9>