DOME Customer Support Knowledge Base - Ethics

Fundamentals for the DOME Customer Support Knowledge Base regarding Ethics. This document set the basis for the further, more detailed Ethics part of the knowledge base, which will be implemented according to the specific characteristics of DOME, as they will be estanlished during the project implementation phases.

- List of relevant documents and resources
- Ethics Management Plan
- AI and Ethics issues in DOME
- Ethics by design

List of relevant documents and resources

List of relevant documents and resources

? Ethics By Design and Ethics of Use Approaches for Artificial Intelligence Version 1.0, European Commission, 25 November 2021

? Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence set up by the European Commission, of 8th April 2019

? Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts COM/2021/206

? Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)

? Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act)

? Convention for the Protection of Human Rights and Fundamental Freedoms, in particular, its Articles 7 and 8.3

? Charter of Fundamental Rights of the European Union, specifically Article 8 concerning the protection of personal data.

Ethics Management Plan

Introduction – DOME Ethical aspects

The DOME solution raises Ethics issues regarding three different aspects:
1. Involvement of human participants;
2. Processing of personal data; and
3. Use of Artificial Intelligence (AI).

Involvement of Human Participants

The involvement of human participants in the DOME activities is very limited, and regards the participation of adult volunteers, capable of providing a valid (informed) consent to their participation to DOME, in requirements assessment and feedback consultations and surveys, and similar tasks. As such, their participation does not require any kind of authorization by Ethics authorities or committees, and is allowed from an Ethical, personal and fundamental rights viewpoint provided that they receive an adequate information sheet, customised to the specific activity in which they are involved, and sign a proper informed consent form.

Requirements, needs, and feedback will be collected mainly from the stakeholders members of the Consortium, most of all in the first phases; therefore, their participation in those tasks is part of their contribution to the solution implementation, and is covered by the project Agreements and internal policies. However, additional information could be collected also by external stakeholders and users.

External stakeholders' personal data will be processed following the rules and procedures described in the DOME Data Management Plan and connected procedures.

a) Criteria and procedures to identify and recruit external participants

The DOME Consortium is committed to complying with ethical principles and applicable international, EU and national law. The Consortium ensures respect for people and for human dignity, protecting the values, rights and interests of the research participants and avoiding any bias.

The responsibility for recruiting research participants ethically lies with the DOME task leaders undertaking the task. Task leaders can approach the PEO and the internal Ethics Helpdesk for advice and assistance during the selection process.

The partners will advertise the DOME activities in an open and transparent manner to recruit the volunteer stakeholders. In practical terms, this will likely involve, but is not limited to: participation to events, workshops and conferences; announcements / adverts through established communications channels (e.g. circular emails); meetings offering an opportunity for Q&A with DOME partners; and direct engagement based on knowledge from existing networks, whether through email or phone.

When conducting any surveys, questionnaires, workshops or webinars where personal information is gathered and stored, the partners will pay attention to privacy, data protection, and data management. External stakeholders' personal data will be processed following the rules and procedures described in the DOME Data Management Plan and connected procedures.

b) Informed Consent Procedures

Informed Consent is a voluntary agreement to participate in specific DOMEt activities such as workshops,

interviews, and surveys, based upon an informed and free decision. Informed consent will be sought in DOME from all external human participants involved in the tasks and activities, as collection of feedback. Regarding partners' staff directly working on DOME (as e.g. staff from organisations/companies member of the DOME Consortium), it is not necessary to seek additional informed consent since their participation is already included in their contractual obligations, and they are fully aware of DOME, its goals, and the activities to be carried out.

Obtaining consent involves first informing the research participants about their rights, the purpose of the project, the procedures that will occur, and the potential risks and benefits of participation.
Before requesting consent, the staff in charge of the activity must make sure that the potential participant has received written and, as appropriate, verbal information. This information must be provided in such a way that the potential participant understands the contents. This means drafting the DOME information sheet in accessible language. The Consortium partners' staff will assure that the participants will have the opportunity to ask questions about DOME and receive detailed and appropriate answers.

