Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable:

D5.1 Initial Ethical Framework

Dissemination Status: Public

Editor: Mišo Mudrić (ISP)

Authors: Joe Levy (CINEDIT), Jaanus Pikani (BMA), Sachin Gaur (BMA), Michelangelo Ceci (CINI), Paolo Mignone (CINI), Constatino Mele (CINI), Bruno Bonomini (CPAD), Giulia Canilli (CPAD), Berta Biescas (INS), Joaquín Luzón Tuells (INS), Eirik Bærulfsen (OSL), Lars Grottenberg (OSL), Ian Simon Gjetrang (OSL), Ravishankar Borgaonkar (STF), Stine S.Kilskar (STF), Johan de Heer (THA), Rafal Hrynkiewicz (THA), Alberto Da Re (UNI), Paolo Mocellin (UPAD), Matteo Bottin (UPAD), Krunoslav Katić (ISP), Jelena Radošević (ISP), Filip Dragović (ISP), Mišo Mudrić (ISP)

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

About IMPETUS

IMPETUS (Intelligent Management of Processes, Ethics and Technology for Urban Safety) is a Horizon 2020 Research and Innovation project that provides city authorities with new means to improve the security of public spaces in smart cities, and so help protect citizens. It delivers an advanced, technology-based solution that helps operational personnel, based on data gathered from multiple sources, to work closely with each other and with state-of-the art tools to detect threats and make well-informed decisions about how to deal with them.

IMPETUS provides a solution that brings together:

- *Technology*: leverage the power of Internet of Things, Artificial Intelligence and Big Data to provide powerful tools that help operational personnel manage physical and cyber security in smart cities.
- *Ethics*: Balance potentially conflicting needs to collect, transform and share large amounts of data with the imperative of ensuring protection of data privacy and respect for other ethical concerns - all in the context of ensuring benefits to society.
- *Processes*: Define the steps that operational personnel must take, and the assessments they need to make, for effective decision making and coordination - fully aligned with their individual context and the powerful support offered by the technology.

Technological results are complemented by a set of *practitioner's guides* providing guidelines, documentation and training materials in the areas of operations, ethical/legal issues and cybersecurity.

IMPETUS places great emphasis on taking full and proper account of ethical and legal issues. This is reflected in the way project work is carried out, the nature of the project's results and the restrictions imposed on their use, and the inclusion of external advisors on these issues in project management.

The cities of Oslo (Norway) and Padova (Italy) have been selected as the site of practical trials of the IMPETUS solution during the project lifetime, but the longer-term goal is to achieve adoption much more widely.

The work is carried out by a consortium of 17 partners from 11 different EU Member States and Associated Countries. It brings together 5 research institutions, 7 specialist industrial and SME companies, 3 NGOs and 2 local government authorities (the trial sites). The consortium is complemented by the Community of Safe and Secure Cities (COSSEC) – a group established by the project to provide feedback on the IMPETUS solution as it is being developed and tested.

The project started in September 2020 with a planned duration of 30 months.

For more information

Project Coordinator: Joe Gorman, SINTEF: joe.gorman@sintef.no
Dissemination Manager: Snježana Knezić, TIEMS: snjezana.knezic@gmail.com

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Executive Summary

This deliverable presents practical advice about ethical, legal and data privacy issues that may arise when using advanced technological solutions to collect, analyse and manipulate data insecurity operations. It is aimed at practitioners with responsibility for security operations, and who could be future adopters of IMPETUS results.

While the official title is "Ethical Framework" the content will be presented externally as "Practitioner's Guidelines", to emphasise the intention of providing practical advice. The term "Ethical Framework" has a specific meaning in the field of ethics. The deliverable *does* contain the type of information that would be expected under that specific meaning but provides other advice in addition.

The deliverable contains:

- An introduction describing the overall approach to ethics within the project. This forms the "requirements" basis for other parts of the deliverable.
- An explanation of the concept of "Practitioner's Guidelines" and what these must cover.
- A set of stand-alone "parts" each offering information/advice on different issues, aimed at different intended readers.

The deliverable will be released in two versions: this initial one (D5.1), followed by the final one (D5.3) towards the end of the project. D5.3 will refine and extend the material presented here by: (a) adding new material; (b) refining material based on practical experience in using the technology during the project; and (c) improving presentational issues to improve communication.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Contents

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Table of Figures

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Tables

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Abbreviations

Abbreviation	Explanation
AAPD	Convention for the Protection of Individual with regard to Automatic Processing of Personal Data
ACM	Association for Computing Machinery
AI	Artificial Intelligence
Art.	Article
AAPD Protocol	2018 Protocol to the Convention for the Protection of Individual with regard to Automatic Processing of Personal Data
CCTV	Closed-circuit television video surveillance system
CEPEJ	European Commission for the Efficiency of Justice
CFREU	Charter of Fundamental Rights of the European Union
CJEU	Court of Justice of the European Union
COSSEC	Community of Safe and Secure Cities
DoA	Description of the Action
DMP	Data Management Plan
CUTLER	Coastal Urban Development through the Lenses of Resiliency
EAB	Ethics Advisory Board
EC	European Commission
ECHR	European Convention of Human Rights
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EGTAI	Ethics Guidelines for Trustworthy AI
EU	European Union
GDPR	General Data Protection Regulation
HECTOS	Harmonized Evaluation, Certification and Testing of Security products
HIC	Human-in-command
HITL	Human-in-the-loop

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Abbreviation	Explanation
HOTL	Human-on-the-loop
IEEE	Institute of Electrical and Electronics Engineers
IMPETUS	Intelligent Management of Processes, Ethics and Technology for Urban Safety
OECD	Organisation for Economic Co-operation and Development
Safe-DEED	Safe Data-Enabled Economic Development
SHERPA	Shaping the ethical dimensions of smart information systems
TFEU	Treaty on the Functioning of the European Union
UDHR	Universal Declaration of Human Rights
UK	United Kingdom
UN	United Nations Organization
WITDOM	Empowering privacy and security in Non-Trusted Environments
WP(s)	Working Package(s)
XAI	Explainable AI

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Definitions

Term	Explanation
AI Ethics	*AI Ethics is generally viewed as an example of applied ethics and focuses on the normative issues raised by the design, development, implementation and use of AI. (EGTAI, 2019:37)*
(de-) Anonymization	*Anonymization is a process by which personal data is irreversibly altered in such a way that a data subject can no longer be identified directly, either by the data controller alone or in collaboration with any other party. (Material 3, D5.1)* *De-anonymization is the reverse process in which anonymous data is cross-referenced with other data sources (publicly available information or auxiliary data) to re-identify the anonymous data source. (Material 3, D5.1)*
Applied ethics	*Applied ethics deals with issues concerning what we are obligated (or permitted) to do in a specific (often historically new) situation or a particular domain of (often historically unprecedented) possibilities for action. Applied ethics deals with real-life situations, where decisions have to be made under time-pressure, and often limited rationality. (EGTAI, 2019:37)*
Artificial intelligence (AI) systems	*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. (EGTAI, 2019:36)*
Data Manipulation	*The term data manipulation refers to all means of data processing. (GDPR, Art. 4(2))*
Data Processing	*Data processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction. (GDPR, Art. 4(2))*
Personal Data	*Personal data means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (GDPR, Art. 4(1))*

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Term	Explanation
Pseudonymisation	*Pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. (GDPR, 2016:4(5))*
Right to Privacy	*Right to private and family life. 1. Everyone has the right to respect for his private and family life, his home and his correspondence. 2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others. (ECHR, 1953)*
Sensitive Personal Data	*Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.*

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. About this Deliverable

Notes to the reader

1. "Framework" vs "Practitioner's Guide"

The official title of this deliverable in the project's DoA uses the term "Ethical Framework". The project also delivers two other "frameworks" – one on Security Operations and one on Cyber Security.

The goal of the three "frameworks" is to provide people outside the project with practical advice, based on work done in the project on the topics of Ethics, Security Operations and Cyber security. The consortium decided that "Practitioner's Guide" would be a better description of the role of these three project results, when presenting them externally. However, the official title of the deliverables remains unchanged in the DoA.

In the case of this deliverable, a possible source of confusion arises because the term "Ethical Framework" has a very specific meaning in the field of ethics. The deliverable *does* describe an "Ethical Framework" in this academic sense, but as a "Practitioner's Guide" also provides supplementary material.

2. Work in progress

This deliverable D5.1 is the *Initial* Ethical Framework. It will evolve into D5.3, to be delivered at the end of the project.

1.1 Intended Readership

The primary intended readership consists of people with any kind of responsibility for security in public spaces, and who are:

- Interested in using results of the IMPETUS project,
- Have specific interests in ethical, legal and data privacy aspects of using the IMPETUS platform in security-related operations.

The primary readership can be split into:

- People outside of the IMPETUS consortium: adopters of project results after completion of the project,
- Members of the IMPETUS consortium involved in the design, testing and evaluation of IMPETUS solutions.

The deliverable (or parts of it) may also be of interest to a wider group: anyone with an interest in ethical, legal and data privacy issues in the context of using advanced technological solutions (including but not limited to use of algorithms, monitoring by electronic means, "Big Data" and "Internet of Things") to collect, analyse and manipulate data in security operations.

1.2 Why would I want to Make Use of this Deliverable?

If you are someone *outside* the IMPETUS consortium wishing to use some or all of the IMEPTUS tools and platform:

- To learn about the general ethical, legal and data privacy issues that you ought to be aware of if you want to deploy the IMPETUS Platform in security operations,
- To learn about any specific issues that may apply in relation to individual tools of the IMPETUS Platform,
- To access practical guidance and support materials that will help you when ensuring compliance with relevant principles and regulation.

If you are someone inside the IMPETUS consortium using the tools as part of project work:

- As for people outside the consortium,
- In addition: To help you provide feedback to tool and platform developers in the project about aspects of

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

their solutions that might need to be designed differently for you to be able to comply with relevant principles and regulation.

If you are outside the consortium,

- To help you design your solution in a way that helps users address/overcome potential ethical, legal or data privacy barriers, and so increase the market potential of your product. ”Users” in this context does not apply just to consortium members: it applies to *any* organisation who might wish to use your product.

1.3 Structure

This document provides an overall introduction consisting of:

1. **CHAPTER 2:** Information on the overall approach to ethics followed in the project. This provides project-internal details that provide insight into the context of the work and the wider scope. Some of the results of this wider work are reported on through other project activities, rather than in this deliverable.
2. **CHAPTER 3:** Information on the scope and rationale for the Practitioner’s Guidelines that are the core of this deliverable.
3. **CHAPTER 4:** Plans for future work to further develop the Practitioner’s Guidelines.

The Guidelines themselves are provided as separate stand-alone documents, each with a specific focus. The idea is that each of these can evolve separately and be refined for the needs of its specific audience. The table below lists the set of materials, also indicating the current status of development of each. The set of documents will grow: see the list provided at the end of the “Next steps” described in Chapter 3.

Part	**Type of Material / Title**	**Description**	**Current Status**
Part 1	Survey Report ”*COSSEC Survey*”	Analysis based on a set of questions to be answered by COSSEC Members	Initial Draft
Part 2	Brochure ”*Public Trust in Digital Platforms and Personal Data Safekeeping*”	Citizens' Guide to the Estonian example of public trust in digital structures, access to data, transparency and personal data safekeeping	Initial Draft
Part 3	Protocol ”*Protocol on Anonymization and Deanonymization*”	Technical protocol on anonymization and deanonymization of data collected in social media	Completed
Part 4	Brochure ”*Human & AI Teaming Brochure*”	Citizens' Guide on Human and AI interoperability, biases, explainability and alignment issues	Completed
Part 5	Report ”*Analysis of Platform's Tools*”	A set of questions for tool developers, analysis of employed technical standards and protocols, of potential use for other WPs	Initial Draft
Part 6	Brochure ”*Alerts and Information Generated by Tools*”	Citizens' Guide on the Interaction between machine learning algorithms and human operators, relevance and consequences of false positive and false negative alerts, wider societal implications	Initial Draft

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Part	Type of Material / Title	Description	Current Status
Part 7	Report "*Human Computer Interaction tool*"	Citizens' Guide on potential biases in human (mental) workload assessments, explainability by assessment, and alignment problems related to HCI tool	Completed
Part 8	Guide "*Thresholds Algorithm Potential Issues*"	Definition of thresholds, tweaking of thresholds, deployers and thresholds' manipulation, relevance of tweaking parameters, list of possible errors, list of possible (relevant) factors and parameters of relevance for algorithm's learning and decision-making, deployers and parameters definition	Completed
Part 9	Brochure "*A guide for citizens: on the ethical and privacy aspects of IMPETUS* ""	Citizens' Guide on the IMPETUS Platform, advantages and better capacity for more efficient workflow, better capacity to avoid mistakes and bias, dangers and fears, possible abuse, safeguards, and similar	Initial Draft
Part 10	Report "Overview of *General Ethical Issues*"	Citizens' Guide on ethical issues arising out of collection and manipulation of data in security operations	Mature Draft
Part 11	Report "*Legal Analysis*"	Report on relevant body of legislation	Initial Draft
Part 12	Report "*Relevant Ethical Guidelines and Principles*"	Citizens' Guide on how the EU Guidelines (and other similar guidelines) are relevant for IMPETUS Platform	Mature Draft

Table 1: Set of Guidelines and Educational Materials

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

2. IMPETUS approach to Ethical Framework

Work on ethics is one of three main pillars of work in the IMPETUS project. The scope of that work goes beyond the guidelines presented in this deliverable. The purpose of this chapter is to describe this more general scope. While some of the project-internal details here may not be directly relevant to practitioners outside the IMPETUS project, they provide an explanation of the overall context, motivation and approach used.

2.1 General Requirements

The smart systems and smart technologies increase the risk of unethical use of personal data. Having that in mind, the IMPETUS Project aims to enhance the cities' resilience regarding security of public spaces by addressing, among other pillars, the ethical aspects of urban security. The ethics pillar is focused on the smart cities' technological capabilities and capacities to collect, transform, and share large amounts of data with the use of advanced machine learning algorithms (often, incorrectly, referred to as the artificial intelligence (AI)). The capacity to collect and handle "Big Data" presents a moral dilemma when faced with the need to protect citizens' privacy and personal data on one side of the spectrum (privacy *in generalis*) and protect citizens' livelihood and property from the other (public security). Therefore, a balance of the noted potentially conflicting interests should be established to simultaneously offer adequate and efficient mechanisms of preserving public security and personal privacy. Ethical issues arising out of the noted conflict of interests require constant study, vigilance, and practical considerations for all the involved stakeholders.

The IMPETUS Ethical Framework, as the third pillar, aims to provide a comprehensive outlook of ethical (and legal) issues arising out of the use of smart cities' technology and personal and non-personal data manipulation. Ethical-grounded research activities and tasks in the IMPETUS Project are divided into several tasks and many deliverables, allocated among various work packages (WPs).

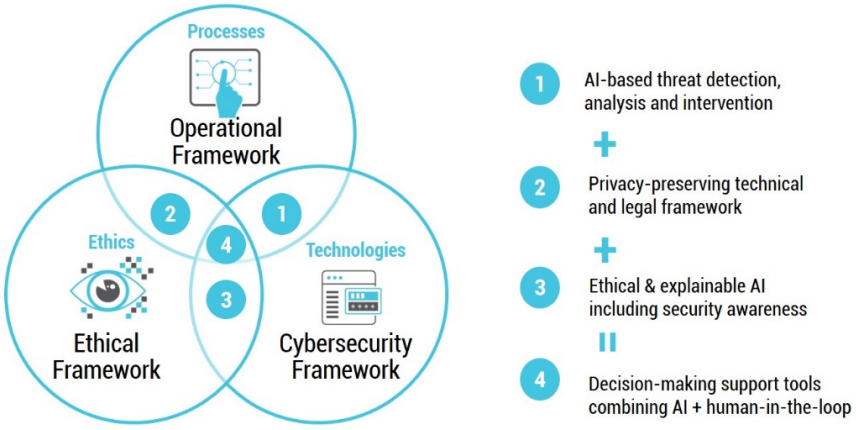


Figure 1: Intersection of Processes, Ethics and Technology (DoA, Part B, 2020:15)

The development of the IMPETUS Ethical Framework will, additionally, be strengthened through the analysis and synthesis of data collected during interviews, workshops and other collaborative activities. The process will involve a wide range of stakeholders, including representatives of local governments, security or emergency management organizations, technological providers and representatives of citizens (participants will be invited through partner cities and through COSSEC). Although aimed primarily at end-users, the development process includes significant resources for the participation of technology providers, with the assumption that relevant partners have a fine understanding of information needs for their solutions and potential risks. They therefore need to participate in the identification of solutions to provide the intended capabilities while ensuring compliance with ethics.

In addition, the introduction of AI raises the risk of additional bias. For example, training an AI on an incorrect data set can increase the potential for stigma and discrimination resulting from being associated with locations or socio-economic categories. Whereas the public, in general, supports the main efforts aimed at enhancing the public security, the data-collection technology utilized in security operations may raise concerns as to the privacy issues. At the same time, the concept of a "*safe city*" refers to a smart city's capacity to utilize the data and technology to keep residents safe. As analysed in D1.2, the Economist's "Safe Cities Index" points to several factors relevant for the pending ethical analysis. Digital security assesses the ability of urban citizens to freely

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

use the internet and other digital channels without fear of privacy violations or identity theft. Personal security takes into consideration policies and decisions such as the level of police engagement and the use of data-driven crime prevention. The technology plays an obvious role in digital security, but new developments, for example, in data mining and AI, are opening some intriguing possibilities in other security pillars. The application of AI to data management aims to improve and enhance digital, health, infrastructure, and personal security, and contribute to creating a generally secure environment. The process will follow the Ethics Guidelines for Trustworthy AI published by the High-Level Expert Group on AI (EGTAI) in order to develop, deploy and operate a trustworthy (i.e., lawful, ethical and robust) AI-based solution.

The IMPETUS Ethical Framework development represents a collaborative process allowing for the construction across perspectives of an understanding of security goals, technological capabilities, and potential impact on citizens, as well as the identification of best practices. The purpose of the IMPETUS Ethical Framework developed by the IMPETUS Project is to serve as a tool for decision-makers, actors of the city investing in and using security and smart city solutions. To be useful, such a decision tool needs to find a balance between simplicity and effectiveness. Like traditional ethical frameworks, it will essentially try to facilitate answering questions about the legal compliance, fairness and perception of the solutions proposed, although in the more specific and novel context of technology-based solutions relying especially on AI. The framework will be location-agnostic in nature and adaptable to the legal and societal context of partner cities because those elements of context vary and do not allow for more specific guidelines. The framework will identify relevant aspects of data processing and utilization during the platform operation and determine the requirements that must be met to have the platform functioning in full compliance with all relevant legal and ethical standards and provisions. The framework will detail all necessary requirements as to the privacy, security, and surveillance issues, in particular regarding the non-voluntary visual and audio capture of personal property, access to personal data, unknown (and especially unintended) surveillance, storage and security of data so collected, access to such data, and similar. The framework will be focused on finding the locally acceptable balance between security needs and information needs for the operation of security and smart city solutions. It will also address the potential repurposing of smart city data (i.e., related to sustainability or transport efficiency goals) for security purposes.

At the same time as the solutions developed by IMPETUS are expected to produce significant societal benefits, elements of the project's work will seek to minimize potential negative impacts. Negative societal impact is understood as the weakening of societal values, in particularly those embedded in the treaties of the European Union (EU) and reflected in European fundamental values such as freedom of association, freedom of expression, protection of personal dignity, privacy and data protection, etc. The IMPETUS approach includes the measures and considerations that place protection of these values throughout its work. In addition, the development of IMPETUS Ethical Framework associated with the technological platform aims precisely at understanding and ensuring the protection of the societal laws, values and goals of its users. After the project is concluded, potential adopters of the platform will therefore be able to refer to this framework and revisit it based on their own contexts.

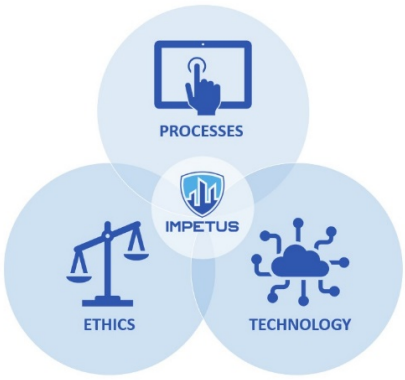


Figure 2: IMPETUS Intersection Diagram, IMPETUS Project Promotion

IMPETUS will develop the IMPETUS Ethical Framework, combining the complementary perspectives of end-users (i.e., partner cities, COSSEC members) and technology providers through data collection, workshops, and other collaborative activities. It should be noted that the IMPETUS Ethical Framework as a third IMPETUS pillar consists of numerous individual and interconnected tasks conducted across various deliverables and work packages, comprising of a wide representation of various ethical (and legal) issues and considerations regarding the use of smart technologies in data collection and manipulation. Table below, points to, at this stage of IMPETUS Project, relevant sections that are representing the IMPETUS Ethical

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Framework.

General Technical, Ethical and Legal Requirements that will form the IMPETUS Ethical Framework	Work Package (WP)
Human agency and oversight: in the project, AI systems will empower human beings, allowing them to make informed decisions and fostering their fundamental rights, and at the same time, proper oversight mechanisms will be ensured using a human-in-the-loop approach. IMPETUS Platform Tools should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that IMPETUS Platform Tools should act as enablers of a democratic, flourishing, and equitable society by supporting the user's agency, fostering fundamental rights and allowing for human oversight.	WP 4
Technical robustness and safety: in the project, AI systems will be resilient and secure, with a fallback plan (degraded mode) in case the systems are not available. A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the principle of prevention of harm. Technical robustness requires that IMPETUS Platform Tools be developed with a preventative approach to risks. In addition, the IMPETUS Platform Tools should be developed in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured.	WP 6
Privacy and data governance: the project will ensure full respect for privacy and data protection, adequate data governance mechanisms, considering the quality and integrity of the data, and ensuring proper access control. Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by IMPETUS Platform Tools. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance considering the domain in which the IMPETUS Platform Tools will be deployed, its access protocols and the capability to process data in a manner that protects privacy.	WP 5 (security operations) WP 11 (other)
Transparency: the project will strive to make the data, system, and AI business models transparent, so AI systems and their decisions should be explained in a manner adapted to the stakeholders concerned. This requirement is closely linked with the principle of explicability and encompasses transparency of elements relevant to IMPETUS Platform Tools: the data, the system, and the business models.	WP 5 (security operations) WP 3
Diversity, non-discrimination, and fairness: the project will avoid unfair bias (such as exacerbation of prejudice and discrimination) by paying particular attention to the data set used to train the AI. To achieve Trustworthy AI, inclusion and diversity must be enabled throughout the entire IMPETUS Platform Tools' life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with the principle of fairness.	WP 3 WP 4

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

General Technical, Ethical and Legal Requirements that will form the IMPETUS Ethical Framework	Work Package (WP)
Societal and environmental well-being: the societal impact of the AI deployed in the project will be carefully considered, such as carbon footprint of the computing resources necessary to run the system efficiently. In line with the principles of fairness and prevention of harm, the broader society, other sentient beings and the environment should be also considered as stakeholders throughout the IMPETUS Platform Tools' life cycle. Sustainability and ecological responsibility of IMPETUS Platform Tools should be encouraged, and research should be fostered into IMPETUS Platform Tools solutions addressing areas of global concern, such as for instance the Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations.	WP 9
Accountability: the project will put in place mechanisms to ensure responsibility and accountability for AI systems and their outcomes. All systems will be fully auditable (assessment of algorithms, data, and design processes). The requirement of accountability complements the above criteria and is closely linked to the principle of fairness. It necessitates that those mechanisms must be put in place to ensure responsibility and accountability for IMPETUS Platform Tools and their outcomes, both before and after their development, deployment, and use.	WP 3
Validation of Ethical Framework Criteria	WP 7

Table 2 2: General Technical, Ethical and Legal Requirements

2.2 Data Collection and Manipulation during Project

Work conducted under WP 11 relates to the part of IMPETUS Ethical Framework dealing with relevant ethical and legal issues regarding the Project activities that involve the collection and manipulation of data. As such, work conducted in WP11 is internal in nature and reserved for Project purposes. The summaries of main general legal and ethical findings may be reproduced for wider audience at a later stage. As such activities are conducted by Project partners as simulations, they do not fall under the security operations activities, but under general data manipulation activities. Therefore, WP 11 is considering all relevant ethical and legal aspects of data manipulation as per GDPR. To that end, WP 11 has already produced a number of deliverables in time for the IMPETUS Projects' pilots and trials.

2.3 Ethical Framework Validation Process

WP 7 is in charge of tools validation from, among other items, ethical perspective. IMPETUS will develop an ethical framework, combining the complementary perspectives of end-users (especially partner cities, COSSEC members) and technology providers through data collection, workshops and other collaborative activities. The process will follow the EGTAI's Trustworthy AI Assessment List in order to develop, deploy and operate a trustworthy (i.e., lawful, ethical and robust) AI-based solution. The Trustworthy AI Assessment list is reproduced in table below.

Trustworthy AI Assessment List
Human Agency and Oversight
Fundamental rights

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q1.- Did you carry out a fundamental rights impact assessment where there could be a negative impact on fundamental rights? Q2. - Did you identify and document potential trade-offs made between the different principles and rights? Q3. - Does the IMPETUS platform interact with decisions by human (end) users (e.g., recommended actions or decisions to take, presenting of options)? Q3.1 -- Could the IMPETUS platform affect human autonomy by interfering with the (end) user's decision-making process in an unintended way? Q3.2 -- Did you consider whether the IMPETUS platform should communicate to (end) users that a decision, content, advice or outcome is the result of an algorithmic decision? Q3.3 -- In case of a chat bot or other conversational system, are the human end users made aware that they are interacting with a non-human agent?
Human agency
Q4 - Is the IMPETUS platform implemented in work and labor process? If so, did you consider the task allocation between the IMPETUS platform and humans for meaningful interactions and appropriate human oversight and control? Q4.1 -- Does the IMPETUS platform enhance or augment human capabilities? Q4.2 -- Did you take safeguards to prevent overconfidence in or overreliance on the IMPETUS platform for work processes?
Human oversight
Q5 - Did you consider the appropriate level of human control for the particular IMPETUS platform and use case? Q5.1 -- Can you describe the level of human control or involvement? Q5.2 -- Who is the "human in control" and what are the moments or tools for human intervention? Q5.3 -- Did you put in place mechanisms and measures to ensure human control or oversight? Q5.4 -- Did you take any measures to enable audit and to remedy issues related to governing AI autonomy? Q6 - Is there is a self-learning or autonomous IMPETUS platform/tool or use case? If so, did you put in place more specific mechanisms of control and oversight? Q6.1 -- Which detection and response mechanisms did you establish to assess whether something could go wrong? Q6.2 -- Did you ensure a stop button or procedure to safely abort an operation where needed? Does this procedure abort the process entirely, in part, or delegate control to a human?
Technical Robustness and Safety

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Resilience to attack and security
Q7 - Did you assess potential forms of attacks to which the IMPETUS platform could be vulnerable? Q7.1 -- Did you consider different types and natures of vulnerabilities, such as data pollution, physical infrastructure, cyber-attacks? Q7.2 -- Did you put measures or systems in place to ensure the integrity and resilience of the IMPETUS platform against potential attacks? Q8 - Did you verify how your system behaves in unexpected situations and environments? Q9 - Did you consider to what degree your system could be dual-use? If so, did you take suitable preventative measures against this case (including for instance not publishing the research or deploying the system)?
Fallback plan and general safety

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q10 - Did you ensure that your system has a sufficient fallback plan if it encounters adversarial attacks or other unexpected situations (for example technical switching procedures or asking for a human operator before proceeding)? Q11 - Did you consider the level of risk raised by the IMPETUS platform in this specific use case? Q11.1 -- Did you put any process in place to measure and assess risks and safety? Q11.2 -- Did you provide the necessary information in case of a risk for human physical integrity? Q11.3 -- Did you consider an insurance policy to deal with potential damage from the IMPETUS platform? Q11.4 -- Did you identify potential safety risks of (other) foreseeable uses of the technology, including accidental or malicious misuse? Is there a plan to mitigate or manage these risks? Q12 - Did you assess whether there is a probable chance that the IMPETUS platform may cause damage or harm to users or third parties? Did you assess the likelihood, potential damage, impacted audience and severity? Q12.1-- Did you consider the liability and consumer protection rules, and take them into account? Q12.2 -- Did you consider the potential impact or safety risk to the environment or to animals? Q12.3 -- Did your risk analysis include whether security or network problems such as cybersecurity hazards could pose safety risks or damage due to unintentional behavior of the AI system? Q13 - Did you estimate the likely impact of a failure of your IMPETUS platform when it provides wrong results, becomes unavailable, or provides societally unacceptable results (for example discrimination)? Q13.1 -- Did you define thresholds and did you put governance procedures in place to trigger alternative/fallback plans? Q13.2 -- Did you define and test fallback plans?
Accuracy

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q14 - Did you assess what level and definition of accuracy would be required in the context of the IMPETUS platform and use case? Q14.1 -- Did you assess how accuracy is measured and assured? Q14.2 -- Did you put in place measures to ensure that the data used is comprehensive and up to date? Q14.3 -- Did you put in place measures in place to assess whether there is a need for additional data, for example to improve accuracy or to eliminate bias? Q15 - Did you verify what harm would be caused if the IMPETUS platform makes inaccurate predictions? Q16 - Did you put in place ways to measure whether your system is making an unacceptable amount of inaccurate predictions? Q17 - Did you put in place a series of steps to increase the system's accuracy?
Reliability and reproducibility
Q18 - Did you put in place a strategy to monitor and test if the IMPETUS platform is meeting the goals, purposes and intended applications? Q18.1 -- Did you test whether specific contexts or particular conditions need to be taken into account to ensure reproducibility? Q18.2 -- Did you put in place verification methods to measure and ensure different aspects of the system's reliability and reproducibility? Q18.3 -- Did you put in place processes to describe when your IMPETUS platform fails in certain types of settings? Q18.4 -- Did you clearly document and operationalise these processes for the testing and verification of the reliability of the IMPETUS platform? Q18.5 -- Did you establish mechanisms of communication to assure (end-)users of the system’s reliability?
Privacy and Data Governance
Respect for privacy and data protection

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q19 - Considering this use case, did you establish a mechanism allowing others to flag issues related to privacy or data protection in the IMPETUS platform's processes of data collection (for training and operation) and data processing? Q20 - Did you assess the type and scope of data in your data sets (for example whether they contain personal data)? Q21 - Did you consider ways to develop the IMPETUS platform or train the model without or with minimal use of potentially sensitive or personal data? Q22 - Did you build in mechanisms for notice and control over personal data in this use case (such as valid consent and possibility to revoke, when applicable)? Q23 - Did you take measures to enhance privacy, such as via encryption, anonymisation and aggregation? Q24 - Where a Data Privacy Officer (DPO) exists, did you involve this person at an early stage in the process?
Quality and integrity of data
Q25 - Did you align your IMPETUS platform with relevant standards (for example ISO, IEEE) or widely adopted protocols for daily data management and governance? Q26 - Did you establish oversight mechanisms for data collection, storage, processing and use? Q27 - Did you assess the extent to which you are in control of the quality of the external data sources used? Q28 - Did you put in place processes to ensure the quality and integrity of your data? Did you consider other processes? How are you verifying that your data sets have not been compromised or hacked?
Access to data
Q29 - What protocols, processes and procedures did you follow to manage and ensure proper data governance? Q29.1 -- Did you assess who can access users' data, and under what circumstances? Q29.2 -- Did you ensure that these persons are qualified and required to access the data, and that they have the necessary competences to understand the details of data protection policy? Q29.3 -- Did you ensure an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed?
Transparency
Traceability

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q30 - Did you establish measures that can ensure traceability? This could entail documenting the following methods: Q31 - Methods used for designing and developing the algorithmic system: Q31.1 -- Rule-based AI systems: the method of programming or how the model was built; Q31.2 -- Learning-based AI systems; the method of training the algorithm, including which input data was gathered and selected, and how this occurred. Q32- Methods used to test and validate the algorithmic system: Q32.1 -- Rule-based AI systems; the scenarios or cases used in order to test and validate; Q32.2 -- Learning-based model: information about the data used to test and validate. Q33 - Outcomes of the algorithmic system: Q33.1 -- The outcomes of or decisions taken by the algorithm, as well as potential other decisions that would result from different cases (for example, for other subgroups of users).
Explainability
Q34 - Did you assess: Q34.1 -- to what extent the decisions and hence the outcome made by the IMPETUS platform can be understood? Q34.2 -- to what degree the system's decision influences the organization's decision-making processes? Q34.3 -- why this particular system was deployed in this specific area? Q34.4 -- what the system's business model is (for example, how does it create value for the organization)? Q35 - Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand? Q36 - Did you design the IMPETUS platform with interpretability in mind from the start? Q36.1 -- Did you research and try to use the simplest and most interpretable model possible for the application in question? Q36.2 -- Did you assess whether you can analyse your training and testing data? Can you change and update this over time? Q36.3 -- Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?
Communication

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q37 - Did you communicate to (end-)users – through a disclaimer or any other means – that they are interacting with an AI system and not with another human? Did you label your IMPETUS platform as such? Q38 - Did you establish mechanisms to inform (end-)users on the reasons and criteria behind the IMPETUS platform's outcomes? Q38.1 -- Did you communicate this clearly and intelligibly to the intended audience? Q38.2 -- Did you establish processes that consider users' feedback and use this to adapt the system? Q38.3 -- Did you communicate around potential or perceived risks, such as bias? Q38.4 -- Considering this use case, did you consider communication and transparency towards other audiences, third parties or the general public? Q39 - Did you clarify the purpose of the IMPETUS platform and who or what may benefit from the product/service? Q39.1 -- Did you specify usage scenarios for the product and clearly communicate these to ensure that it is understandable and appropriate for the intended audience? Q39.2 -- Considering this use case, did you think about human psychology and potential limitations, such as risk of confusion, confirmation bias or cognitive fatigue? Q40 - Did you clearly communicate characteristics, limitations and potential shortcomings of the IMPETUS platform? Q40.1 -- In case of the system's development: to whoever is deploying it into a product or service? Q40.2 -- In case of the system's deployment: to the (end-)user or consumer?
Diversity, Non-Discrimination and Fairness
Unfair bias avoidance

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q41 - Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design? Q41.1 -- Did you assess and acknowledge the possible limitations stemming from the composition of the used data sets? Q41.2 -- Did you consider diversity and representativeness of users in the data? Did you test for specific populations or problematic use cases? Q41.3 -- Did you research and use available technical tools to improve your understanding of the data, model and performance? Q41.4 -- Did you put in place processes to test and monitor for potential biases during the development, deployment and use phase of the system? Q42 – Considering this use case, did you ensure a mechanism that allows others to flag issues related to bias, discrimination or poor performance of the IMPETUS platform? Q42.1 -- Did you establish clear steps and ways of communicating on how and to whom such issues can be raised? Q42.2 -- Did you consider others, potentially indirectly affected by the IMPETUS platform, in addition to the (end)-users? Q43 - Did you assess whether there is any possible decision variability that can occur under the same conditions? Q43.1 -- If so, did you consider what the possible causes of this could be? Q43.2 -- In case of variability, did you establish a measurement or assessment mechanism of the potential impact of such variability on fundamental rights? Q44 - Did you ensure an adequate working definition of ”fairness” that you apply in designing this IMPETUS platform? Q44.1 -- Is your definition commonly used? Did you consider other definitions before choosing this one? Q44.2 -- Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness? Q44.3 -- Did you establish mechanisms to ensure fairness in your IMPETUS platform? Did you consider other potential mechanisms?
Accessibility and universal design

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q45 - Did you ensure that the IMPETUS platform accommodates a wide range of individual preferences and abilities? Q45.1 -- Did you assess whether the IMPETUS platform is usable by those with special needs or disabilities or those at risk of exclusion? How was this designed into the system and how is it verified? Q45.2 -- Did you ensure that information about the IMPETUS platform is accessible also to users of assistive technologies? Q45.3 -- Did you involve or consult this community during the development phase of the IMPETUS platform? Q46 - Did you take the impact of your IMPETUS platform on the potential user audience into account? Q46.1 -- Did you assess whether the team involved in building the IMPETUS platform is representative of your target user audience? Is it representative of the wider population, considering also of other groups who might tangentially be impacted? Q46.2 -- Did you assess whether there could be persons or groups who might be disproportionately affected by negative implications? Q46.3 -- Did you get feedback from other teams or groups that represent different backgrounds and experiences?
Stakeholder participation
Q47 - Did you consider a mechanism to include the participation of different stakeholders in the IMPETUS platform's development and use? Q48 - Did you pave the way for the introduction of the IMPETUS platform in your organisation by informing and involving impacted workers and their representatives in advance?
Societal and Environment Well-Being
Sustainable and environmentally friendly AI
Q49 - Did you establish mechanisms to measure the environmental impact of the IMPETUS platform's development, deployment and use (for example the type of energy used by the data centres)? Q50 - Did you ensure measures to reduce the environmental impact of your IMPETUS platform's life cycle?
Social impact

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q51 - In case the IMPETUS platform interacts directly with humans: Q51.1 -- Did you assess whether the IMPETUS platform encourages humans to develop attachment and empathy towards the system? Q51.2 -- Did you ensure that the IMPETUS platform clearly signals that its social interaction is simulated and that it has no capacities of ”understanding” and ”feeling”? Q52 - Did you ensure that the social impacts of the IMPETUS platform are well understood? For example, did you assess whether there is a risk of job loss or de-skilling of the workforce? What steps have been taken to counteract such risks?
Society and democracy
Q53 - Did you assess the broader societal impact of the IMPETUS platform’s use beyond the individual (end-) user, such as potentially indirectly affected stakeholders?
Accountability
Auditability
Q54 - Did you establish mechanisms that facilitate the system’s auditability, such as ensuring traceability and logging of the IMPETUS platform’s processes and outcomes? Q55 - Did you ensure, in applications affecting fundamental rights (including safety-critical applications) that the IMPETUS platform can be audited independently?
Minimising and reporting negative Impact
Q56 - Did you carry out a risk or impact assessment of the IMPETUS platform, which takes into account different stakeholders that are (in)directly affected? Q57 - Did you provide training and education to help developing accountability practices? Q57.1 -- Which workers or branches of the team are involved? Does it go beyond the development phase? Q57.2 -- Do these trainings also teach the potential legal framework applicable to the IMPETUS platform? Q57.3 -- Did you consider establishing an ‘ethical IMPETUS platform review board’ or a similar mechanism to discuss overall accountability and ethics practices, including potentially unclear grey areas? Q58 - Did you foresee any kind of external guidance or put in place auditing processes to oversee ethics and accountability, in addition to internal initiatives? Q59 - Did you establish processes for third parties (e.g., suppliers, consumers, distributors/vendors) or workers to report potential vulnerabilities, risks or biases in the IMPETUS platform?
Documenting trade-offs

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Trustworthy AI Assessment List
Q60 - Did you establish a mechanism to identify relevant interests and values implicated by the IMPETUS platform and potential trade-offs between them? Q61 - How do you decide on such trade-offs? Did you ensure that the trade-off decision was documented?
Ability to redress
Q62 - Did you establish an adequate set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact? Q63 - Did you put mechanisms in place both to provide information to (end-)users/third parties about opportunities for redress?

Table 3 3: Trustworthy AI Assessment List

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

3. Practitioner's Guidelines on AI Ethics in Security Operations

The goal of this deliverable is to provide practical guidelines and practitioners' support materials covering ethical issues relevant for deployment of the IMPETUS Platform in live security operations. The advice is also applicable to other technologies providing similar functionalities and facing similar challenges. This complements other work in the project offering advice and support on privacy-preserving mechanisms.

In general, IMPETUS follows EGTAI in order to develop, deploy and operate a trustworthy (i.e., lawful, ethical and robust) AI-based solution: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) societal and environmental well-being and (7) accountability. This work will be guided by societal resilience concepts. It will emphasize the potentially different impact smart city solutions might have on different groups of the local populations. Differences considered include (but are not limited to) gender, ethnicity, socio-economic conditions, digital access, and literacy.

The Practitioner's Guidelines on AI Ethics in Security Operations are of two main types:

- Overall guidelines - processes to follow, dos and don'ts, role definitions, and similar, and
- Tool documentation - when to use which IMPETUS tools, and how to use them effectively.

The deliverable addresses five interconnected elements: security operations, use of algorithms, data collection and manipulation, personal data and right of privacy, practitioner's guidelines – as shown in the table below.