Human participants will be provided with adequate information on DOME and their involvement in it. In order to help partners in this process, a DOME Information Sheet and a Consent Form template for partners to give to external participants involved in DOME will be provided. The templates can be adjusted and customised if necessary and appropriate.
The participants will be informed that their personal data will remain confidential and properly protected. Informed consent for participation must be signed, dated and kept on file.

The consent of the data subject is one of the available legal grounds for the processing of personal data under the General Data Protection Regulation (Regulation (EU) 2016/679) 'GDPR', Article 6.
As such, informed consent for human participation in DOME activities overlaps with, but is distinct from, consent for the processing of personal data. In many cases, participating in a DOME activity will also involve the processing of personal data (for example, processing names and contact details to arrange a workshop or technology testing). However, these two consent requirements are conceptually and legally distinct. Owing to the mentioned significant conceptual overlapping, it is usually clearer and more practical if these two types of informed consent are dealt with together within the same information sheet and informed consent form.

'Consent' under the GDPR, Article 4 is defined as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her".

For consent to the processing of personal data to be 'informed', the data subject must similarly be provided with detailed information about the envisaged data processing in an intelligible and easily accessible form, using clear and plain language.

The Article 7 of the GDPR indicates the exact conditions required for a valid consent:

1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.
2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.
3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.
4. When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

The informed consent should also include the prior provision of information to the data subject on the data subject's rights as guaranteed by the GDPR and the EU Charter of Fundamental Rights, in particular the right to withdraw consent or access their data, the procedures to follow should they wish to do so, and so forth.

As required by the Article 13 GDPR, the information sheet accompanying the informed consent form should

contain in an accessible way information on the following data subject rights:
• The right of access by the data subject (Article 15 GDPR);
• The right to rectification (Article 16 GDPR);
• The right to erasure (‘right to be forgotten’), including the withdrawal of consent (Article 17 GDPR);
• The right to restrict the processing (Article 18 GDPR);
• The right to data portability (Article 20 GDPR);
• The right to lodge a complaint with a supervisory authority (Article 57 GDPR).

Use of Artificial Intelligence

In the development of all AI DOME models, partners fully respect and comply with rules, regulation and recommendations regarding the use and implementation of AI tools, first of all specifying the process for how each model will be evaluated within DOME and therefore the method for ascertaining whether it is able to justify the results.
Specific attention will be given to data processing, since data, and usually personal data, are a key ingredient for AI. The Article 22 of the GDPR explicitly prohibits decisions affecting a data subject solely based on automated decision-making, unless authorised by a Union or Member State law with suitable safeguards (Article 22(2)(b)) or by explicit consent (Article 22(2)(c)). Data subjects also have the right in such cases to obtain from the data controller under Article 14 of the GDPR, “meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.” Of course, other important aspects of AI linked to other regulations beyond the protection of personal data will be taken into consideration as well in subsequent versions of this document.
A detailed analysis and description of the rules, legal basis, principles, procedures, and methodologies through which Ethics issues related to AI will be addressed in DOME is provided in next Subsection.