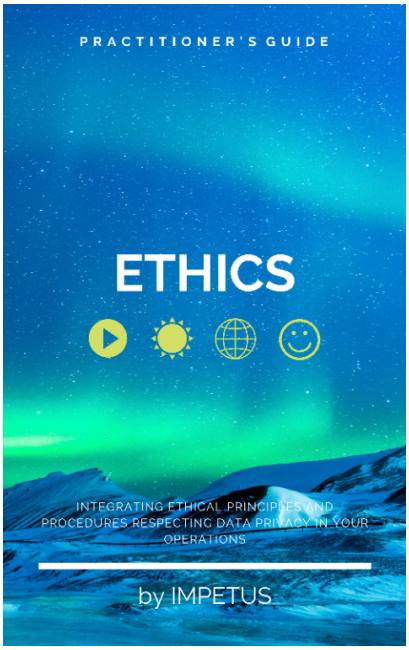


Figure 3: Practitioner's Guide on Ethics, IMPETUS Project Promotion[1]

Interconnected Elements	**Description**
Security operations	The ethical analysis is solely considering data collected and manipulated during the security operations (i.e., run or controlled by police departments, intelligence agencies, and similar)
Use of Algorithms	The ethical analysis focuses on such data collection and manipulation aided by or independently run by algorithms
Data Collection and Manipulation	As per relevant technology (both hardware and software that are made available for IMPETUS Project), focus in placed on the IoT system's Integration Layer and Intelligence Layer (as per D1.2 Requirements for public Safety Solutions)
Personal Data and Right of Privacy	The ethical analysis exclusively focuses on personal data

[1] The illustration shows one possible way that that the Guidelines might be packaged: as a textbook with an attractive cover. However, no decision has so far been reached on the best way to package and present the advice.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Interconnected Elements	Description
Practitioner's Guidelines	The ethical analysis is presented in the form of guidelines (short, summary overview of main issues, recommendations for practitioners), and not as a framework study

Table 44: Interconnected Elements

The Practitioner's Guidelines on AI Ethics in Security Operations focus on the following areas, also illustrated in the figure below:

- Security-related data manipulation considerations
- AI-related data manipulation considerations
- Focused on the technology utilized during IMPETUS Project
- Focused on personal data
- Short and practical learning materials and guidelines

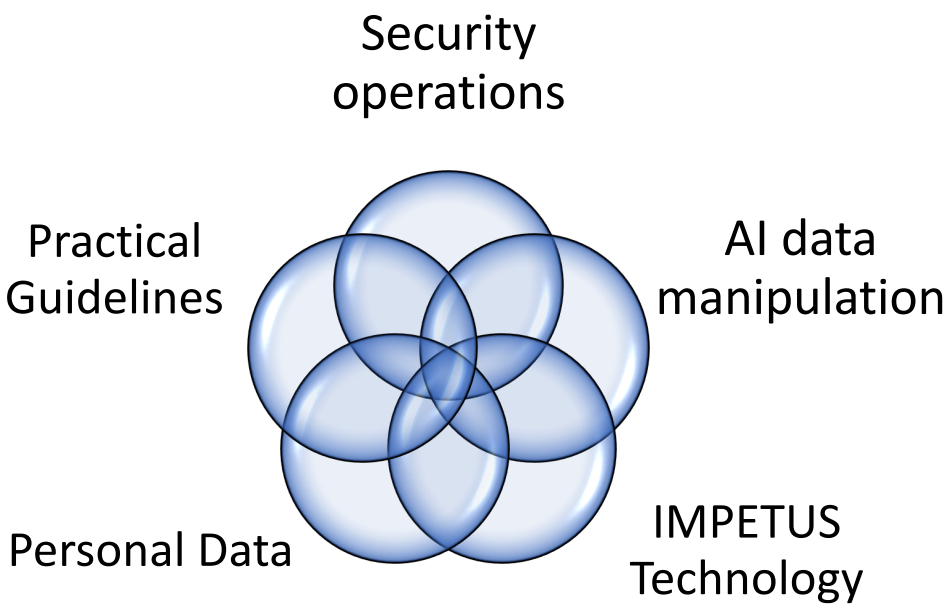


Figure 4: Focus areas

The Practitioner's Guidelines are heavily influenced by experiences preparing for and carrying out practical work in the by the pilot cities, and on the methods followed in carrying out that work. They are also heavily influenced by the tools developed and used in the project.

The materials presented here are practical in nature, short and informative. It is envisaged that there will be 27 individuals materials: 13 Reports, 8 Brochures, 5 Guides and 1 Protocol. But the structure may evolve: some materials may merge, some may change, and some new ones may get introduced.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

4. Next Steps

The work on individual materials will continue in deliverable D5.3. Given the nature of IMPETUS Project and the IMPETUS Platform that is being developed, practical guidelines are predominantly connected to the Platform itself and the tools it will be utilizing. Thus, many of the guidelines can be considerably matured after experience is gained in using the tools.

The following materials will be updated:

- Survey Report "*COSSEC Survey*" – analysis based on a set of questions to be answered by COSSEC Members,
- Brochure "*Public Trust in Digital Platforms and Personal Data Safekeeping*" – citizens' Guide to the Estonian example of public trust in digital structures, access to data, transparency and personal data safekeeping,
- Report "*Analysis of Platform's Tools*" – a set of questions for tools' developers, analysis of employed technical standards and protocols, of potential use for other WPs,
- Brochure "*Alerts and Information Generated by Tools*" – citizens' Guide on the Interaction between machine learning algorithms and human operators, relevance and consequences of false positive and false negative alerts, wider societal implications,
- Brochure "*A guide for citizens: on the ethical and privacy aspects of IMPETUS*" – citizens' Guide on the IMPETUS Platform, advantages and better capacity for more efficient workflow, better capacity to avoid mistakes and bias, dangers and fears, possible abuse, safeguards, and similar,
- Report "*Overview of General Ethical Issues*" – citizens' Guide on ethical issues arising out of collection and manipulation of data in security operations,
- Report "*Legal Analysis*" – report on relevant body of legislation,
- Report "*Relevant Ethical Guidelines and Principles*" – citizens' Guide on how the EU Guidelines (and other similar guidelines) are relevant for IMPETUS Platform.

The following new materials will be created:

- Report "*Report on Public Surveys*" – report on ethical issues arising out of various types of public surveys conducted by ISP under D1.2,
- Report "*Practitioner's Guide to Ethics*" – report on potential ethical issues arising from tools' evaluation and analysis present in D5.1 / D5.3, as well as general ethical considerations,
- Video Guide "*Data Flow Guide*"" – simplified explanation of data collection, data analysis, data flow, data accessibility, data storage and data utilization,
- Guide "*Data Flow Guide*" – simplified explanation of data collection, data analysis, data flow, data accessibility, data storage and data utilization,
- Report "*CINEDIT Tool Evaluation and Analysis*" – evaluation of pilot project and use of the tool, lessons learned,
- Guide "*Social Media Detection Tool*" – citizens' Guide on the tool,
- Brochure "*Guide on Data Anonymization, Deanonymization and Pseudoanonymization*" – citizens' Guide on Data Anonymization, Deanonymization and Pseudoanonymization,
- Report "*INS Tool Evaluation and Analysis*" – evaluation of pilot project and use of the tool, lessons learned,
- Brochure "*Set of Challenges for Deployers - Preparation Phase*" – technical infrastructure – what technical requirements need to be secured, training of human operators, use of tools, technology utilization risk assessment (technology, data, legislation, ethics); Ethical agenda – collection of data, citizen information; Policy agenda – city council and mayor, relationship with a national agenda and national bodies,
- Report "*THA Tool Evaluation and Analysis*" – review of the three AI - ethics related issues from the brochure in the context of all the tools used in the Impetus platform and the test findings concerning their use in the two partner cities,
- Guide "*Risk Values' Thresholds Algorithm Manual*" – citizens' Guide on the Tool,

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

- Report "*UPAD Tool Evaluation and Analysis*" – evaluation of pilot project and use of the tool, lessons learned.

Work in the next period will also consider presentational issues in more detail: how can the materials be communicated in an attractive, easily accessible format that makes it easy for practitioners to find the advice and help they need. Section 1.3 on "Structure" will be extended and improved to provide a basic "readers guide" describing how the various parts complement each other, and indicating the targeted readership of each.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Members of the IMPETUS consortium

SINTEF	SINTEF, Strindvegen 4, Trondheim, Norway, https://www.sintef.no	Joe Gorman joe.gorman@sintef.no
Institut Mines-Télécom	Institut Mines Telecom, 19 place Marguerite Perey, 91120 Palaiseau, France, https://www.imt.fr	Joaquin Garcia-Alfaro joaquin.garcia_alfaro@telecom-sudparis.eu
UNÎMES	Université de Nimes, Rue du Docteur Georges Salan CS 13019 30021 Nîmes Cedex 1, France, https://www.unimes.fr	Axelle Cadiere axelle.cadiere@unimes.fr
cini	Consorzio Interuniversitario Nazionale per l'Informatica, Via Ariosto, 25, 00185 – Roma, Italy, https://www.consorzio-cini.it	Donato Malerba donato.malerba@uniba.it
Università degli Studi di Padova	University of Padova, Via 8 Febbraio, 2 - 35122 Padova, Italy, https://www.unipd.it	Giuseppe Maschio giuseppe.maschio@unipd.it
Entrepreneurship Development Centre for BIOTECHNOLOGY and MEDICINE	Biotehnoloogia ja Meditsiini Ettevõtluse Arendamise Sihtasutus, Tiigi 61b, 50410 Tartu, Estonia, https://biopark.ee	Sven Parkel sven@biopark.ee
SIMAVI Software Imagination & Vision	SIMAVI, Complex Victoria Park, Corp C4, Etaj 2, Șos. București – Ploiești, nr. 73 – 81, Sector 1, București, Romania, https://www.simavi.ro	Gabriel Nicola Gabriel.Nicola@simavi.ro Monica Florea Monica.Florea@simavi.ro
THALES	Thales Nederland BV, Zuidelijke Havenweg 40, 7554 RR Hengelo, Netherlands, https://www.thalesgroup.com/en/countries/europe/netherlands	Johan de Heer johan.deheer@nl.thalesgroup.com
CINEDIT INTELLIGENT VIDEO ANALYTICS	Cinedit VA GmbH, Poststrasse 21, 8634 Hombrechtikon, Switzerland, https://www.cinedit.com	Joachim Levy j@cinedit.com

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

INSIKT INTELLIGENCE	Insikt Intelligence, Calle Huelva 106, 9-4, 08020 Barcelona, Spain, https://www.insiktintelligence.com	Dana Tantu dana@insiktintelligence.com
SIXGILL	Sixgill, Derech Menachem Begin 132 Azrieli Tower, Triangle Building, 42nd Floor, Tel Aviv, 6701101, Israel, https://www.cybersixgill.com	Benjamin Preminger benjamin@cybersixgill.com Ron Shamir ron@cybersixgill.com
XM CYBER	XM Cyber, Galgalei ha-Plada St 11, Herzliya, Israel https://www.xmcyber.com	Lior Barak lior.barak@xmcyber.com Menachem Shafran menachem.shafran@xmcyber.com
Comune di Padova	City of Padova, via del Municipio, 1 - 35122 Padova Italy, https://www.padovanet.it	Enrico Fiorentin fiorentine@comune.padova.it Stefano Baraldi Baraldis@comune.padova.it
Oslo	City of Oslo, Grensen 13, 0159 Oslo, Norway, https://www.oslo.kommune.no	Osman Ibrahim osman.ibrahim@ber.oslo.kommune.no
Institut za sigurnosne politike ISP	Institute for Security Policies, Kruge 9, 10000 Zagreb, Croatia, http://insigpol.hr	Krunoslav Katic krunoslav.katic@insigpol.hr
TIEMS	International Emergency Management Society, Rue Des Deux Eglises 39, 1000 Brussels, Belgium, https://www.tiems.info	K. Harald Drager khdrager@online.no
UniSMART Fondazione Università di Padova	Unismart – Fondazione Università degli Studi di Padova, Via VIII febbraio, 2 - 35122 Padova, Italy, https://www.unismart.it	Alberto Da Re alberto.dare@unismart.it

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

5. Specific guidelines – appended as separate documents

Each of the separate, stand-alone parts of the deliverable providing advice on specific topics (as listed in the table in section 1.3) will be made available at the end of the project as separate resources using the project website and/or other dissemination mechanisms. For the purposes of reporting at this intermediate stage, they are appended to this document.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 1: COSSEC Survey

Status: Initial Draft

Analysis based on a set of questions to be answered by COSSEC Members.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Tables

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Report on COSSEC Survey

This material contains the questionnaire to be issued to the COSSEC network in the Fall/Winter period 2021-2022. Once the survey has been conducted, the feedback will be thoroughly analysed.

	QUESTIONNAIRE ON ETHICAL ISSUES
1)	**Have you encountered ethical issues regarding some of your applications/products/projects so far?**
[] Yes, I have encountered ethical issues. [continue] [] No, I have never encountered ethical issues. [go to question number 5]	
2)	**What was your ethical issue related to?**
… (max 100 words)	
3)	**Why was it an issue?**
… (max 100 words)	
4)	**How did you manage to solve it?**
… (max 200 words)	
5)	**Are you able to identify or suggest any ethical issues that might be encountered in the future applications of IMPETUS?**
… (max 200 words)	
6)	**How would you manage these, or some of these, issues?**
… (max 200 words)	
7)	**What topic do you think could be useful to discuss in an open session with COSSEC members and IMPETUS partners?**
… (max 100 words)	
8)	**What will be the focus of the open session? Underline the possible involved partners.**
… (max 100 words)	

Table 1: Questionnaire on Ethical Issues

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 2: Public Trust in Digital Platforms and Personal Data Safekeeping

Status: Initial Draft

Citizens' Guide to the Estonian example of public trust in digital structures, access to data, transparency and personal data safekeeping.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Public Trust

1.1 Public Trust in Digital Platforms and Personal Data Safekeeping: Citizens' Guide to the Estonian example of public trust in digital structures, access to data, transparency, and personal data safekeeping.

As the societal uptake of digital systems is increasing, we are also becoming increasingly vulnerable to cyber-attacks, data breaches and giving away too much information about ourselves to third parties. In such a context, it is very important for governments and local bodies to embody security and privacy by design principles and rely on system architecture which empower citizens by putting them in control of their data.

The Estonian example is one such leading example in the European context, where citizens have adopted the digital platforms by a great majority. From online tax filing to voting for elections, Estonia provides almost all possible services online.

There are two fundamental building blocks in the Estonian rollout which can be considered as important factors in avoiding security and privacy breach incidents and building of high level of trust. The Estonian system relies on a state of art architecture which is decentralized block chain type systems which is called X-Road. The system is based on sound mathematical and cryptographic fundamentals which minimizes the threat to the Estonian state on the technical level.

However, that is not enough. Estonia also enables through a portal access to citizens of all their records online and a log of who has accessed their data. Such a transparency for citizens is very powerful and prevents abuse by people in high offices who may have access to such data to act responsibly.

This Guide will explore the level of general public's trust into digital safekeeping of personal data on the Estonian example (Estonia being one of the front leaders in digital agenda). The guide will attempt to draw conclusions relevant for IMPETUS Platform's users and other interested stakeholders.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 3: Protocol on Anonymization and Deanonymization

Status: Completed

Technical protocol on anonymization and deanonymization of data collected in social media.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Table of Figures

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Tables

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Technical protocol on anonymization and deanonymization of data collected in social media

1.1 Introduction

Data anonymization has been defined as a "process by which personal data is irreversibly altered in such a way that a data subject can no longer be identified directly, either by the data controller alone or in collaboration with any other party". Therefore, we consider data anonymization as the process of either encrypting or removing personally identifiable information from datasets, so that the people whom the data describe remains anonymous. Pseudonymized data can be restored to its original state with the addition of information which then allows individuals to be re-identified, while anonymized data can never be restored to its original state. Our tool has adopted the pseudonymization approach by default since some use cases consider that certain authorised parties (such as emergency services) would need to identify threatening individuals during an emergency. De-anonymization is the reverse process in which anonymous data is cross-referenced with other data sources (publicly available information or auxiliary data) to re-identify the anonymous data source.

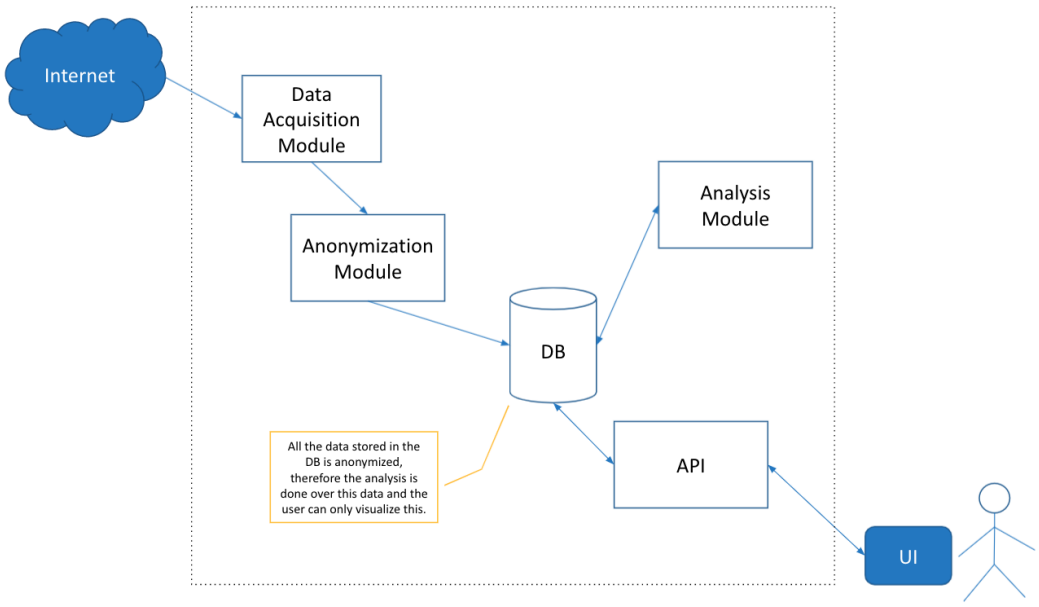


Figure 1: High level Description of Anonymization Process

The Social Media Detection Tool uses Advanced Encryption Standard (AES). A unique key is created for each client that is then encrypted with a private key and stored in K8s (Kubernetes) Vault for security.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

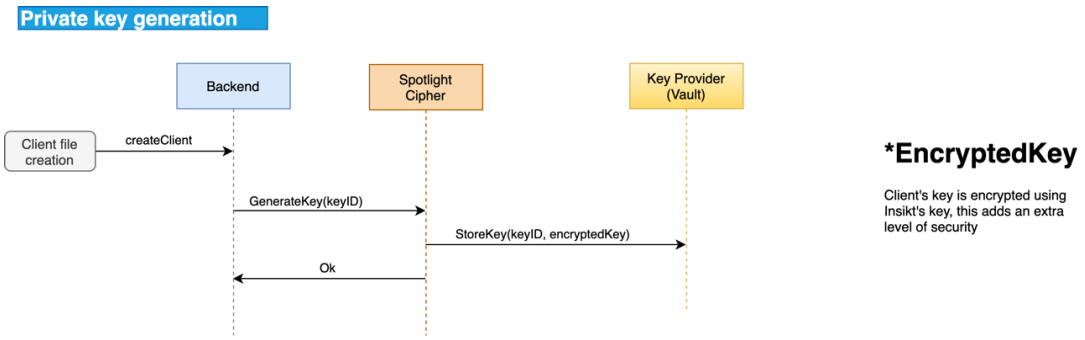


Figure 2: Private Key Generation Process

1.1.1.1 Data Anonymization

Spotlight, or SMD (Social Media Detection Tool), offers user data anonymization through symmetric encryption of predefined personal data (Advanced Encryption Standard). We consider this process as anonymization. Only the authorised users would be able to have access to a private key that would allow them to deanonymize in specific situations. The generated client key is encrypted using Insikt's key, adding an extra level of security.