AI and Ethics issues in DOME

AI and risk assessment – Legal approach to AI – related risks

The analysis, assessment, and management of AI in DOME (including NLP, machine learning, and Ethics-by-design in AI) are based on the guidance provided by the Ethics Guidelines for Trustworthy AI and the Ethics By Design and Ethics of Use Approaches for Artificial Intelligence document of the EU Commission.
As clearly stated in its introduction, the Ethics By Design and Ethics of Use Approaches for Artificial Intelligence note concerns all research activities involving the development or/and use of artificial intelligence (AI)-based systems or techniques, including robotics. It builds on the work of the Independent High-Level Expert Group on AI and their mentioned Ethics Guidelines for Trustworthy AI as well as on the results of the EU-funded SHERPA and SIENNA projects.
The central approach of this note is based on Ethics by Design (see Subsection 1.3). The aim of Ethics by Design is to incorporate ethical principles into the development process allowing that ethical issues are addressed as early as possible and followed up closely during project activities. It explicitly identifies concrete tasks which can be taken and can be applied to any development methodology. However, the advised approach should be tailored to the type of activity being proposed keeping also in mind that ethics risks can be different during the research/design phase and the deployment or implementation phase.
Furthermore, from a practical viewpoint, the principles and methodology of the Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment and of the HRESIA (Human Rights, Ethical, and Social Impact Assessment) model and approach will be followed.
The European AI strategy places trust as a prerequisite to ensure a human-centric approach to AI, considering AI not an end in itself, but a tool that has to serve people with the ultimate aim of increasing human well-being.
In 2019, the European Commission created a High-Level Expert Group on Artificial Intelligence (AI HLEG), comprising representatives from academia, civil society, and industry to provide recommendations on future related policy development and on ethical, legal and societal issues related to AI, including socio-economic challenges. In April 2019, the AI HLEG presented the Ethics Guidelines for Trustworthy AI. According to that document, Trustworthy AI has three components, which should be met throughout the system's entire life cycle:
(1) It should be lawful, complying with all applicable laws and regulations,
(2) It should be ethical, ensuring adherence to ethical principles and values, and
(3) It should be robust; both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm.
Regarding the respect of indivisible rights set out in international human rights law, the EU Treaties and the EU Charter, the following families of fundamental rights are particularly apt to cover AI systems :
1. Respect for human dignity. Human dignity encompasses the idea that every human being possesses an “intrinsic worth”, which should never be diminished, compromised or repressed by others – nor by new technologies like AI systems. In this context, respect for human dignity entails that all people are treated with respect due to them as moral subjects, rather than merely as objects to be sifted, sorted, scored, herded, conditioned or manipulated. AI systems should hence be developed in a manner that respects, serves and protects humans’ physical and mental integrity, personal and cultural sense of identity, and satisfaction of their essential needs.
2. Freedom of the individual. Human beings should remain free to make life decisions for themselves. This entails freedom from sovereign intrusion, but also requires intervention from government and non-governmental organisations to ensure that individuals or people at risk of exclusion have equal access to AI’s benefits and opportunities. In an AI context, freedom of the individual for instance requires mitigation of (in)direct illegitimate coercion, threats to mental autonomy and mental health, unjustified surveillance, deception and unfair manipulation. In fact, freedom of the individual means a commitment to enabling individuals to wield even higher control over their lives, including (among other rights) protection of the freedom to conduct a business, the freedom of the arts and science, freedom of expression, the right to private life and privacy, and freedom of assembly and association.
3. Respect for democracy, justice and the rule of law. All governmental power in constitutional democracies must be legally authorised and limited by law. AI systems should serve to maintain and foster democratic processes

and respect the plurality of values and life choices of individuals. AI systems must not undermine democratic processes, human deliberation or democratic voting systems. AI systems must also embed a commitment to ensure that they do not operate in ways that undermine the foundational commitments upon which the rule of law is founded, mandatory laws and regulation, and to ensure due process and equality before the law.
4. Equality, non-discrimination and solidarity - including the rights of persons at risk of exclusion. Equal respect for the moral worth and dignity of all human beings must be ensured. This goes beyond non-discrimination, which tolerates the drawing of distinctions between dissimilar situations based on objective justifications. In an AI context, equality entails that the system's operations cannot generate unfairly biassed outputs (e.g. the data used to train AI systems should be as inclusive as possible, representing different population groups). This also requires adequate respect for potentially vulnerable persons and groups, such as workers, women, persons with disabilities, ethnic minorities, children, consumers or others at risk of exclusion.
5. Citizens' rights. Citizens benefit from a wide array of rights, including the right to vote, the right to good administration or access to public documents, and the right to petition the administration. AI systems offer substantial potential to improve the scale and efficiency of government in the provision of public goods and services to society. At the same time, citizens' rights could also be negatively impacted by AI systems and should be safeguarded. When the term "citizens' rights" is used here, this is not to deny or neglect the rights of third-country nationals and irregular (or illegal) persons in the EU who also have rights under international law, and – therefore – in the area of AI systems.

On the above mentioned groups of fundamental rights mentioned in the document, are rooted the main Ethical principles relevant in the Context of AI systems. Many of these are to a large extent already reflected in existing legal requirements for which mandatory compliance is required and hence also fall within the scope of lawful AI, which is Trustworthy AI's first component.

1. Respect for human autonomy

The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems. AI systems may also fundamentally change the work sphere. It should support humans in the working environment, and aim for the creation of meaningful work.

2. Prevention of harm

AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity. AI systems and the environments in which they operate must be safe and secure. They must be technically robust and it should be ensured that they are not open to malicious use. Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems. Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens. Preventing harm also entails consideration of the natural environment and all living beings.

3. Fairness

The development, deployment and use of AI systems must be fair. While we acknowledge that there are many different interpretations of fairness, we believe that fairness has both a substantive and a procedural dimension. The substantive dimension implies a commitment to: ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation. If unfair biases can be avoided, AI systems could even increase societal fairness. Equal opportunity in terms of access to education, goods, services and technology should also be fostered. Moreover, the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice. Additionally, fairness implies that AI practitioners should respect the principle of proportionality between means and ends, and consider carefully how to balance competing interests and objectives.31 The procedural dimension of fairness entails the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them.32 In order to do so, the entity accountable for the decision must be identifiable, and the decision-making processes should be explicable.

4. Explicability

Explicability is crucial for building and maintaining users' trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be

duly contested. An explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred to as ‘black box’ algorithms and require special attention. In those circumstances, other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate.
Based on the abovementioned fundamental rights and ethical principles, the Ethics Guidelines for Trustworthy AI set out seven key requirements that AI systems should meet in order to be trustworthy. The list of requirements is non-exhaustive, and it includes systemic, individual and societal aspects :
1. Human agency and oversight (including fundamental rights, human agency and human oversight): AI systems should support human autonomy and decision making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user’s agency and foster fundamental rights and allow for human oversight. In situations where AI may negatively affect fundamental rights, a fundamental rights impact assessment should be undertaken. Users should be able to make informed autonomous decisions regarding AI systems. Human oversight may be achieved through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL) or human-in-command (HIC) approach. ‘HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system’s operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation.’
2. Technical robustness and safety (including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility): A crucial component of achieving trustworthy AI is technical robustness, which is closely linked to the principle of prevention of harm. AI systems, like all software systems, should be protected against vulnerabilities that can allow them to be exploited by adversaries, e.g., by hacking. AI systems should have safeguards that enable a fallback plan in case of problems. Moreover, AI requires a high level of accuracy, which pertains to an AI system’s ability to make correct judgements, for example, to correctly classify information into the proper categories, or its ability to make correct predictions, recommendations, or decisions based on data or models. It is also critical that the results of AI systems are reproducible, as well as reliable.
3. Privacy and data governance (including respect for privacy, quality and integrity of data, and access to data): AI systems must guarantee privacy and data protection throughout a system’s entire lifecycle. The quality of the data sets used is paramount to the performance of AI systems. When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. Such issues need to be addressed prior to training any given data set. In addition, the integrity of the data must be ensured. In any organisation that handles individuals’ data (whether someone is a user of the system or not), data protocols governing data access should be put in place.
4. Transparency (including traceability, explainability and communication): The data sets and the processes that yield the AI system’s decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard to allow for traceability and an increase in transparency. This also applies to the decisions made by the AI system. This enables identification of the reasons why an AI-decision was erroneous which, in turn, could help prevent future mistakes. Traceability facilitates auditability as well as explainability. Explainability concerns the ability to explain both the technical processes of an AI system and the related human decisions (e.g., application areas of a system). Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. AI systems should not represent themselves as humans to users; humans have the right to be informed that they are interacting with an AI system. The AI system’s capabilities and limitations should be communicated to AI practitioners or end-users.
5. Diversity, non-discrimination and fairness (including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation): Trustworthy AI must enable inclusion and diversity throughout the entire AI system’s life cycle, ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with the principle of fairness. Trustworthy AI requires avoidance of unfair bias, accessibility and universal design and stakeholder participation. Data sets used by AI systems (both for training and operation) may suffer from the inclusion of inadvertent historic bias, incompleteness and bad governance models. This could lead to unintended (in)direct prejudice and discrimination against certain groups or people, potentially exacerbating prejudice and marginalisation. The way in which AI systems are developed (e.g., algorithms’ programming) may also suffer from unfair bias. This could be counteracted by putting in place oversight processes to analyse and address the system’s purpose, constraints, requirements and decisions in a clear and transparent manner. Moreover, hiring from diverse backgrounds, cultures and disciplines can ensure diversity of opinions and should be encouraged. AI systems should not have a one-size-fits-all approach and