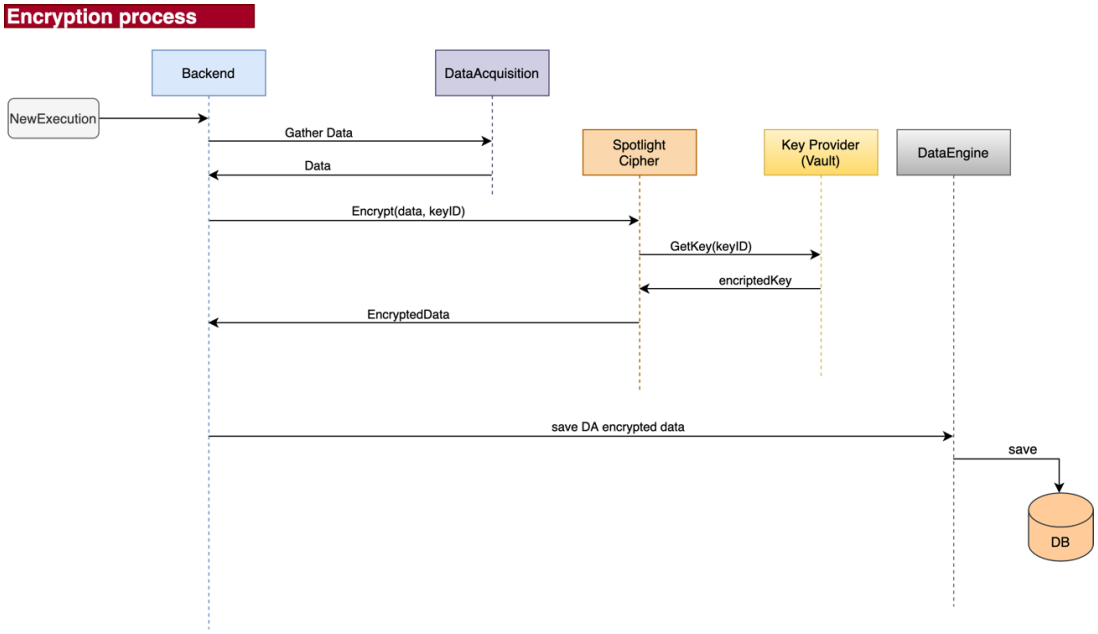


Figure 3: Full Encryption Process Diagram

Data anonymization will be performed in the back-end system after data acquisition and before bulk data into the database. That way, data will be stored and remain encrypted in the database.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

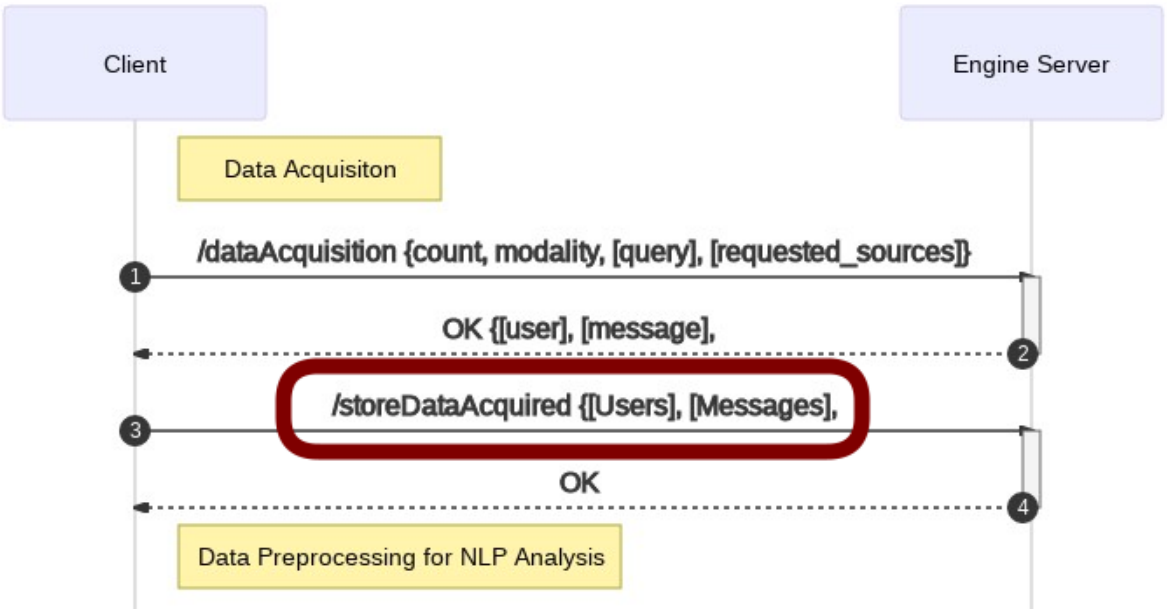


Figure 4: Encryption of Data after Data Acquisition

1.1.1.2 Data Anonymized

The below table gives an example of a database containing anonymized data.

Table	Attribute
Author	name
Author	nickname
Author	URL
Author	location
Author	gender
Author	birthday
Author	email
Author	website
Author	work
Author	university
Author	occupation
Author	phone
Author	address
Author	lat
Author	lng
Author	hometown_address
Author	hometown_lat
Author	hometown_lng

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Table	Attribute
Message	replytoauthor
Message	URL

Table 1: Data Anonymized in the Database

1.1.1.3 Data De-Anonymization

Data will only be revealed after an authorised user asks for specific user's information and introduces the authorisation password. Because of the GDPR, the de-anonymization process will be done one author at a time. There is no option of de-anonymizing a complete project altogether.

1.1.1.4 De-Anonymization Process

The process for de-anonymizing data is as follows:

- In the dashboard there is a button that when the user clicks on it a modal window appears.
- In the modal window there is:
  - A dropdown list with text input so that the user can look up and write the encrypted Username of the author that wants to be de-encrypted.
  - A password textbox so that the user will have to reintroduce her/his login password to reverify her/his identity.
- There will be a call to the backend endpoint with the information:
  - User's token.
  - User's password.
  - Project ID.
  - Author's UUID.
  - Reason text.
- The backend will verify:
  - If the user has permits to de-anonymize a project.
  - The password is correct.
  - The project is of her/his property.
- After verification, the backend will decipher the data:
  - Will look up the UUID in the Author table in the DB and will de-encrypt all the information regarding the Author.
  - Will look up all the Messages in the DB with the Author UUID and will de-encrypt the information except the ReplyToAuthor Field.
- The backend will store in a table in the DB a log with:
  - Timestamp.
  - Spotlight user ID.
  - Project ID.
  - Author UUID.
  - Reason.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

- The backend will return a JSON to Frontend (dashboard) with all the information that was de-anonymized.

The table below refers to the de-anonymized database data.

Table	Attribute
Author	name
Author	nickname
Author	URL
Author	location
Author	gender
Author	birthday
Author	email
Author	website
Author	work
Author	university
Author	occupation
Author	phone
Author	address
Author	lat
Author	lng
Author	hometown_address
Author	hometown_lat
Author	hometown_lng
Message	replytoauthor
Message	URL

Table 2: Data after De-Anonymization

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

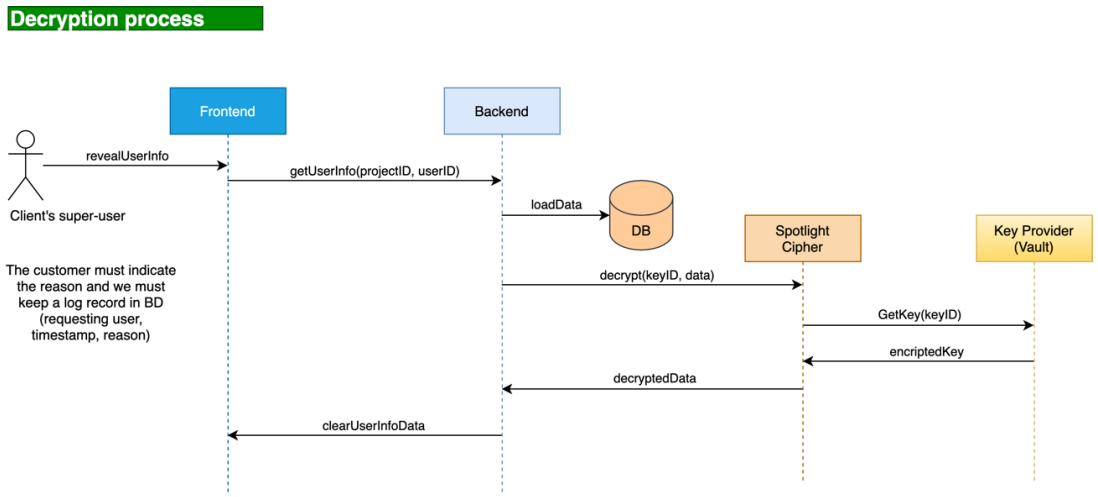


Figure 5: Decryption Process

The tool evaluation analysis will follow the IMPETUS Project's trials and evaluate the tools utilization and successful anonymization and de-anonymization procedures.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 4: Human & AI Teaming Brochure

Status: Completed

Citizens' Guide on Human and AI interoperability, biases, explainability and alignment issues.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Human & AI Teaming Brochure

1.1 Summary of Content

Covering the following topics:

a) AI and Human biases
b) Mutual explainability
c) Alignment problem

This brochure focuses on three ethics related issues with AI, which are broadly relevant. Our take on this is the human-machine teaming perspective, so seeing an AI as a team member, just like the human. First we talk about which biases each AI has, where they originate (mainly limited sample size and biased model). We mention the fact that these same biases are also present in humans. Then we describe the issue of explainability: which goes both ways; how does the AI explain itself to the human and how do humans make clear to an AI what our motivations and objective functions are. Finally, we discuss the alignment problem, which asks "how we can build AI that does what we want it to do, as opposed to building AI that will compromise our own values by accomplishing tasks that may be harmful or dangerous to us". This is the issue where ethical implications of use of AI are most explicitly discussed.

1.2 Introduction

The focus of this brochure is on two ethical issues related to AI. Our take on these is the human-machine teaming perspective, that is one where we see an AI as a team member, working alongside human team mates on designated tasks and pursuing certain goals. First, we talk about which biases in AI models and their originate, which are the data (limited sample size) and model bias. We observe that the same type of biases is also present in humans and elaborate on these biases and their possible consequences. Then, we describe the issue of bi-directional explainability, that is, how can an AI model be explained or explain itself to the human and how can humans make their motivations and reasoning clear to an AI model. We further sketch out some of the 'Explainable AI' (XAI) techniques, used by the various stakeholders.

1.3 Artificial Intelligence and (Human) Biases

Artificial intelligence (AI) is becoming increasingly present and important in our world. It is an artificial creation of intelligence, an attempt to create intelligence. One of the definitions [1] of intelligence is: "The ability to acquire, understand, and use knowledge." This, like most of other definitions of intelligence, is based on the ability of our (human) brains. In the search for answers to how intelligence can be created, it is imperative to understand how intelligence works. One of the aspects of intelligence is especially relevant in AI modelling as it currently functions and is used: the gathering and processing of information with the goal of creating knowledge, which can then be interpreted, understood, or used in tasks such as decision making. In the process of gathering, filtering, and interpreting information, the human brain is prone to taking shortcuts. These shortcuts are often legitimate (due to limited time and resources used for cognition), sometimes necessary (to be able to react to changes quickly enough in the environment, for example). In situations where large amounts of information need to be processed and we can afford a lengthy deliberation of the information, these shortcuts can be inconvenient or even dangerous.

One of the ways this 'shortcut taking' manifests itself is by what is called a bias. There are many known biases in human cognition [2, 3]. In the context of AI, biases related to information selection, filtering and

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

interpretation are of relevance. For example, take the so called “availability bias”, which means considering only the readily available information, or the “confirmation bias”, which is a tendency to look for information that confirms already present beliefs. Such biases show that our judgments, choice, and decision-making behaviours could be skewed and may lead to conclusions that could be false.

The AI models are created by human intelligence and appear to have somehow inherited some of these cognitive biases. The information, or strictly speaking, the data, which contains information, is what can be called the “input” of an AI model. These data are then processed by the AI. The “outputs” can be manyfold, ranging from simple filtering and views on these data, through deriving (statistical) properties of the data, to generated predictions, solutions or even decisions, based on these data. Two places where these biases originate are the data itself and the (AI) model used for processing the data to create the desired output.

As is the case with humans, no AI knows everything, though the amount of data an AI model can handle is vastly larger than that of a single human. This means that the data consists of only a part of all the potential data that could be collected. It is desired to form conclusions and make decisions based on what is assumed to be valid if ALL the data were considered. When only a limited portion of all the data is available it means that generalizations must be made. Generalizations can be valid if the available portion of the data is representative of the whole and can be called “unbiased” or “fair”. Therefore, the choice of which portion of the data is used decides how valid or skewed (wrong) the conclusions will be. The smaller the available portion of the data, the higher the chance to be wrong. This prompts the quest for unbiased data which is the data portion to enable valid generalizations. It is therefore imperative to choose the “right” data portion to serve as input for an AI model.

The AI model encompasses how the data is interpreted and which conclusions (or other outputs) will be drawn based on the data. The model decides for each new data point, whether it confirms or contradicts the modelled beliefs about the data. This belief is based on the input (like data points and corresponding labels) used for generating the model. If in the process of data labelling, which is done by humans, incorrect or contradictory labels are assigned, these beliefs can be skewed. This is where the model can be(come) biased, consequently producing conclusions which could be wrong.

Both in the case of data (portion) selection and creating AI models, it is important to understand the processes or the way these things are done, no matter whether it is done by a human or AI. Both are biased and prone to making incorrect assumptions or other mistakes. Therefore, transparency and control of these processes is key to make sure both the data and the model used by an AI model will be unbiased.

1.4 Mutual Explainability

The value of transparency in the design of AI models extends to its use. Wherever an AI model is used, it ultimately impacts humans, whether it’s the society as a whole or on a personal scale. Especially when AI is used as a decision support tool, or even elevated to the level of a teammate, working alongside the humans, which is increasingly likely with AIs growing complexity and capability.

In any form of such cooperation between humans and AIs, communication is required, and transparency in information use, reasoning and intent are key components of effective communication. Since communication is a two-way process, both involved parties (the human and the AI model) must be able to understand each other and explain themselves when prompted. The issue of explainability of AI is often referred to as “explainable AI” and abbreviated as “XAI”.

In case of cooperation between humans, our information processing and reasoning are similar. For any misunderstanding, it’s usual for humans to ask for additional explanation, making it possible to interactively clear up the issues in a (verbal) exchange. Such exchange of information, queries and explanations between humans and AI’s is more challenging, given the AI’s limited and artificial nature, the challenge of understanding (the explanations) of AI is higher than would be the case in a cooperation between humans alone.

Much of the past effort has been focussed on explaining the (workings of) AI to humans. Since AI models are artificial creations, this has been largely successful, though increasing complexity of AI (especially with Deep

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Learning algorithms), complexity of the problems it tackles or the limits of human comprehension, means it becomes harder to understand the AI's "reasoning" process. This also means that, considering earlier discussed biases, it becomes harder to judge whether the AI's output is valid or correct.

The other part of the two-way communication is the challenge of AI understanding the human, and the humans explaining themselves to AIs. Human insight into an AI model is usually more extensive than the other way around. Humans can look up information about the data used by AI and its reasoning algorithm, whereas such information of the human is incomplete or unavailable to the AI. This is often the cause of misunderstandings between humans and AI. If an AI model does not (or is not able to) inquire into underlying goals, intentions or beliefs behind human's queries or commands, it can have a hard time fulfilling these requests.

A complicating factor to the above communication challenge, are the (unpredictable) dynamics. This is less of an issue with AIs, since (disregarding occasional failure in operation), they are fairly stable. Humans' mood, patience or even cognitive ability can vary over time, however. Since the AI does not possess any means (sensors) to detect such changes, the humans can seem erratic and unpredictable (within certain bounds) to the AI. To compensate for this, new ways should be developed for AIs to be able to detect and interpret such variations in human behaviour, and abilities for AIs to (be able to) inquire for explanations, as well as being able to explain themselves and ensure a stable and solid communication.

It is worth noting that here, we focus on the user of AI, who is also the recipient of its output. There are (many) other stakeholders, with different relations with the AIs, with different needs for explainability, which are not covered here.

Much has been done in the explainable AI (XAI) domain to structure and analyse the issues as well as provide (practical) solutions or suggestions for approaches to explainability. There are two ways to approach explainability: globally and locally. Global explanations are focussed on providing insights into the workings of an AI model in general, so answering the question "how does this AI model produce the predictions". Local explanations are instead only interested in specific cases, so answering the question "how did this particular prediction come to be". It has been observed that the global explanation is of particular interest to Machine Learning experts, AI model builders or regulators, while the end user (often the recipient of the AI) is more often concerned with local explanations.

Such explanations can be provided on different levels and in multiple ways. Two noteworthy and prevalent ones, used particularly by users interested in local explanations, are the so called "feature importance" and "counterfactual explanations". Feature importance is akin to how humans often explain themselves, that is by providing the (most important) reasons for their decisions or actions. Model features are the characteristic properties of the (input) data, that contributed (most) to the model predictions (output). Identifying these key characteristics provides insight into the AI model. Counterfactual explanations are local explanations that identify alternative ("counterfactual") input data that closely resemble the actual input data used to generate the particular prediction being investigated. This answers the question of "how should the input data need to change in order to get a different output from the AI model". This technique is also useful to investigate model robustness, security, or fairness issues.

An issue that complicates explainability lies in the (mistaken) assumptions about the workings of AI models. Humans (often) expect an explanation that involves a reasoning of how the outcome (model output) is causally related to the input, whereas (most) AI methods are based on correlational inference, that is they base their predictions on correlations in the data. This mismatch makes understanding each other difficult, and it is the lack of understanding that is usually the reason behind the queries for explanations.

Whether an explanation is accepted, often depends on how plausible or trustworthy the explanation seems. If the provided (local) explanations align with our expectations, we are inclined to accept them and to trust the (particular output of the) AI model. Opinions of subject matter experts are still the reference (or the "golden standard") here: if the experts agree with the explanations, the model can be accepted as sound or trustworthy. This partly explains the preference for explanation techniques which are similar to the ones used by human (experts).

Finally, it should be noted that, from the Human-Machine Teaming perspective, the XAI research and methods have (until now) been focussed on the explainability of AI towards humans, and much less on the reverse, that

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

is human explainability towards AI.

1.5 Selected Sources of Inspiration and Reference

[1] https://www.thefreedictionary.com/intelligence

[2] https://en.wikipedia.org/wiki/List_of_cognitive_biases

[3] Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

[4] Bhatt U et al. 2020 Explainable machine learning in deployment. In Proc. of the 2020 Conf. On Fairness, Accountability, and Transparency, pp. 648–657. https://doi.org/10.1145/3351095.3375624

[5] Aitken, M., Toreini, E., Carmichael, P., Coopamootoo, K., Elliott, K., & van Moorsel, A. (2020). Establishing a social licence for Financial Technology: Reflections on the role of the private sector in pursuing ethical data practices. Big Data & Society. https://doi.org/10.1177/2053951720908892

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 5: Analysis of Platform's Tools

Status: Initial Draft

A set of questions for tool developers, analysis of employed technical standards and protocols, of potential use for other WPs.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Tables

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Analysis of Platform's Tools

To ascertain potential ethical and legal issues in connection to the use of IMPETUS Platform's tools, it is necessary to collect certain information as to the nature of such tools and nature of data being manipulated (collected, analysed, stored, shared, etc.). To that end, a survey has been prepared. The survey contains a set of questions for tools' developers. The feedback will be thoroughly analysed.

Questionnaire for Technology Partners	
No.	Question
1.	What sort of data are your tools/equipment capable of collecting?
2.	What sort of data are your tools/equipment going to collect for the project IMPETUS? (pilot activities, other activities)?
3.	Does your tool collect personal data? If the answer is yes, please specify the nature of personal data and their volume:
a.	Are there any Documents which certify the adoption of the principle of "Privacy by Design"? Understand the goal of the tool and which actions have been put in action by the data controllers to protect the privacy of data subjects.
b.	Has a suitable legal basis been identified for the processing of personal data? Have the data subjects provided some authorization to use the data? Which purpose does this authorization refer to?
c.	Can personal data be disclosed by the tool/IMPETUS platform?
4.	What sort of data do you anticipate your tools/equipment will collect through the IMPETUS Platform once it is completed?
5.	Are you providing software or hardware solutions, or both? Are the privacy roles defined? Is every person involved as data manager/developer who works on the project educated in privacy matters? NDA should be signed by the developers/ data managers.
6.	If you are providing software (either individually or packed with hardware):
a.	Does that software contain machine learning, deep learning or other kinds of advanced algorithms?
b.	Can you briefly describe how such algorithm(s) handle:
i.	Data collection?
ii.	Data analysis?
iii.	Data protection (if applicable, and related to data storage, data access and data sharing)?
iv.	Decision-making (if applicable) / Human-in-the-Loop or similar concepts?
v.	Learning capacity?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Questionnaire for Technology Partners	
c.	Does your software provide a solution where an advanced algorithm (machine learning, deep learning or other):
i.	Independently makes decision based on collected and analysed Big Data?
ii.	Raises alerts to be decided upon by human operators?
iii.	Analyses the data without any direct feedback issued to the human operator? In this case, make sure that the algorithm allows, according to transparency, an effective control by the data subjects.
d.	Is the human operator capable of understanding the feedback information received by an advanced algorithm (either a decision, proposal for a decision, alert, or other kind of feedback) and understanding how an advanced algorithm has produced that feedback?
7.	Are you providing a product (i.e., you sell a tool or a piece of equipment) or a service (you are collecting and analysing data, or doing something different for the client)? In this case, understand and define the type of personal data and draw up, here too, privacy policies.
8.	In case you are selling a product/solution or a limited service, do you provide training services for buyers/deployers (for their personnel that will be operating your tools/equipment) of your product/solution?
9.	Can your tools/equipment function as a standalone product completely and solely utilized by the user/deployer, or does it require your (developer) presence and activity?
10.	Do you provide support services that include items like backup service, technical service and other? Do you have any data breach procedure defined?
11.	Is any of the previously mentioned data, collected by your tools/equipment, personal data or can it potentially lead to the identification of individuals?
12.	In case your tools/equipment are collecting personal data:
a.	What systems (inherent to your tools/equipment) do you have in place to protect the privacy of individuals? Here, for accountability (principle foreseen by the privacy discipline), you should provide a short report.
b.	Are individuals whose data is collected, analysed and stored informed prior to such operations?
c.	Are individuals whose data is collected, analysed and stored informed during such operations?
d.	Are individuals whose data is collected, analysed and stored informed after such operations?
e.	Do your employees have access to personal data? If so, have they been trained in accordance with the provisions of the privacy legislation? NDA should be signed by the employees who have access to the data
f.	Do your clients have access to personal data? If so, what are the security measures in place?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Questionnaire for Technology Partners	
g.	Does an advanced algorithm (machine learning, deep learning or other) have access to personal data? In this case, make sure that the algorithm allows, according to transparency, an effective control by the data subjects.
h.	If either the advanced algorithm or your employees have access to personal data, how do you ensure the protection of personal data and privacy of persons whose data has been accessed? If yes, please indicate the list of security measures, according to the GDPR.
13.	Do you have specialized protocols in place concerning (if applicable):
a.	GDPR compliance?
b.	Other legislation compliance (if having worked with public security bodies' in security operations)?
c.	Data collection?
d.	Data analysis?
e.	Data storage?
f.	Data access?
g.	Data storage?
h.	Data anonymization/de-anonymization/pseudonymization?
i.	Cybersecurity?

Table 1: Questionnaire for Technology Partners

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 6: Alerts and Information Generated by Tools

Status: Initial Draft

Citizens' Guide on the Interaction between machine learning algorithms and human operators, relevance and consequences of false positive and false negative alerts, wider societal implications.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Alerts and Information Generated by Tools

This content/material represents a citizens' guide on the interaction between machine learning algorithms and human operators, relevance and consequences of false positive and false negative alerts, as well as wider societal implications.

The guide will cover the perspective of citizens and wider society when decision making is influenced by machine learning algorithms which otherwise was sole purview of human authorities. For example, an item carried by a citizen in a city like Oslo that may be wrongly classified as a gun, could lead to public embarrassment, especially if such persons are accompanied by family and friends. Or even weirder, imagine a scenario where a t-shirt (clothing) of the person contains a picture which the algorithm wrongly detects as an object that they are carrying, would it be possible that such a person be stopped by police every time that they wear such a t-shirt? This may lead to harassment or public shaming in some ways. Especially if the person comes from an ethnic minority which may further heighten the decision-making bias. AI algorithms such as in cars like Tesla are already shown to be vulnerable to such sticker-based attacks.

When Google launched their image labelling feature in the Google photos product, it led to huge embarrassment for the company, as it wrongly classified black women as Gorillas. There are several ways by which we can improve. This may be better checked with diversity in training of input data, cities may consider asking the AI solution providers on providing metadata of data trained to see if data is diverse or over represents certain class over others. The solution providers could check if there are no race or gender bias in the trained systems by providing testcases. Also, in the post deployment phase one should consider the false positive cases a bit more carefully from the wider societal perspective so that the AI tools may not end up becoming a harassment mechanism for citizens.

The guide will be completed following the IMPETUS Project’s trials and subsequent analysis.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 7: Human Computer Interaction tool

Status: Completed

Citizens' Guide on potential biases in human (mental) workload assessments, explainability by assessment, and alignment problems related to HCI tool.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Table of Figures

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Human Computer Interaction Tool

1.1 Summary of content

a) Potential biases in human (mental) workload assessments (model is based on calibration tasks; bio sensing is individually determined – and not averaged over subjects)

b) Explainability by assessment (signal quality / assessment context eg. based on other IMPETUS tools)

c) Alignment problem related to HCI tool

In this document, we first describe the HCI tool, its intended context of use and purpose. Then, we address the three AI related topics from the Human & AI Teaming brochure and talk about how the HCI tool deals with them. We attempt to limit the biases by having a work assessment model that is personal (based on individual biomarkers) and basing the model on the operator's actual daily tasks/activity.

The HCI tool deals with explainability in three ways: The AI model is fully documented, based on literature and previous experimental findings (high level AI and human explainability), it reports back its reliability of signal measurement (low level AI explainability) and the suggested inclusion of human input and input from other Impetus tools (via alerts), telling the AI when an unusual occurrence has taken place, takes care of the third point (human explainability towards AI). Finally, the alignment problem is addressed in the design rationale of the HCI tool and its intended use, especially the use of the assessment feedback and how this relates to operator performance, subsequently leading to improved level of security in a smart city context.

1.2 Introduction

In this document, we first describe the HCI tool, its purpose and intended context of use. Then, we address the two AI-related topics from the “Human and AI teaming” brochure and discuss how the HCI tool deals with them.

1.3 Biases

We limit the biases in the HCI tool by implementing a workload assessment model that is personal (based on biomarkers belonging to the individual) and based on the operator's actual daily tasks/ activities.

1.4 Explainability

The HCI tool deals with explainability in three ways. First, the AI model is fully documented, based on literature and previous experimental findings (high level AI and human explainability). Second, the model reports back its reliability of signal measurement (low level AI explainability) and the (optional) inclusion of human input and/or input from other Impetus tools (via alerts). Human input, which tells the model when an unusual occurrence has taken place, for example, takes care of the third point, that is human explainability towards AI.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1.5 HCI Tool

The Human-Computer Interaction (HCI) tool is a human support tool, aimed specifically at professionals working in high stress, high risk and high impact settings, where time is of essence, such as operators of command and control (C2) rooms. The tool comprises hardware (biosensors and computers) and dedicated software, which aims at assessing the operator's mental workload and stress level and to provide feedback, especially when those levels cross undesired thresholds. The achieve this, the HCI tool uses personalized models, created and calibrated specifically for each individual user, using the data collected by the biosensors.

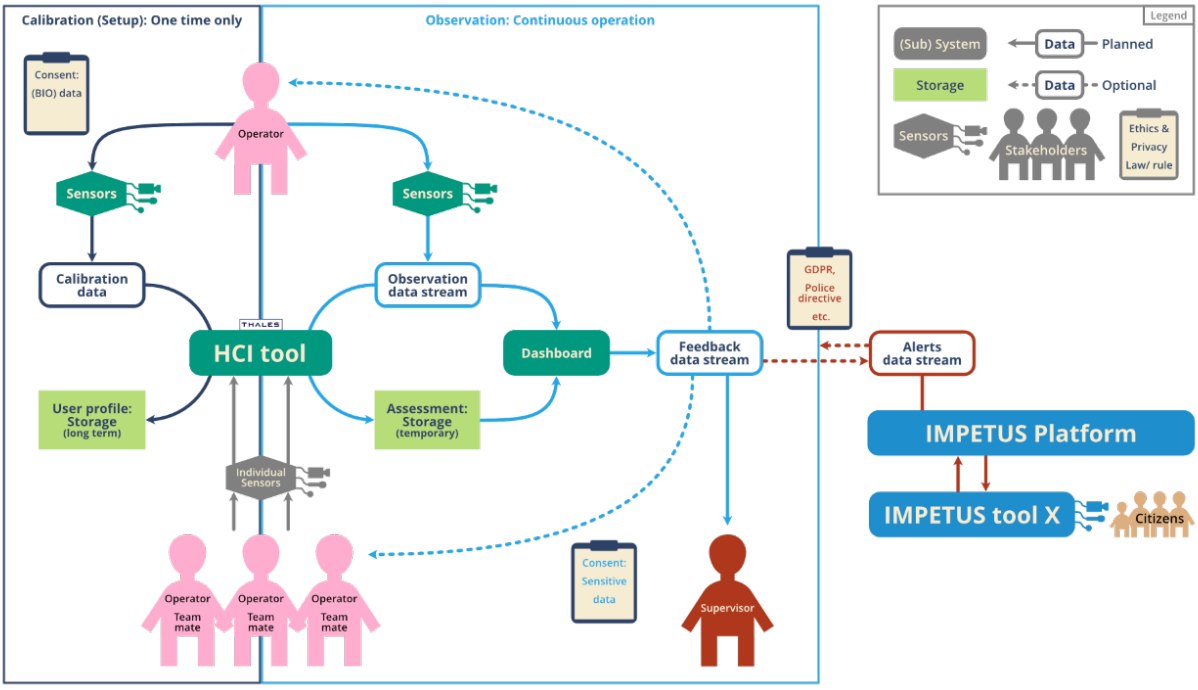


Figure 1: HCI Tool

The graphic below summarizes this continuous loop of data collection, processing and feedback generation to the individual user, who (in most cases) is part of a team.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

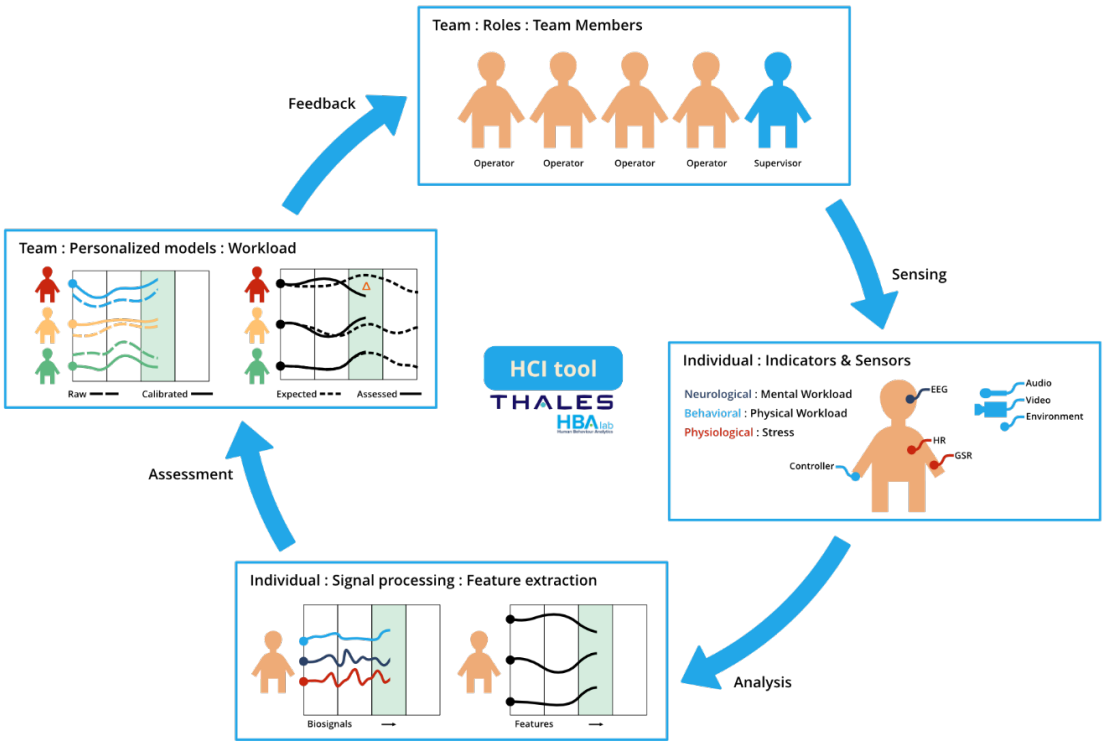


Figure 2: Continuous Loop of Data

The issues, related to AI described in the "Human & AI Teaming" brochure were data and model biases and mutual explainability between the human user and the AI in human-machine teaming. Here, we describe how these issues are addressed by the Human-Computer Interaction (HCI) tool.

1.6 Data and model biases

The HCI tool can be seen as a two-way communication system. The sensors measure the user's bio-signals and send these data to the tool. There, the model is employed to process and interpret these data to generate an assessment of the user's mental workload. This assessment is then communicated back to the user. Both the data and the model should be unbiased.

The HCI tool's personalized model is created prior to the measurement-assessment-feedback loop. To create these models, a specially designed calibration task is performed by the users. This task is designed to mimic their normal tasks/ activities, as close as possible. While the user is performing the calibration task, biodata (physiological and neurological) and behavioural (self-assessment and observation) are collected. These data are particular to the individual and his/her natural working environment and task. The model personalization is based on these data and so a unique model is created for each individual. Once the model is created, it can be used in the HCI tool during the continuous loop described above. The data collected during the use of the HCI tool is of the same nature as the data collected during the calibration. It is therefore specific and limited to the individual it's collected from and the task they perform. This way, the biases in the data and in the HCI tool models are minimized.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1.7 Transparency and explainability

As explained above, the HCI tool is a two-way communication system. Biodata from the user is communicated to the tool and assessment of mental workload is communicated back to the user. In both communication streams transparency and explainability is guaranteed. In the case of the HCI tool, explainability extends further than to the user alone: the technical expert or maintenance personnel require different levels of explainability, which the tool provides.

The HCI tool deals with explainability in the following manner:

There is full transparency about the data acquisition (sensors specifications), data storage and processing pipeline as well as the biomarkers used for assessment generation by the model. The calibration task itself is documented, as is the rationale behind its design. The task uses the same sensors and collects the same data as the HCI tool during its operation (assessment and feedback generation).

The model (used for the generation of the assessment) is based on scientific literature and previous experimental findings and its algorithms (including any ML components) are documented. One of the most used XAI (Explainable AI) techniques, namely feature importance, is also available to gain insight in the workings of the HCI tool. There is full transparency about which features the workload assessment is based on, for each sensor (bio data modality) employed by the HCI tool.

All this information attributes to global model explainability and is more likely to be requested by expert users, regulatory bodies or system integrators, rather than the end user (the recipient of the assessment).

The feedback generated by the HCI tool, is visual and provides the following information supporting explainability:

- The current assessment as well as the recent (high-level) data it is based on. This provides high-level insight into the generation process and subjective reliability of the model.
- In case of detection of unusual sensor readings, the tool provides (real-time) feedback in the form of alerts, to proactively inform the users of possible malfunctions or incorrect use of the sensors. This provides insight into the reliability of the signal measurement and as such, acts as a factor enabling explanation of the model outputs (i.e., the assessment).

These contribute to local explainability and are usually mostly requested and consumed by the end users and recipients of the model's outputs.

Additionally, the HCI tool could incorporate other information as input, during the generation of assessment. This operation could be manual (done by the human) or automatic (determined by other means used within the same and context and given in the form of alert messages from other context sensitive tools). Such additional data would inform the model if an unusual occurrence, which has not been modelled, has taken place and can thus more fully explain the data coming from the biosensors. This addresses the point of human explainability towards AI.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 8: Potential Issues of the Thresholds of Risk Egress Algorithm

Status: Completed

Dissemination Status: Public

n

Description of the Risk Egress Algorithm, definition of emergency threshold, and list of potential ethical issues regarding the identification of the threshold values, the threshold manipulation, and the value of the risks that may influence the risk egress assessment

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Table of Figures

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Thresholds of Risk Egress Algorithm - Potential Issues

1.1 Crowd Simulations

The simulation of crowds and dynamics of groups of people is of crucial interest in the framework of safety and security of Smart cities. As an example, the PTRO (Physical Threat Response Optimization tool) in IMPETUS will be supported by specific simulations of crowd movement. They can be used to optimize their management but also to enhance the planning step of authorized and even unauthorized events. The simulations are aimed at proposing suitable guidelines and management indications to be implemented before, during, and after a specific scenario.

IMPETUS may benefit from structured information related to:

- the time required for a group of people to leave a specified area
- the dynamic that follows an initiating event (fire, explosion, gun shot, …)
- the planning of escape routes and the evaluation of suitability of emergency corridors

1.2 Frequency of Simulations

Simulations are expected to support the management of crowds in the case of hazardous scenarios. According to the PTRO discussion with partners, the aim is to run simulations on reference agreed scenarios based on the city of Padova and Oslo, with special focus on the location of the pilot. These scenarios are formulated according to the following parameters:

- Total number of people involved in the reference scenario (up to 5000 people for the Padova scenario), divided into different classes of numerosity (1000, 2000, 3000, 4000, 5000)
- Availability/unavailability of escape routes

In this way, related management indications are obtained and included in the Impetus platform through the PTRO tool interface. Indications include egress time, most suitable escape routes to be used, and people density maps. Different simulated scenarios are linked to alerts of different degree (green, yellow, orange, red) to drive the best actions and preventive/mitigative actions. These simulations are intended as pre-loaded, and no real-time update is expected. Reference scenarios are discussed among Impetus project partners, and emergency operators. From an ethical perspective, no relevant issues are expected because all simulations are anonymous. No details on people orientation, race, and habits are used and information required for the simulations are general and only related to:

- Total number of people
- Draw/map of the context to be simulated only with respect to geometric configuration and escape routes
- Presence of obstacles and/or hindered escape routes

However, if people counters are expected to be put in place in relevant areas of the cities, ethical issues should be discussed and addressed. This eventuality and the availability of such a technology is still under discussion within the Impetus consortium.

1.3 Algorithm and Approaches to be Used

Different algorithms are available both in literature and on the commercial framework to address the topic of crowd simulation. The basic common idea of algorithms is to model the people flow via hydraulic approaches

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

that can provide the expected statistical trajectory and behaviour of a single person among other people. Psychological elements related to "crowd behaviour" are included in such a way that the mere hydraulic modelling approach is supplemented with additional aspects included during the flow of many people. The following aspects affect the choice of a specific algorithm:

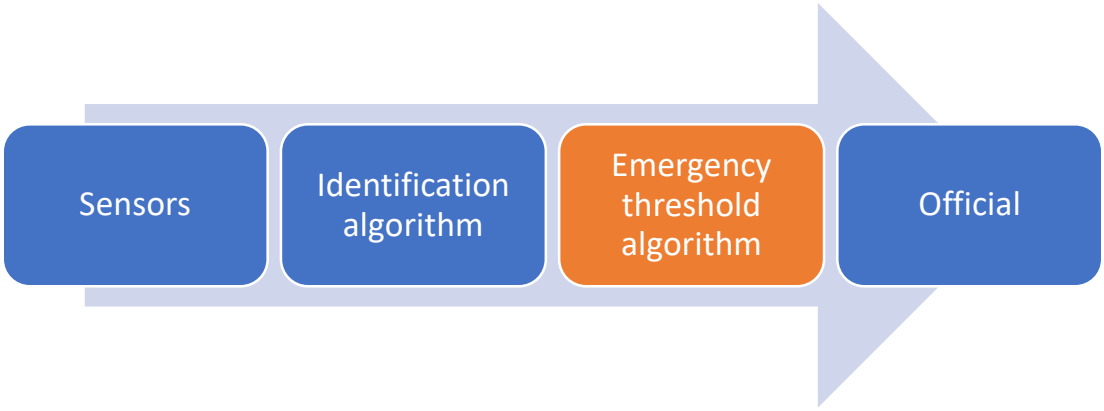
- Number of people (5000 people can be tracked with detailed approaches)
- Extent of geometry (detailed geometries may ask for improved model capabilities. Padova and Oslo reference areas can be still approached with detailed algorithms)
- Computational burden (this topic is relevant only for real-time updates. This option is beyond current intentions)

1.4 Links to other IMPETUS Platform Tools/Technologies

The activity of basic crowd simulation, identification of reference scenarios, and the formulation of guidelines fall within the PTRO tool purposes. A proper interface will be provided by the technological partners and may consist of pre-formulated indications, maps with relevant indications to be implemented and guidelines. Depending on the readiness/availability of sensors, CCTV, … the crowd simulation activity can benefit from retrieved sensors data that may drive to the most pertaining pre-simulated scenario and related indications.

1.5 Emergency Threshold Algorithm

It is difficult to identify if an element detected via the sensors installed in the city is a hazard. For sure, if a camera identifies a person holding a gun, this event is clearly a hazard. In other scenarios, a certain information may result in a "light" hazard, not dangerous enough to alert the emergency services, but if this "light" hazard is added to other "light" hazard, the overall sum of the contributions may result in a dangerous scenario, which requires the emergency services to be alerted.