should consider Universal Design principles addressing the widest possible range of users, following relevant accessibility standards. It is advisable to consult stakeholders who may directly or indirectly be affected by the system throughout its life cycle.
6. Societal and environmental well-being (including sustainability and environmental friendliness, social impact, society and democracy): Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as the UN's Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations. The effects of AI on people's physical and mental wellbeing, society and democracy must be carefully monitored and considered.
7. Accountability (including auditability, minimisation and reporting of negative impact, trade-offs and redress): Trustworthy AI necessitates mechanisms to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use. It requires (a) auditability, which entails the enablement of the assessment of algorithms, data and design processes, (b) minimisation and reporting of negative impacts through impact assessments used (e.g., red teaming or forms of algorithmic impact assessment) both prior to and during the development, deployment and use of AI systems, (c) trade-offs should be addressed in a rational and methodological manner within the state of the art, (d) when unjust adverse impact occurs, accessible mechanisms should be foreseen that ensure adequate redress.
In July 2020, the AI HLEG released a self-assessment list to help AI developers assess their tools against the Ethics Guidelines for Trustworthy AI . This list will guide all AI developers through DOME.
Finally, the Institute of Electrical and Electronic Engineers (IEEE), a global professional organisation working towards technology standards for human benefit, has a global Initiative on the Ethics of Autonomous and Intelligent Systems. In 2019, the IEEE released Ethically Aligned Design: A Vision for Prioritising Human Well-being with Autonomous and Intelligent Systems outlining a number of general principles to guide technology developers in the design and implementation of Autonomous and Intelligent Systems (A/IS) . These principles, that will be taken into consideration throughout DOME, are the following:
1. Human Rights. A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
2. Well-being. A/IS creators shall adopt increased human well-being as a primary success criterion for development.
3. Data Agency. A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.
4. Effectiveness. A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.
5. Transparency. The basis of a particular A/IS decision should always be discoverable.
6. Accountability. A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
7. Awareness of Misuse. A/IS creators shall guard against all potential misuses and risks of A/IS in operation.
8. Competence. A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