The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Figure 1: Emergency Threshold Algorithm

1.6 Number of Control Points

The city/the area to be monitored is discretized in multiple points, each of which is used in the threshold algorithm. The number of control points affects the precision of the algorithm itself: with very low points it is difficult to identify a hazard with precision. These points may be in road intersections, in places where a sensor is located (e.g., a camera), or where specific information are located (e.g., the GPS coordinates where the phone carrier information are provided). The definition of the control points may result in ethical issues:

- Who decides where the points are located? Is it limited to the sensor positions?
- Is it possible to modify the points? Who has the right to do it?
- Can "fictious" points be included in the algorithm? For example, the emergency services may be interested in the hazard level of a location without sensors: if the hazard of the nearby points "spread" on the territory and their different contribution sums in such fictious points, a finer control on the territory can be performed. The choice of the placement of such fictious points may result in ethical issues (e.g., a higher control at the addresses of certain persons)?

1.7 Hazard Threshold Values

The algorithm uses Gaussian curves to estimate the hazard level at a control point. Different Hazard thresholds may be set, based on the gravity of the hazard. This aspect is for great relevance for ethical issues: like PID Control (Partial, Integral, Derivative Control), the threshold values are to be calibrated based on the experimental data and on the experience. As a result:

- Who chooses the initial hazard threshold? How can it influence the response of the emergency services?
- How many thresholds are to be set? Which official is in charge of the reception of a specific threshold?
- Who can modify the threshold? Are they editable by all the emergency services?
- Different emergency services require different thresholds?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

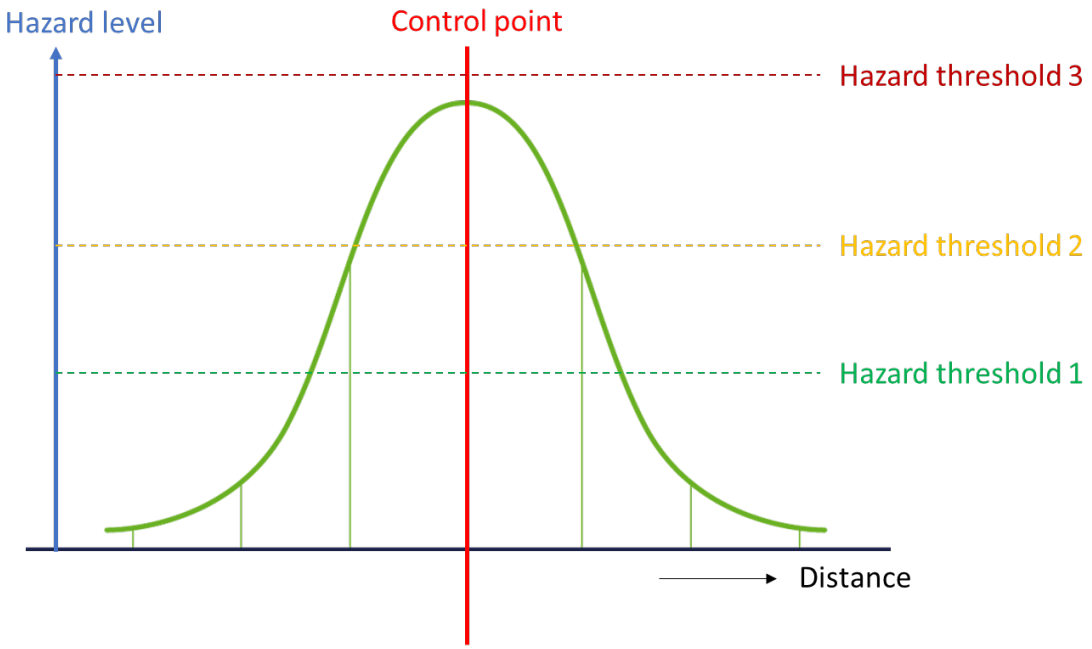


Figure 2: Hazard Threshold Values

1.8 Algorithm Estimation

The algorithm itself is useless without the input from the sensors and/or data aggregators. However, it is the algorithm duty to handle the inputs and translates the received data in "hazard level". Such translation is directly linked to the hazard threshold values, since different translation may result in different difficulty in reaching the thresholds. For example, let us consider a Hazard threshold 1 of 20 (dimensionless), and a single control point in which is installed a camera that simply counts the number of persons in a small area. If each person increases the hazard level by 1, when 20 persons are in the area, Hazard threshold 1 is reached. On the contrary, if one person increases the hazard level by 0.5, 40 persons are required to reach the threshold. This aspect is further complicated if multiple hazard sources are considered: each one of the sources can be handled separately, but the designer has to consider all the possible sources to calibrate each "hazard importance". As a result:

- Who chooses the "hazard importance"?
- The different hazard sources may vary its "hazard importance" based on the control point location?
- Who has the right to modify the values?

1.9 Mathematical Aspects

As previously stated, the algorithm uses Gaussian curves to estimate the hazard level at a control point. However, there are different parameters in the mathematical description of the Gaussian curve:

- Standard deviation (i.e., how largely it spreads when the distance from the control point increases?)
- Symmetry (i.e., the curve spreads evenly across all directions?)
- Direction of spread (i.e., the curve spreads in all directions or only along the roads of the city?)

The last point is the result of a simple consideration: if some sensors are watching a city centre road (let us call it road A), it is more likely for a hazard to spread along the direction of the road and not perpendicular to it (since there are buildings on both sides of the road). In this sense, a road which is parallel to road A (e.g.,

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

road B) but is divided from the first one via a dense number of buildings, may be not influenced by its hazards. For sure, this aspect depends on the hazard itself: a car accident is confined on the road where the accident occurs; a fire may spread through the adjacent roads through the buildings. All these considerations are to be discussed in an ethical point of view.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 9: A guide for citizens: on the ethical and privacy aspects of IMPETUS

Status: Initial Draft

Citizens' Guide on the IMPETUS Platform, advantages and better capacity for more efficient workflow, better capacity to avoid mistakes and bias, dangers and fears, possible abuse, safeguards, and similar.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. A guide for citizens: on the ethical and privacy aspects of IMPETUS

A guide for citizens: on the ethical and privacy aspects of IMPETUS intends to represent a set of short questions and answers related to the key ethical and privacy aspects of IMPETUS Platform. It's purpose is to provide a summary and informative explanation on how IMPETUS Platform affects citizens' personal data, and what are the main benefits of utilizing the IMPETUS Platform.

The current content is the initial draft of the issues and information that will form part of the guide. Once all relevant information has been identified, it will be transformed into the frequently asked questions format.

The general information explaining the relationship with privacy and ethical principles followed by IMEPTUS is important to increase citizen's trust and confidence in a city council or government. The main objective of this document is to act as a practical guide for city residents who want to know how IMPETUS platform tools follow ethical and privacy aspects to assist in enhancing the resilience of smart cities in the face of security threats in public spaces. It focuses on technologies and data utilized in the IMPETUS tools. The practical guide provides a simple overview of the IMPETUS platform architecture, describes how ethical and privacy principles are followed in collecting, processing, and securing data, and describes the incorporated ICT technologies.

This document presents an initial draft of a city resident's guide on how IMPETUS takes care of ethical and data privacy issues affecting citizens' rights. The current version of the document outlines a number of criteria for creating a practical guide, outlines a methodology and and discusses potential digital formats (brochure or video).

The following criteria are considered for drafting a practical guide:

1. Identify and gather information on national and European laws applicable for IMETUS platform
2. From each tool used in the IMPETUS platform: investigation of how they collect data and the system implemented to guarantee resepect of privacy and ethics. The more focus will be given on the distinction between data (anonymized) and information (that may be related to a specific individual). In any case, we evaluate how they act in order to guarantee citizens privacy. The tentative methodology is as follows:
    a. Type of data collected
    b. Information collected
    c. Possible risks and threats to privacy and ethics → A scenario that describes an event and its consequences, valued for its severity and probability
    d. Risk management → all the coordinated activities implemented to address these risks
    e. Guidelines for a correct use of the tools on an ethical level
3. Study privacy management guidelines to be used even in other smart city platform/data and bigdata collection situations.
4. Converting information from step 1 and step 2 into non-technical (self-explanatory) format to be communicated to the citizens: As some tools may be felt as very intrusive, we need to be bulletproof on a privacy level even when communicating the project to citizens. Some concerns may be raised for the collections of large amounts of data sets, and we need to explain to the citizen how the platform works and how we manage to protect citizens data. To this regard, a brochure seems a useful tool, where it will be explained in a non-technical (but still not too vague) way the use and the benefits of IMPETUS. An animation video may be very useful as well. Furthermore, the goal is not only to reassure citizens about their privacy protection, but also to highlight all the benefits, both in terms of efficacy and of ethics, that IMPETUS aims at bringing with its tools. IMPETUS, by providing to the

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

emergency operators a more complete and precise set of information, will help the operators to face in the best way possible emergency situations, avoiding bias or misunderstanding due to lack or partial information.

Both the brochure and the video-based approaches are to explain to the citizens what IMPETUS as a platform is, which tools the police/municipality will use, and which benefits it will bring to the city. The brochure consists of following steps for the development -

- Practical example of how the platform and its tools will work. We may use the scenarios (or part of them) that will actually take place in the IMPETUS exercises. Using these examples we would like to underline that the goal is to make the public security or safety entities (such as police) more efficient and more "precise" (with following national regulations). In this first part, the goal is to focus on highlighting improvements and advantages, not to raise any fear or doubt about ethical and privacy issues.
- Short description of the tools if allowed to disclose by national regulations
- A list of the legal compliances, national laws and local regulation
- List some Frequently Asked Questions (FAQs) from a citizen's perspective about ethics and privacy such as:
  - Will the police collect data about citizens?
  - Can it be used to control them?
  - Will citizen's data be collected by some private (non-governmental) bodies or enterprises?
  - How is privacy guaranteed if data is processed and stored?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 10: Overview of General Ethical Issues

Status: Mature Draft

Citizens' Guide on ethical issues arising out of collection and manipulation of data in security operations.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Table of Figures

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Tables

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Overview of General Ethical Issues

1.1 Introduction

The purpose of the following report is to examine potential general and individual ethical requirements relevant both in the IMPETUS Project's context and in broader considerations. The report will analyse the relevant wider issues in connection to the use (collection and manipulation) of "Big Data" in security operations. The ethical analysis in question analyses the implications of current capacities in data gathering to the personal data rights in the context of security and intelligence data-gathering operations.

The initial research into general ethical and legal issues (general analysis of smart cities, smart technologies, potential stakeholders, etc.) has identified a number of general ethical issues and considerations, as reported below. Further analysis will follow the city pilots.

1.2 General Ethical Questions

The following is an excerpt from D1.2:

"*The concept of a smart city, among others, includes tools and methods to enhance urban areas' security (Clever et al., 2018; Vogiatzaki et al., 2020). One typical example of an intelligent security hub is the City Office of Homeland Security, City of New Orleans (New Orleans, 2021), United States. The system pulls live feed from 400 city-owned and 150 business-owned and private homes-owned closed-circuit television video surveillance systems (CCTV)s. The system is connected to the emergency telephone service, with the built-in automatic screening of all locations relevant to incoming emergency alerts or calls for assistance. Thus, the system allows immediate real-time visual access to the emergency location (situational awareness) before emergency service arrival. Also, it enables historical footage, which is particularly relevant for investigatory purposes (investigation enhancement). The Artificial Intelligence (AI) system utilized by the real-time crime centre is an advanced machine learning algorithm capable of analysing the recorded footage without human interference (descriptive artificial intelligence). It offers advanced identification methods in real-time and historical footage (including shrivelled targets' behavioural aspects; Timmermans (ed.), 2009). The AI system is non-stop active (non-stop surveillance). The algorithm learns to recognize and separate normal from unusual behaviour/motions in specific areas/circumstances by monitoring the stock footage. The data collection is thus enhanced.*

A real-time footage recording is analysed, allowing subject profiling (subject profiling; the so-called laser analysis: behavioural patterns, associations, property, interests, and others). A question is raised to what extent is the AI algorithm in charge of recognizing suspicious occurrences and altering human operators of suspicious circumstances. This issue is related to concerns regarding the use of algorithms (and bias) in crime prediction and facial recognition, predictive policing in general, place-based predictive policing (high-risk areas crime patterns), and the notion of pre-emptive justice. The data utilized to generate early profiles is based on the already existing data in police records. This practice theoretically enabled the transfer of the established police record bias (the so-called over-policing in areas usually characterized as minority areas, low-income areas, and similar) into AI algorithms (Ferguson, 2017). With the advancement of technology, the body-worn cameras, and other smart devices (the so-called connected police officers) and drones (utilized by police) will easily incorporate the AI technology.

Besides, the AI system will become increasingly interconnected with various data sources (Security Internet of Things; i.e., license plates registry, parking systems, hotel registrations, transport systems' datasets, other publicly and privately owned CCTVs, various sensors, social networks, and others). The social network data set is particularly data abundant. Besides the usual exchange of posts and messages, the metadata, cookies, web scraping, and text mining can reveal identifiable data, such as user information, user location, and others (Menzer et al., 2015). The described technology, to a large extent already employed in practice, brings the surveillance and information-gathering capabilities to a new level. Such a database is continuously growing, incorporating information on surveillance targets and many other individuals not necessarily to any extent a

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

part of criminal investigations."

Based on the above-described security hub, this segment will evaluate the Project's pilots and endeavour to critically analyse factors represented in the table below. The analysis will proceed with the completion of pilots.

Consideration Points	Potential ethical and Legal Issues
"Live Feed from CCTVs" Live feed is collected from both public bodies owned and privately owned CCTVs. Public bodies owned CCTVs are not restricted to security organizations' systems but encompass the entire CCTV system operated by the city authorities and possibly other public organization.	The live feed is providing raw data, including a variety of personal data so collected, by means of accessing and utilizing both private and public data collection mechanisms and tools. Various questions arise, including: • What are the legal grounds for security apparatus to access both public and private CCTVs feeds? • What policies and protocols are in place that regulate the use of such data? • Are there any specific restrictions regarding the manipulation of such data? • Who supervises such activities? • Are such activities non-stop, or is the raw data analysed only when necessitated by a security alert? • Does the system utilize a machine learning algorithm, and what protocols are in place regarding the use and functioning of such an algorithm?
"Situational Awareness" The system enables a real-time situational awareness feed.	The system basically allows a plethora of actors to access raw data. Various questions arise, including: • Are there safeguards and protocols in place concerning the protection of personal data accessed by individual actors who use the system? • Are actors properly trained and informed concerning the sensitive nature of personal data? • Is the access to situational awareness feed restricted by any means?
"Historical Footage" The system enables access to historical footage.	The system records live feed and makes it available for later analysis and investigation. Various questions arise, including: • Are there any protocols in place concerning the storage, access, deletion, and manipulation of data? • Are there any protocols in place concerning the right to use such data for specific purposes at a later stage? • Are data subjects at any point informed on the fact that their data has been recorded during live feed, analysed, stored, or deleted?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Consideration Points	Potential ethical and Legal Issues
"Descriptive Artificial Intelligence" The system utilized advanced machine learning algorithms capable of analysing data without human interference.	To be considered if applicable.
"Behavioural Analysis and Subject Profiling as Laser Analysis, and Predictive Policing"	To be considered if applicable.
"Non-Stop Mass Surveillance"	To be considered if applicable.
"Inter-Connectivity with other Smart Systems"	To be considered if applicable.

Table 1: Potential Ethical and Legal Issues I

1.3 General Ethical Questions

The public, in general, supports the primary efforts aimed at enhancing public security. However, the data-collection technology utilized in security operations may raise concerns about privacy issues. Timely and efficient law enforcement operations may require intrusions into privacy, requiring precise and detailed analysis of the available ethical and legal guidelines and rules on how the engaged actors must handle such data processing. Ethical considerations revolve over the principal conflict of interests between collecting, analysing, and sharing big data on one side, and the need to protect personal data on the other. Noted strife is enhanced by the ever-increasing smart city capabilities of gathering big data as challenged by the civilizational strive to promote data rights as human rights. The purpose of an ethical analysis is to reconcile the two, at first glance, opposing forces, and to explore venues of co-existence, as represented in the table below.

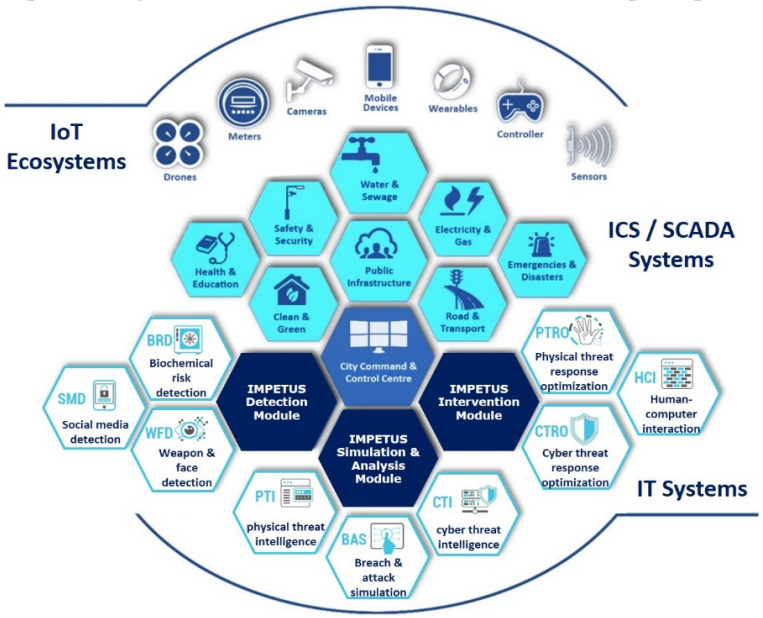


Figure 1: IMPETUS Urban Safety Modules and Tools (DoA, Part B, 2020: 7 *et seq.*)

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Conflicting Interests	
Security Interests	Privacy Interests
Necessity of Identification	Right to Anonymity
Access to Data	Control of Data and Information Symmetry
Arbitrary *post hoc* Information on Data Access	Informed Access with Consent
Storing and Use of Data for other Purposes	Right to be Forgotten
National and Public Security Interests	Limitation of Governmental Surveillance
Right of Secrecy	Right to Redress
System of Immunity and Impunity	Responsibility and Liability

Table 2: Conflicting Interests

1.4 Privacy and Personal Data in a Non-Security Environment

The following segment analyses the manipulation of personal (and non-personal) data in a non-security environment.

The all-out collection and manipulation of data are particularly pronounced in the differentiation between targeted surveillance versus mass surveillance (Cate, Dempsey (eds.), 2017). In the context of modern security challenges, it would be irresponsible to neglect and abandon the means and tools in data mining and analysis offered by the AI (EC, 2020b). At the same time, it will be necessary to understand what makes the mass and targeted surveillance justifiable and what are the concrete benefits of employing AI algorithms in such operations. This evaluation's outcome must have a fair and beneficial effect on the security and intelligence sector and society. A positive impact on society (common good principle) must outweigh all negative aspects (von Silva, Larsen, 2011). Potential threats and risks must be acknowledged and mitigated to the best extent possible (Floridi et al., 2018). The common good principle should point to a particular value of general appreciation that gives justification for reducing other values, principles, and rights (what the European Data Protection Supervisor (EDPS) referred to as the so-called big data protection ecosystem (EDPS, 2015)).

The Convention for the Protection of Individual with regard to Automatic Processing of Personal Data (APPD, 1981) with its latest Protocol from 2018, recognizes the need to reconcile the right to personal data protection with other fundamental human rights, thus placing the data rights into fundamental rights' category. The personal data implies any information pertaining to an identified or identifiable individual (Art, 1. APPD; co-opted by the General Data Protection Regulation (GDPR; GDPR, 2016) in Art. 4(1)), irrespective of its nature (WP Article 29, 2007). The term identifiable may relate to anonymized data that can be re-personalized. As a vast aray of data can be adjoined to the personal data category (including items such as the metadata, the IP address, and similar), a large section of big data collection and manipulation pertains to the noted category (Voigt, von dem Bussche, 2017:240 et seq.). The non-personal data is regulated by the Regulation (EU) 2018/1807 on the non-personal data (Regulation (EU) 2018/1807, 2018).