Ethics by design

Historical and Theoretical overview

The Ethics by Design approach will be used throughout DOME in addressing technical Ethics issues, and most of all those raised (or potentially raised) by AI, in accordance with the Ethics By Design and Ethics of Use Approaches for Artificial Intelligence notes and guidelines.
The adoption of the Ethics by Design approach, however, does not preclude additional measures to ensure adherence to all major AI ethics principles and compliance with the EU legal framework, in order to guarantee full ethical compliance and implementation of the ethical requirements.
Ethics by Design has been largely influenced by both Privacy by Design and Value Sensitive Design as an approach to a design process. Privacy by Design involves incorporating privacy concerns across the design process. A key part of the Privacy by Design methodology is the regular implementation of privacy design strategies to make technology development more privacy compliant. With Ethics by Design, there is the implementation of Ethical design requirements to try to make technology more Ethical, or more aligned with Ethical standards.
Value Sensitive Design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process .
In Value Sensitive Design, technology development can be analysed in terms of what is ‘good’ or ‘important’, as determined by stakeholders, and alternative design options can be developed. Designs are developed using an investigation consisting of three phases: conceptual, empirical and technological. These investigations are intended to be iterative, allowing the designer to modify the design continuously.
The process can lead to new design solutions that balance priorities from different values and mitigate harms. It is worth noting that Ethical values can be used as the main drivers of this type of design process.
Value Sensitive Design is subjected to criticisms most of all in relation to the adopted methodology. First, it is implemented on analysing a technology that already exists, developing/proposing alternative designs. However, if a harmful technology is already developed, this is negative per se, ethically speaking, whether or not that technology is actually used. To fulfil Ethical values it is preferable that harmful technologies are not developed at all. Furthermore, by focussing on what stakeholders define as ‘good’, there is little consideration of how the development of the technology aligns with guiding principles of what is ‘right’: something that is good for one person might not be the right thing to do if there are negative consequences for others. As such, Value Sensitive Design can be considered as a starting point, but additional features must be added to make the outcomes more Ethical.
As stated in the Ethics By Design and Ethics of Use Approaches for Artificial Intelligence , for many AI projects, the relevant ethical issues may only be identified after the system’s deployment (making it very useful for Value Sensitive Design), while for other projects these might be revealed during the development phase. Ethics by Design is intended to prevent ethical issues from arising in the first place by addressing them during the development stage, rather than trying to fix them later in the process. This is achieved by proactively using the principles as system requirements. What is more, since many requirements cannot be achieved unless the system is constructed in particular ways, ethical requirements sometimes apply to development processes, rather than the AI system itself.
The aim of Ethics by Design is to make the systems designers think about and address potential ethics concerns, while they are developing a system.

Ethics by Design requirements and principles in DOME
The planned development, implementation, and use of AI tools in DOME is actually limited to the adoption of a chatbot for customer service.
Ethics by Design tools and methodologies, as presented below, will be followed during the planning, development, and use phase of the envisaged chatbot, including aspects of NLP functionalities development and machine learning.
In the Ethics By Design and Ethics of Use Approaches for Artificial Intelligence , Ethics by Design is described with a five-layer model. This model is similar to many others in Computer Science: higher levels are more abstract, with increasing levels of specificity going down the levels.

In the above mentioned guidelines, Ethics by Design is premised on the basis that development processes for AI and robotics systems can be described using a generic model containing six phases, which can be considered parts of a sequential process or can be iterative or even incremental. By mapping the development methodology to the generic model used here, the relevant ethical requirements can be determined. Once this has been accomplished, the Ethics by Design will be embedded into the development methodology as tasks, goals, constraints and the like. The chance of ethical concerns surfacing is thus minimised because each step in the development process will contain measures to prevent them arising in the first place.

The six tasks in the generic model are:

1. Specification of objectives: The determination of what the system is for and what it should be capable of doing.
2. Specification of requirements: Development of technical and non-technical requirements for building the system, including initial determination of required resources, together with an initial risk assessment and cost-benefit analysis, resulting in a design plan.
3. High-level design: Development of a high-level architecture. This is sometimes preceded by the development of a conceptual model.
4. Data collection and preparation: Collection, verification, cleaning and integration of data.
5. Detailed design and development: The actual construction of a fully working system.
6. Testing and evaluation: Testing and evaluation of the system.

During the DOME implementation, whenever appropriate the partners will apply the generic model methodology, as explained in detail in the Ethics By Design and Ethics of Use Approaches for Artificial Intelligence guidelines, page 13/22.