The following is an excerpt from D1.2:

"*The right of privacy, recognized as a universal human right by the Universal Declaration of Human Rights (Art. 12; UDHR, 1948), is, among other sources, defined by the European Convention of Human Rights (Art. 8; ECHR, 1953). The Convention (as interpreted by the European Court of Human Rights) places an obligation on the State to protect its citizens against unjustified intrusions into their private affairs. Such intrusions are only permitted if prescribed by law and required by exceptional circumstances (i.e., national*

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

security, prevention of crime, citizens' protection, and similar). Such exceptions are scrutinized by the Court of Justice of the European Union (CJEU), particularly regarding the data protection defined by the Charter of Fundamental Rights of the European Union (CFREU, 2012). CRFREU requires (Art. 8) consent for data collection or some other legitimate basis and stipulates the subject's right to access collected data and the right to rectification. In cases where such rights are to be limited or excluded (Art. 52), the limitations or exclusions must be prescribed by law, must the necessary (sufficient grounds) and proportionate (choice and severity of measures), must be foreseeable (to a certain degree), and must fulfil broader goals of general interest (in the public interest).

The APPD Protocol reinforces the afore-mentioned criteria by stipulating (Art. 14) that any data processing activities justified by national security or defence purposes must be subjected to a regulated independent review and supervision. GDPR also stipulates several legal grounds (Art. 6(1), Art. 10, and others), including the public interest and the exercise of official authority (Handbook, 2018). Art. 23 GDPR details several reasons for a valid exclusion or limitation of subject's rights, including national security, defence, public security, criminal proceedings, critical public interests, and others. Similarly, the Directive 2002/58/EC on privacy and electronic communications (Directive on privacy, 2009) completely sets its application aside regarding the matters of state security, public security, defence, and criminal law.

Based on Art. 23 GDPR (and other similar rules and other legal documents), national legislation usually adopts separate acts and statues concerning the security and intelligence operations, criminal proceedings, public security issues, national security issues, and others (European Data Protection Board (EDPB); EDPB, 2020).

Therefore, the regulated aspects of personal data and right to privacy may or may not be relevant for each jurisdiction, as most relevant rights are either limited, restricted, or altogether excluded (AccessNow, 2019). Indeed, most European Union (EU) Member States and other states will have enacted specialized laws and status concerning the operation of their security apparatus, police force, data secrecy, and similar. In principle, the matters of national security remain under the purview of states (Member States in the EU context; as per Art. 4(2) of the Treaty on the Functioning of the European Union (TFEU, 2012))."

1.5 Stakeholders

In the modern-day environment, various actors can get involved in security and intelligence data collection and manipulation. The primary layer consists of public security agencies and institutions (police department, security apparatus, fire department, supranational bodies, and others.). The second layer consists of relevant public administration bodies (transportation authority, various city offices, and others). The third layer consists of private entities contractually or non-contractually engaged directly or indirectly in security operations. The fourth layer represents ordinary citizens and legal persons involved either in autonomous or automatic data sharing/gathering.

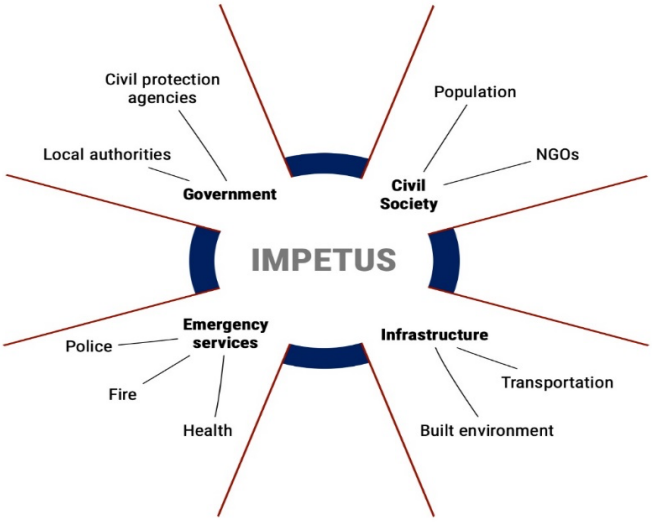


Figure 2: IMPETUS Stakeholders (DoA, Part B, 2020: 7 *et seq.*)

Whereas the security services primarily use anonymous data to identify surveillance targets, private companies use anonymous data (Artificial Intelligence Committee, 2017) for commercial purposes. Individuals whose data is collected and analysed become objects rather than users. Such data may nor may not be re-identified. Questions are raised on whether private entities engaged in security and intelligence data collection and manipulation operations can avail of such data for other purposes.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Stakeholder	Role	Examples
Regulators	Regulate the implementation of security technology in public places	• Policy makers at the city, national and European level
Decision-makers	Decide on investment in security technology solutions	• City managers, smart city managers • Regional & national security agency managers • Critical infrastructure managers
Security actors	Use the technologies in security operations	• Police and other security agencies, contractors • City, regional, national level • Operators and managers
Emergency actors	Interact with primary users in managing security events	• Emergency agencies' actors and managers • Critical infrastructure operators
Population	Residents of the smart cities, impacted by the use of the technology	• Citizens (including citizen groups)

Table 3: Stakeholders of Public Safety Solutions (D1.2)

1.6 References

- AccessNow. One Year Under the EU GDPR, An Implementation Progress Report: State of play, analysis, and recommendations. AccessNow.org, 2019
- Article 29 Working Party, Opinion 4/2007 on the concept of personal data. 01248/07/EN WP 136.
- Artificial Intelligence Committee, AI in the UK: ready, willing and able? Report of Session 2017-19 - published 16 April 2017 - HL Paper 100
- Asilomar AI Principles (2017). Principles developed in conjunction with the 2017 Asilomar conference.
- Association for Computing Machinery (2018). ACM Code of Ethics and Professional Conduct: Affirming our obligation to use our skills to benefit society.
- Bundesministerium des Innern, für Bau und Heimat, Bundesministerium der Justiz und für Verbraucherschutz (2018). The Federal Governments key questions to the Data Ethics Commission. 5 June 2018
- Cate, F.H., Dempsey, J.X. (eds.) (2017). Bulk Collection: Systematic Government Access to Private-Sector Data. Oxford: Oxford University Press
- Charter of Fundamental Rights of the European Union, OJ C 326, 26.10.2012, p. 391–407.
- Clever, S., Crago, T., Polka, A., Al-Jaroodi, J., Mohamed, N. (2018). Ethical Analyses of Smart City Applications. *Urban Sci*. 2018, 2, 96
- Coeckelbergh, M. (2020). AI Ethics. Cambridge: The MIT Press.
- Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

(CETS No. 108).

- Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Protocol 2018 (CETS No. 223).
- Corea, F. (2019). An Introduction to Data: Everything You Need to Know about AI, Big Data and Data Sciences. Cham: Springer Nature.
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector, OJ L 201, 31.7.2002, p.37, amended by: Directive 2006/EC of the European Parliament and of the Council of 15 March 2006, OJ L 105, p. 54, and, Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009, OJ L 337, p. 337.
- Van Eck, G.J.R. (2018). Emergency calls with a photo attached: The effects of urging citizens to use their smartphones for surveillance. In: Newell, B.C., Timan, T., Koops, B.-J. (eds) (2018) Surveillance, Privacy and Public Space. Routledge Publishing, 2018
- Ethics Advisory Group 2018 Report, Towards a digital ethics, available at: https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf (12th January 2021)
- European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. The European Agenda on Security. COM(2015) 185 final
- European Commission, 2020b. White Paper: On Artificial Intelligence – a European approach to excellence and trust, COM(2020) 65 final
- European Convention of Human Rights, Council of Europe, 1953
- European Data Protection Board (2020). Statement on restrictions on data subject rights in connection to the state of emergency in Member States. Adopted on 2 June 2002.
- European Data Protection Supervisor, Guidelines 10/2020 on restrictions under Article 23 GDPR, Version 1.0. 15 November 2020
- European Data Protection Supervisor, Opinion 4/2015, Towards a new digital ethics Data, dignity and technology, 2015, available at: https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf (12th January 2021)
- European Union Agency for Fundamental Rights, Council of Europe (2018). Handbook on European data protection law, 2018 edition. Luxembourg: Publications Office of the European Union.
- Feldstein, S. (2019). The Global Expansion of AI Surveillance. Washington: Carnegie Endowment for International Peace
- Ferguson, A. G. (2017). The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. New York: New York University Press.
- Floridi, L. *et al.*, AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines (2018) 28:689–707.
- G20 (2019). Ministerial Statement on Trade and Digital Economy. G20 Osaka Summit, G20 Trade Meetings
- Menzer, S., Rubba, C., Meißner, P., Nyhuis, D. (2015). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. West Sussex: John Wiley & Sons, Ltd
- Milaj, J., van Eck, G.J.R. (2019). Capturing license plates: police-citizen interaction apps from an EU data protection perspective. International Review of Law, Computers and Technology, 25 March 2019
- Office of Homeland Security & Emergency Preparedness, City of New Orleans (New Orleans), available at: https://www.nola.gov/homeland-security/real-time-crime-center/ (12th January 2021)

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

- 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 (GDPR), and repealing Directive 95/46/EC, OJ L 119, 4.5.2016
- Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ L 303, 28.11.2018, p. 59–68
- von Silva, B., Larsen, T. (2011). Setting the Watch: Privacy and the Ethics of CCTV Surveillance. Portland: Hart Publishing.
- Timmermans, H. (ed.) (2009). Pedestrian Behavior: Models, Data Collection and Applications. Bingley: Emerald Group Publishing Limited
- Treaty on the Functioning of the European Union, OJ C 326, 26.10.2012
- United Nations Organization, Universal Declaration of Human Rights, General Assembly resolution 217 A
- United Nations Organization, High-Level Committee on Management. Personal Data Protection and Privacy Principles. 36th Meeting, October 2018
- Vogiatzaki, M., Zerefos, S., Tania, M.H. (2020). Enhancing City Sustainability through Smart Technologies: A Framework for Automatic Pre-Emptive Action to Promote Safety and Security Using Lighting and ICT-Based Surveillance. *Sustainability* 2020, 12, 6142.
- Voigt, Paul, von dem Bussche, Axel. The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer. Cham: International Publishing, 2017
- Zwitter, A. (2014). Big Data Ethics. Big Data & Society. July-December 2014; 1-6

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 11: Legal Analysis

Status: Initial Draft

Overview of the key legislation relevant for personal data manipulation in security operations.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Tables

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Legal Analysis

1.1 Introduction

The following segment analyses the manipulation of personal data in security operations. This report will analyse the relevant body of EU law relevant for security operations' data collection, analysis, sharing, storing and similar. Unlike the GDPR that is relevant for personal data protection in non-security operations, all personal data utilized for security (and related) purposes must adhere to a different set of individual norms. The legal scrutiny of AI-powered systems (AI regulation) is only beginning to emerge (as noted by Asilomar AI principles; Asilomar, 2017). Introducing AI systems into everyday operations will likely require an overhaul of existing legal principles and norms (as indicated in United Kingdom (UK) Parliament Artificial Intelligence Committee, Chapter 8; Artificial Intelligence Committee, 2017; EC, 2020a). The summarized legal analysis will include all the legislation previously mentioned in this section and several additional legal instruments, as enumerated in the following excerpt from D1.2. The full analysis will be completed following the IMPETUS Platform trials.

"*The Directive (EU) 2016/680 (Police Directive, 2016), which regulates the (partially) automated processing of natural persons' personal data in criminal investigations and sanctions, public security operations (safeguarding and prevention of crime) law enforcement purposes by competent authorities (public security authorities, other public authorities, as well as private entities statutory entrusted with security operations). The Police Directive follows the GDPR logic (both are a part of the same legislative package) when determining the relevant definitions of terms such as personal data, data processing, pseudonymization, controller, and processor (as discussed above). The Directive establishes several relevant principles, such as lawful and fair processing, legitimate grounds (legality), specific and explicit purpose (necessity), adequate and not excessive (proportionality), accurate and up-to-date, secured and purpose restricted.*

This Directive is relevant for the IMPETUS Project's context, especially considering the security and intelligence data collection and manipulation activities. It stipulates rules concerning automated individual decision-making, rights of the data subjects (general, information, access, and others), obligations of controllers and processors, data protection officers, supervision, data security, access to data, and data transfer, and others. As is the case with the GDPR, the Police Directive also allows for limitations and exclusions in cases of criminal proceedings, national security, and others, and keeping in mind that this is a directive (unlike the GDPR), the changes of national law variety in regulating these discrepancies and exclusion is increased. The Police Directive includes data processing operations aimed at preventing threats to public security. Additionally, it promotes cross-border cooperation (exchange of data) and establishes a compensation scheme for breach cases (with the general requirements informing the public and individuals regarding data processing).

It should be noted that, depending on a case-to-case evaluation, both the GDPR and Police Directive can be relevant for each security data gathering, processing, and handling operation. The applicable norm depends on who the relevant actors are, how they are involved in the process, and what sort of information is being collected. The noted applicability criteria are of particular importance for private actors who participate in security operations. Private actors could be providing specific services, such as online cloud storage or secure communications, hardware solutions, software solutions, support solutions, and others. Alternatively, private actors could be engaged through public-private partnerships, including a delegation of certain powers and duties to private entities (Purtova, 2018). The noted services must be aligned with general GDPR requirements and, if applicable, particular Police Directive specifications. Also, such activities are likely to be further regulated by other relevant legislation, such as the rules on technical compatibility and security (i.e., Regulation (EU) 2018/389 with regard to regulatory technical standards for strong customer authentication and common and secure open standards of communication). The IMPETUS will consider the legal framework established by the Regulation (EU) 2018/1725 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, especially in the extent that the Regulation creates rules additional to those already established by the Police Directive and are of interest when relevant EU bodies (i.e., Eurojust)

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

conducting law enforcement operations are to be, potentially, adjoined to the IMPETUS platform operations.

It must be noted that the highlighted legal sources are based on the general requirements and policy set by TFEU (namely Chapter 4 and Chapter 5, Title V, Part Three, TFEU), requiring judicial and police cooperation in law enforcement activities and operations. Furthermore, the European Agenda of Security (Security Agenda, 2015), among other items, highlights the use of several additional legal instruments, including the Directive (EU) 2018/843 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing (Directive (EU) 2018/843, 2018), the Directive (EU) 2019/1153 laying down rules facilitating the use of financial and other information for the prevention, detection, investigation or prosecution of certain criminal offences (Directive (EU) 2019/1153, 2019), and, Regulation (EU) 2015/847 on information accompanying transfers of funds (Regulation (EU) 2015/847, 2015), allowing the Financial Intelligence Units and other competent public authorities access to sensitive information.

It will be useful to analyse the possible ramifications of the Proposal for a Regulation concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (providing clear rules with regard the persons' privacy (i.e., traffic data, location data, secure communication service, and similar; Proposal on Regulation on Privacy and Electronic Communications, 2017), allowing so-called backdoor access by law enforcement and affiliated agencies to individuals' data). Similarly, the Proposal for a Regulation on European Production and Preservation Orders for electronic evidence in criminal matters (Proposal on Regulation on Electronic Evidence in Criminal Matters, 2018), as well as the Proposal for a Directive laying down harmonised rules on the appointment of legal representatives for the purpose of gathering evidence in criminal proceedings (Proposal for Directive on Legal Representatives in Evidence Gathering, 2018), allow for more direct means available to law enforcement and affiliated agencies to connect to individuals' data (Boehm, 2012). Depending on the relevant stakeholders, the ethical analysis may include further legal analysis of instruments such the Directive 2016/1148 on network and information systems (Directive (EU) 2016/1148, 2016), the Regulation (EU) 2019/881 on cybersecurity (Regulation (EU) 2019/881, 201)."

The table below provides a brief legal survey of Italian and Norwegian law (as per pilot cities).

Brief Legal Survey	
Italian Law; answered by CPAD	
Briefly describe the surveillance legal framework (relevant acts and statutes, relevant case practice).	
On a European level: GDPR 679/2016; European Directive for police forces 680/2016 On a national level: Personal Data Protection Code d.lgs 196/2003, then integrated with d. lgs 101/2018 to adapt the national legislation to the GDPR, D.lgs 51/2018 to adapt the national law to the 280/2016 Directive On a local level: Municipal Regulation Video Surveillance Sysyem "Padova Città Sicura"(Padova Safe City) since 2008 (now about to be updated). In general we can say that our national and local legislation is totally compliant with the GDPR and 280/2016 directive. *d.lgs can be translated as national law, even though it indicates a specific approval procedure.	On a European level: GDPR 679/2016 On a national level: The Personal Data Act (Personopplysningsloven) and The Personal Data Regulations (Personopplysningsforskriften). The Working Environment Act (Arbeidsmiljøloven § 9-6 Camera surveillance) The national legislation is in general in compliance with GDPR.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Brief Legal Survey	
Which public stakeholders (ie, security agencies, intelligence agencies, and similar) are entitled to utilize surveillance? Are such stakeholders allowed to share data and information between each other?	
All the police forces (Local Police, National Police, Carabinieri Corp, Finance Guard and other police corps) are allowed to utilize surveillance materials and they can share information between each other, even though there is no real need to share as they have direct access to it.	Potential and criminal offences can be shared with the police, and the Police may require information through investigation. It is not allowed to share surveillance between different stakeholders.
When and to what purpose are public stakeholders allowed to utilize surveillance? To what extent is such data usable in criminal investigations and court proceedings?	
Normally the cameras keep the images for 7 days, and then they automatically erased them. In case of felony or damage, or on request of citizens, the police forces are allowed to watch them and, if they show a crime, the images can be kept as proof.	All stakeholders may utilize surveillance within the legal frame; there must be a necessity for the surveillance. Legally deployed surveillance may be used in court proceedings. Both public and private stakeholders must delete surveillance data within 7 days, except approved application for storage of surveillance of 30 days under the criminal act.
Can public stakeholders utilize private means of surveillance (ie, access to the data collected by privately-owned CCTVs)? If so, how is this regulated, and to what extent is such data usable in criminal investigations and court proceedings?	
Yes, the police forces can ask private citizens to utilize the images of their private surveillance, and citizens can' t refuse. The images achieved are usable in investigations and trials.	Yes, Same as answer 3.
Which private stakeholders are allowed to utilize surveillance, what sort of surveillance, and what are the rules concerning the data so acquired?	
All private citizens can use video surveillance in their property and the area nearby, but there must be a clear indication of it. On the contrary private citizens can't have access to public surveillance videos, they can only report a crime and let the police watch the videos.	Yes, Same as answer 3.
Briefly describe the data protection and data privacy legal framework (relevant acts and statutes, relevant case practice).	

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Brief Legal Survey	
On a European level: GDPR 679/2016, European Directive for police forces 680/2016 On a national level: Personal Data Protection Code d.lgs 196/2003, then integrated with d. lgs 101/2018 to adapt the national legislation to the GDPR, D.lgs 51/2018 to adapt the national law to the 280/2016 Directive As previously mentioned, our legislation is compliant with the European one, so practice cases are always based on the rule of the explicit consent. Otherwise even police can't have access to private data, with the only exception of security issues.	Same as answer 1.
Is the collection and/or analysis of publicly available data (World Wide Web) regulated in your jurisdiction?	
Yes, by the GDPR. In general there is the rule of the explicit consent.	Processing of data will be subject to data protection act if it contains into personal information or can reveal identifiable information through aggregation. The processing of Big Data may trigger obligations and rights pursuant to the Norwegian Personal Data Act. The prerequisite here is that the processing concern personal data. The Act namely applies to the "processing of personal data wholly or partly by automatic means", pursuant to Section 3 of the Act. That the processing of Big Data takes place by automatic means lies in the nature of such data (Datatilsynet, 2013).

Table 1: Brief Legal Survey

Finally, the legal analysis will consider the *de lege ferenda* ramifications of the current Proposal for a Regulation laying down harmonised rules on artificial intelligence (COM(2021) 206 final), through which the Commission plans to harmonise rules on artificial intelligence.

1.2 References

- Artificial Intelligence Committee, AI in the UK: ready, willing and able? Report of Session 2017-19 - published 16 April 2017 - HL Paper 100
- Asilomar AI Principles (2017). Principles developed in conjunction with the 2017 Asilomar conference.
- Boehm, F. (2012). Information Sharing and Data Protection in the Area of Freedom, Security and Justice: Towards Harmonised Data Protection Principles for Information Exchange at EU-level. Berlin: Springer-Verlag
- Commission Delegated Regulation (EU) 2018/389 of 27 November 2017 supplementing Directive

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

(EU) 2015/2366 of the European Parliament and of the Council with regard to regulatory technical standards for strong customer authentication and common and secure open standards of communication, OJ L 69, 13.3.2018, p. 23–43

- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector, OJ L 201, 31.7.2002, p.37, amended by: Directive 2006/EC of the European Parliament and of the Council of 15 March 2006, OJ L 105, p. 54, and, Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009, OJ L 337, p. 337.
- Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119 4.5.2016, p. 89
- Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, OJ L 194, 19.7.2016, p. 1–30
- Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU, OJ L 156, 19.6.2018, p. 43–7
- Directive (EU) 2019/1153 of the European Parliament and of the Council of 20 June 2019 laying down rules facilitating the use of financial and other information for the prevention, detection, investigation or prosecution of certain criminal offences, and repealing Council Decision 2000/642/JHA, OJ L 186, 11.7.2019, p. 122–137
- Ethics Advisory Group 2018 Report, Towards a digital ethics, available at: https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf (12th January 2021)
- European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. The European Agenda on Security. COM(2015) 185 final
- Evas, Tatjana. European framework on ethical aspects of artificial intelligence, robotics and related technologies: European added value assessment. European Parliament: European Parliamentary Research Service, PE 654.179, 2020
- OECD (2020). The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector, available at: www.oecd.org/finance/Impact-Big-Data-AI-in-the-Insurance-Sector.htm
- Proposal for a Directive of the European Parliament and of the Council laying down harmonised rules on the appointment of legal representatives for the purpose of gathering evidence in criminal proceedings, COM/2018/226 final - 2018/0107 (COD)
- Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), COM/2017/010 final - 2017/03 (COD)
- Purtova, N. (2018). Between GDPR and the Police Directive: Navigating through the Maze of Information Sharing in Public-Private Partnerships. International Data Privacy Law (2018)
- Proposal for a Regulation of the European Parliament and of the Council on European Production and Preservation Orders for electronic evidence in criminal matters, COM/2018/225 final - 2018/0108 (COD)
- Regulation (EU) 2015/847 of the European Parliament and of the Council of 20 May 2015 on

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

information accompanying transfers of funds and repealing Regulation (EC) No 1781/2006, OJ L 141, 5.6.2015, p. 1–18

- Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ L 295, 21.11.2018, p. 39–98
- 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), OJ L 151, 7.6.2019, p. 15–69
- Treaty on the Functioning of the European Union, OJ C 326, 26.10.2012
- Young, M., Katell, M., Krafft, P.M. (2019). Municipal surveillance regulation and algorithmic accountability. Big Data & Society, July-December 2019: 1-14.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Grant number: 883286
Project duration: Sep 2020 – Aug 2022
Project Coordinator: Joe Gorman, SINTEF

Horizon 2020: Secure societies
SU-INFRA02-2019
Security for smart and safe cities, including for public spaces
Project Type: Innovation Action

http://www.impetus-project.eu

IMPETUS Project Deliverable: D5.1 Initial Ethical Framework

Part 12: Relevant Ethical Guidelines and Principles

Status: Mature Draft

Citizens' Guide on how the EU Guidelines (and other similar guidelines) are relevant for IMPETUS Platform.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Tables

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1. Relevant Ethical Guidelines and Principles

The IMPETUS Project has determined that the development of an ethical framework must follow the Ethics Guidelines for Trustworthy AI (EGTAI, 2019), prepared by the High-Level Expert Group on Artificial Intelligence. The EU promulgates the EGTAI principles as the centre-piece of its AI Ethics strategy (Coordinated Plan, 2018). The EGTAI determines that a notion of a trustworthy artificial intelligence (AI) implies several components. The AI system must be lawful (compliance with regulations), ethical (adherence to values and principals) and robust (socio-technical adaptability). The AI system must also fulfil several specific requirements (human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability). Other similar standards and guidelines will be taken into consideration when completing this report after the IMPETUS trials.

1.1 EU Ethics Guidelines for Trustworthy AI

The following is an excerpt from D1.2.

"*EGTAI states that ethical principles and values must be ever-present in all AI development and utilization stages. EGTAI highlights four fundamental principles: respect for human autonomy, prevention of harm, fairness, and explicability. The noted fundamental principles must be assessed in the IMPETUS Project's context. The following considerations are relevant for further analysis.*

The respect for the human autonomy principle is particularly emphasized through the need to secure human oversight over the AI system's operations. The prevention of harm principle relates to the secure and safe AI system's functioning (elimination of malicious use or exploitation (Brundage, 2018)). Additionally, it underlines the importance of preventing negative impacts resulting from the asymmetry of power and asymmetry of information. The latter is especially relevant in connection to vulnerable groups (inclusivity principle).

The fairness principle focuses on eliminating bias and stigmatization (prevention of unlawful and unfair use of data), stipulating the necessity of formulating clear procedural rules for human oversight of AI operations. The fairness principle additionally requires the proportional application of measures.

The proportionality principle is reflected through the practice of undertaking limited measures to the extent necessary to achieve specific goals (scope limitation). The choice of measures undertaken should be guided by the desire to protect fundamental rights and ethical values (qualitative/severity measures' limitation). It also addresses people's potential right to redress the consequences of AI and human operators' activities. The latter is connected with the principle of explicability that emphasizes the need for transparency of AI means and operations. EGTAI urges caution when employing AI systems, as such practices may have unwanted and unexpected consequences for society (i.e., the effect on distributive justice).

The noted principles must be further examined in the IMPETUS Project context, emphasizing the public-to-citizen relationship and security and intelligence operations' sensitivity and confidentiality. The ethical analysis should discern to what extent are the noted principles applicable in the IMPETUS Project context. EGTAI itself gives an example of "predictive policing", stating that the "… surveillance activities [may] impinge on individual liberty and privacy" (EGTAI, 2019:13), thus emphasizing a possible overlap or conflict between the relevant principles. EGTAI warns that an ethical framework cannot be constructed based on the principles alone but rather through an evidence-based reflection. The latter corresponds to the various data collection activities and pilot projects (to be) conducted through the IMPETUS Project.

As per EGTAI, fundamental human rights are an inseparable component of a trustworthy and lawful AI. Therefore, when assessing big data, the AI system must incorporate safeguards to protect personal dignity and identity, considering people whose data is being captured and manipulated, not merely as objects but subjects with individual rights. EGTAI states that whenever there is a risk of an adverse effect on fundamental rights, each AI system operation must be preceded by a fundamental rights' impact assessment. In addition, EGTAI further notes that whenever an AI system affects fundamental rights, there ought to be an independent

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

audit system in place (external auditors). The freedom of the individual incorporates, among others, the right to be protected from unjustified surveillance, the right to private life, and the right to data privacy. EGTAI warns that mass or targeted surveillance, when conducted without a legitimate purpose and justification and by using disproportionate means, can lead to fundamental rights endangerment.

EGTAI stipulates that certain fundamental rights cannot be subjected to a trade-off and remain within the ethically acceptable norms. In cases where such measures are not ethically acceptable, the data collection and manipulation should not proceed. Where data collection and manipulation activities potentially infringe on private interests, such action must be identified, acknowledged, evaluated, and justified. All data collection and manipulation activities should be documented, and procedural and ethical rules on data collection and manipulation admissibility should be continuously reviewed and evaluated.

Whereas EGTAI insists on dispute settlement mechanisms available to the third-party stakeholders (not only the compensation but all other means of rectification), it remains questionable and has to be further examined to what extent can such a mechanism be implemented in the IMPETUS Project's context. EGTAI suggests that public sector data is preferred to personal data. The latter indicates the extent to which the proportionality principle should be assessed in the IMPETUS Project context, having in mind both the available public and private technological means of collecting data.

Additionally, the level of human intervention in AI operations in the IMPETUS Project context must be further examined. Given the sensitive and often confidential security and intelligence operations, a human operator's role in the AI governance mechanism can be steered towards critical procedural instances when relevant decisions are made. Keeping in mind the scope and quantity of big data, the human-in-the-loop (HITL) approach, as EGTAI suggests, is unrealistic. It would be logistically impossible to have human operators remain responsible for each decision to collect and manipulate data. The human-on-the-loop (HOTL) suggests a human operator's access during the AI system design and monitoring capacity. Keeping in mind the nature of data collection and manipulation in the IMPETUS Project's context, the human-in-command (HIC) approach is likely the most suitable. The HIC level of interaction allows the human operator's control over the AI system, including the choice of overriding or influencing (human discretion) a particular decision made by the AI, full access, and oversight. Keeping in mind the scope of collected data and analytical processes, it is inevitable that a large volume of data analysis will remain under AI's purview, resorting to black-box post-hoc investigation in cases of unwanted occurrences (Leslie, 2019).

In the IMPETUS Project's context, it is vital to assess what sort of measures should be left for the human operator to decide upon, instigate or suspend. The latter is closely connected to the explainability principle, requiring that the human operator can understand and explain (in plain language) the AI system's and human operator's decision-making process and understand to what extent the AI algorithm influences both decision-making processes. Whereas the former is a technical issue, the latter deserves further evaluation during the IMPETUS Project."

1.2 Other Relevant Guidelines

Other HORIZON 2020 programme and similar projects dealing with the ethics of data collection and manipulation have adopted different standards and followed different guidelines. For example, the Safe-DEED project has adopted the principles of autonomy, justice, beneficence, non-maleficence, and responsibility (Safe-DEED, 2020). The European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems (published by the European Commission for the Efficiency of Justice (CEPEJ); CEPEJ, 2018) refers to respect for fundamental rights, non-discrimination, quality and security, fairness and transparency, and user control. The Alan Turing Institute has developed the FAST Track Principles set: fairness, accountability, sustainability, and transparency (Leslie, 2019). The noted and other principles (see: HECTOS, 2015; CUTLER, 2018; SHERPA, 2019; WITDOM, 2017; IEEE, 2019; Montreal, 2018; UN, 2018; OECD, 2019; G20, 2019; ACM, 2018; Berkman Klein, 2020; Also see: (Lorenz, 2020)) can, to a certain extent, be compared with that utilized by EGTAI. The full analysis will be conducted after the completion of IMPETUS trials.

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1.3 Relevant Ethical Requirements

The list of requirements was established in D1.2 and refers potential ethical issues relevant to the IMPETUS Project's context. Some issues refer to requirements that only need to be confirmed as fulfilled (i.e., various technical standards, procedures, and similar). Others issues problems require ethical evaluation and analysis. The majority of the indicated points are already a part of multiple deliverables through the IMPETUS Project.

1.3.1 List of Stakeholders

The ethical analysis shall include a list of stakeholders. The list shall consist of all parties involved in the IMPETUS Project and endeavour to have a broader list of interested stakeholders. The list of stakeholders must appropriately categorize and classify all stakeholders. The guiding questions to be considered are as follows:

List of Stakeholders
Who are the affected stakeholders? Can all of them be categorized into one of the following (EGTAI) categories: developers (research, design, development of AI system), deployers (public or private organizations that utilize AI system for themselves or as a service to third persons), end-users (engaged with the AI system), society at large (all third parties affected by AI system)?
Are there relevant stakeholders on the side of developers? If so, are they public or private entities? What kind of an AI system are they developing? Do they already have a legal and ethical framework in place concerning the collection and manipulation of data?
Are there relevant stakeholders on the side of deployers? If so, are they public or private entities? What kind of an AI system are they utilizing, and to what purpose? Do they already have a legal and ethical framework in place concerning the collection and manipulation of data? Do they solely rely on the AI system (full automatization), or do they insist on human operators/oversight/control?
Are there relevant stakeholders on the side of end-users? If so, are they public or private entities? What kind of an AI system are they utilizing, and to what purpose?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

List of Stakeholders
Are there relevant stakeholders on the side of third parties? If so, are they public or private? How are they affected? Are they aware that they are affected? To what extent are they affected?

Table 1: List of Stakeholders

1.3.2 Fundamental Rights' Impact Assessment

The ethical analysis shall assess the likelihood of risks to fundamental data rights and means to prevent such occurrences. The guiding questions to be considered are as follows:

Fundamental Rights' Impact Assessment
What sort of data can be collected (visual data, auditory data, biometric data, genetic data, documentary data, ethnical data, racial data, social data, religious data, health data, private data, and other data)? What sort of data is relevant to the IMPETUS Project's context?
Can data rights be considered as fundamental (human) rights?
What is considered a fundamental data right concerning collecting and manipulating personal data in the IMPETUS Project context? To what extent are the GDPR and similar legislation enabled rights enforceable in the IMPETUS Project context?
Does the concept of fundamental data rights recognize the difference between data collected from publicly available sources and private sources?
What are the possible risks to fundamental data rights?
Are such risks acceptable, and can a limitation or exclusion of fundamental data rights be justified by higher goals/principles?
What are the implications of the fact that anonymized (de-identified) personal data can get re-personalized (re-identified)?
Does the technology in question allow for pseudonymization technics leading to untraceable sets of data?
Are there certain fundamental data rights that preclude the collection and manipulation of personal data?
Are there any boundaries to what extent private data can be collected and manipulated?
To what extent is collecting and manipulating data for commercial purposes different from collecting and manipulating data for security and intelligence purposes?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Fundamental Rights' Impact Assessment
Should human operators control private data collection and manipulation, or can specific activities be fully automated? Does the moment when anonymized data needs to be re-personalized constitute a valid junction for a human operator to take over the analysis from the AI system? Does the AI system decide which anonymized data needs to be re-personalized or simply make recommendations to that effect? Should the AI system continue with independent analysis and decision-making after the data has been re-personalized?
Are the third-party stakeholders (citizens and private entities whose data is collected and manipulated) entitled to be informed on such activities? If so, at what time and to what extent is this information to be released? Are there any plausible exclusions to such a rule? Should a human operator be in charge of informing the third-party stakeholders, or can this procedure be delegated to the AI system? Does it make a difference whether the data collected on a particular individual was relevant for the conducted investigation/surveillance?
If the collection and manipulation of data have led to a particular decision legally affecting individuals, can such decisions be based on automated processing? Or must the human operator's consideration precede such a decision?
To what extent should private stakeholders be involved in data collection and manipulation in security and intelligence operations? Should the private entities be allowed to act as processors on behalf of security and intelligence agencies acting as controllers?
Is there any audit or external feedback mechanism in place, and is there an oversight system in place?

Table 2: Fundamental Rights' Impact Assessment

1.3.3 Deployer's Obligations

The ethical analysis shall thoroughly examine the ethical issues connected to the deployers category. The guiding questions to be considered are as follows:

Deployer's Obligations
To what extent should the AI system's operations be directed, controlled, or supervised by human operators?
Accordingly, what category of the AI system's governance mechanism should be employed?
Does the deployer have an established audit mechanism? Is the audit handled internally, or is there an established external audit body?

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Deployer's Obligations
Does the deployer have an established supervision mechanism? Is the supervision handled internally, or is there an established external audit body?
Does the deployer have an established oversight mechanism? Is the oversight handled internally, or is there an established external oversight body?
Is the deployer's operation regulated, or does the issue require further regulation?
Has the deployer developed a data storage protocol, data access protocol, AI system's algorithm integrity and reliability protocol, AI system's algorithm decision-making protocol, human operator's decision-making protocol, data quality and integrity protocol, data processing protocol, data sets and processes traceability protocol, and other similar protocols and standard (code) of conduct specifications?
In line with the previous point, does the AI system's algorithm have clear rules on dividing surveillance-related relevant from irrelevant data? Are the two sets stored jointly or separately? Can a human operator access both sets or just the relevant data set? Should the human operator be able to access both sets? Should the irrelevant data set be stored for a specified amount of time, or should it be deleted immediately after classification?
In line with the previous points, does the AI algorithm have clear rules on pseudonymization and the creation of traceable and untraceable data sets? Are the two sets stored jointly or separately? Can a human operator access both sets or just the traceable data set? Should the untraceable data set be stored for a specified amount of time, or should it be deleted immediately after classification?
Should the AI system be deciding what data goes into traceable and what data goes into the untraceable category? Should all data not falling under the red-flag alert system automatically be pseudonymized and directed to the untraceable data set (green-flagged, cleared data)? Should a human operator nevertheless have the option to reverse the course and move data freely from one category to another?
Has the deployer developed an AI system's negative impacts' management system (identification, assessment, documentation, and minimization)?
To what extent does the AI system's algorithm influence the human operator decision-making process?

Table 3: Deployer's Obligations

1.3.4 General Ethical and Legal Issues

The ethical analysis shall thoroughly examine the broader ethical issues arising out of the IMPETUS Project's context. Such topics include but are not limited to the following questions:

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

General Ethical and Legal Issues
To what extent does the use of big data in security and intelligence operations widen the asymmetry of information and power between the public security departments and agencies and the general public?
To what extent does the use of big data in security and intelligence operations belittle the concept of data rights as fundamental human rights?
To what extent do the security exigencies justify private data infringement, access to personal information, use of private collection mechanisms (i.e., mobile phones, private CCTVs, and similar)?
To what extent does the security clearance negate third-party stakeholders' right to be informed over legal or illegal transgressions into their private domain?
To what extent should security and intelligence operations concerning data collection and manipulation be regulated?
To what extent should security and intelligence operations concerning data collection and manipulation be supervised and oversight? Should the oversight and supervision bodies be internal or external, or both? What should be the composition of such bodies?
To what extent should the redress right (right to claim for damage compensation) be warranted where harm has resulted from unlawful/unfair/unethical collection and data manipulation during the security and intelligence operation? To what extent should the redress right be warranted where harm has resulted from the AI system's malicious use during the security and intelligence operation? To what extent should the redress right be warranted where harm has resulted from the AI system's malfunction during the security and intelligence operation?
Concerning the previous point, who should be the responsible party (i.e., the agency conducting data collection and manipulation, the agency providing hardware/software, the agency in charge of the overall investigation, ministry of interior, state)? Does the existence of such a right require the presence of a mandatory liability insurance policy?
Should the security or intelligence agency collecting and manipulating data have experts specialized in ethical issues, or should each operator be trained in ethics?
For how long should the anonymized and re-personalized data be kept, can it be shared with other agencies, and can it be used for purposes other than the initial investigation?
Should consideration be made concerning the restrictions imposed on public bodies when collecting and manipulating data be equally applied to the private sector?
Should each public security department and agency run separate real-time security and intelligence data collection and manipulation centre, or should an emphasis be placed on joint centre(s)? What is the benefit of singular, and what are the benefits of multiple centres?

Table 4: General Ethical and Legal Issues

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

1.1 References

- Association for Computing Machinery (2018). ACM Code of Ethics and Professional Conduct: Affirming our obligation to use our skills to benefit society.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A.C., Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI (Berkman Klein). Berkman Klein Center for Internet & Society at Harvard University. Research Publication No. 2020-1
- Brundage, M. and others (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation. Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, OpenAI.
- Coastal Urban Development through the Lenses of Resiliency (CUTLER) (2018). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770469
- Empowering privacy and security in Non-Trusted Environments (WITDOM) (2017). This project has received funding from the European Union's Horizon 2020 research and innovation programme (H2020-ICT-2014-1) under grant agreement No. 64437.
- European Commission. Communication from the Commission to the European Parliament, the European Council and the Council, the European Economic and Social Committee and the Committee of the Regions, Coordinated Plan on Artificial Intelligence, COM(2018) 795 final
- European Commission. Independent High-Level Expert Group on Artificial Intelligence, European Commission. Ethics Guidelines for Trustworthy AI. Brussels, 2019
- European Commission for the Efficiency of Justice, European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. 31st CEPEJ plenary meeting, Strasbourg, 2018.
- G20 (2019). Ministerial Statement on Trade and Digital Economy. G20 Osaka Summit, G20 Trade Meetings
- Harmonized Evaluation, Certification and Testing of Security products (HECTOS). Project funded by the European Community's Seventh Framework ProgrammeFP7/2007-2013 under Grant Agreement No 606861, 2015
- Institute of Electrical and Electronics Engineers (IEEE) (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. First Edition.
- Leslie, D. (2019). Understanding artificial intelligence ethics safety: A guide for the responsible design and implementation systems in the public sector. The Alan Turing Institute.
- Lorenz, P. (2020). AI Governance through Political Fora and Standards Developing Organizations: Mapping the actors relevant to AI governance. Berlin: Stiftung Neue Verantwortung
- Montreal Declaration for a Responsible Development of Artificial Intelligence (2018). Announced at the conclusion of the Forum on the Socially Responsible Development of AI.
- OECD (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449
- Safe Data-Enabled Economic Development Horizon 2020 research and innovation programme (Safe-DEED), Grant Agreement No. 825225
- Shaping the ethical dimensions of smart information systems (SIS) – a European perspective (SHERPA) (2018). his project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641
- United Nations Organization, High-Level Committee on Management. Personal Data Protection and

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.

Privacy Principles. 36th Meeting, October 2018

The research leading to these results has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation (H2020) under grant agreement n° 883286.