D9.8 Updated framework for assessing direct ethical, legal and societal impact of risk based screening concepts

Document Date: 30/11/2021

Work Package 9: Ethics and Data Protection

Document Dissemination Level: Public

TRESSPASS
robus**T** **R**isk bas**E**d **S**creening and alert **S**ystem for **PASS**engers and luggage
is funded by the Horizon 2020 Framework Programme of the EU for Research and Innovation.
Grant Agreement number: 787120

ABSTRACT

The TRESSPASS project aims to develop, demonstrate and validate a single cohesive risk based border management concept for air, maritime and land border crossing points. A crucial part of this goal is an Ethics and Data Protection by Design (EDPbD) approach that the project follows.

The purpose of the TRESSPASS project design is to modernize and simplify security checks at border crossing points through AI technology under the assumption that the way border control is practiced does not meet the requirements of increasingly larger traffic volume anymore. Therefore, the idea is to develop an analytic framework to assess and to quantify risk in order to shorten border control procedures and adjust them individually to people based on specific indicators that can be measured as described in chapter 3.1.1 of this deliverable. Based on the estimated risk the types and number of security checks required for the traveller can be chosen. The goal is to maintain the best security level while optimizing the travel flows. This new approach and the use of AI technology may come with certain risks of their own and with concerns regarding e.g. fundamental rights. Therefore, an impact assessment for all technologies and developments is required.

As outlined in the previous deliverables D 9.6 *Typology of ethical, legal and societal issues of risk based screening* and D9.7 *Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts*, this deliverable aims to fill the need for a comprehensive framework for impact assessment (D9.7) which refers to the typology of ethical, legal and societal issues of risk based border management (D9.6). The framework enables a comparative assessment of different border check procedures and allows a better understanding of the risks and trade-offs that come with the implementation of risk based border checks.

The basis for the assessment in this deliverable are the Guidelines for Trustworthy AI which were released in 2018 by the High-Level Expert Group on Artificial Intelligence (AI HLEG) and specifically the *Assessment List for Trustworthy AI* (ALTAI) which is part of the Guidelines. The Guidelines aim to foster secure, ethically approved and innovative AI development in Europe. They are based on the research of the European Group on Ethics in Science and New Technologies. The foundation of the compiled ethics Guidelines is that "AI systems need to be human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom"[1]. AI HLEG was established by the European Commission specifically to formulate and enable the implementation of this vision. ALTAI is meant as a specific tool for self-evaluation for stakeholders during the process of designing, development and deploying of AI systems within the guidelines. As key requirements of Trustworthy AI, it specifies 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. The *Guidelines for Trustworthy AI* and especially ALTAI are foremost focussed and based on the principle of the protection of people's fundamental rights.

This document consists of five chapters. Chapter one presents the project background, the aim of the present deliverable and the input and output to it. Chapter two provides an outline of the assessment methods, a conclusion of previous findings and a description of the

[1]Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.

methodology used for the present assessment. Chapter three expounds ethics consideration in risk based border control methods in general, discusses the basic idea and distinguishes between three different kinds of risk based checks and screenings. Chapter four delivers the ethics assessment for all technology concepts within the TRESSPASS project. The deliverable continues with the integration with WP6 results on acceptability in chapter five and closes with the conclusion.

Project Information

Project Name	robusT Risk basEd Screening and alert System for PASSengers and luggage
Project Acronym	TRESSPASS
Project Coordinator	National Center for Scientific Research "Demokritos", EL
Project Funded by	European Commission
Under the Programme	Horizon 2020 Secure Societies
Call	H2020-SEC-2016-2017 (SECURITY)
Topic	SEC-15-BES-2017 "Risk based screening at border crossing"
Funding Instrument	Innovation Action
Grant Agreement No.	787120

Document Information

Document reference	**D9.8**
Document Title	**Updated framework for assessing direct ethical, legal and societal impact of risk based screening concepts**
Work Package reference	Task 9.3
Dissemination Level	Public
Author(s)	**Elisa Orru (ALU-FR), Marianne Grzondziel (ALU-FR)**
Contributor(s)	**Georgia Lavranou (NIUM), Hans Tangelder (VV), Mostafa Shahriari (SCCH), Evangelos Mantas (INF), Marcin Kowalski (WAT), Jeroen van Rest (TNO)**
Document Review Status	☒ Consortium ☒ WP leader ☒ Technical Manager ☒ Quality and Risk Manager ☒ Ethical Advisory Board ☐ Security Advisory Committee ☒ Project Coordinator

List of Acronyms and Abbreviations

ACRONYM	EXPLANATION
AI	Artificial Intelligence
BCP	Border Crossing Point
BCSEP	Shared Evaluation Platform for Border Crossings
BG	Border Guard
CCTV	Closed-circuit television
CHAT	Cultural-historical activity theory
CONOPS	Concept of Operations
C2	Command and Control System
DFA	Data Fusion & Analytics
DMS	Distributed Messaging System
DoW	Description of Work
DPO	Data Protection Officer
DRAS	Dynamic Risk Assessment System
EC	European Commission
EDPbD	Ethics and Data Protection by Design
ELSA	Ethical, Legal, Societal Aspects
EU	European Union
FP7	Seventh Framework Programme of the EU
H2020	Horizon 2020 Programme
KPI	Key Performance Indicator
LEA	Law Enforcement Agency
MMCAT	Multi Modal Communication Analysis Tool
VTC	Video Tracking Component
NGO	Non-governmental organization
PERSONA	Privacy, Ethical, Regulatory and Social “No-gate crossing point solution” Acceptance (research project)
RBBM	Risk based border management
RFID	Radio-frequency identification
SOTA	State of the art
SPOT	Screening of Passengers by Observation Techniques

TRESSPASS	robusT Risk basEd Screening and alert System for PASSengers and luggage
XP-DITE	Accelerated Checkpoint Design Integration Test and Evaluation (research project)

Table of Contents

1 INTRODUCTION

1.1 Background

The development, demonstration and validation of a cohesive risk based border management (RBBM) concept for air, maritime and land border crossing points is the aim of TRESSPASS. TRESSPASS is an innovation action project that focusses on border control tasks such as immigration control, customs and smuggling prevention, police searches but also on approaches for cross-border crime and terrorism prevention. An "ethics and data protection by design" concept based on in-depth ethical research is underlying to the TRESSPASS framework.

To ensure a procedure like this a range of "what-if"-scenarios was developed in order to provide decision makers and stakeholders (including BCP designers, legislators and members of the traveling public) with the necessary background in order to identify the differences between risk based border crossing procedure and current forms of customs and border checks with regard to ethical, legal and societal aspects – and how one could minimize such impact during the design of the single components and of BCPs.

Based on the above methods, contributions to the tech-focused work packages of the project are made. The main objective is to implement a "value-sensitive design during component development so as to mitigate or reduce ethical impact by changing how the specific technologies are designed".[2]

1.2 Aim of this document

The aim of this document is to update and finalise the ethical impact assessment started in T9.2 "Identification of relevant ethical, legal and societal risks" and carried on in T9.3 "Framework for assessment of relevant ethical, legal and societal impact".

This deliverable focuses on the assessment of single TRESSPASS components and provides a semi-quantitative output on the expected ethical impact of each of them. Being all the components developed in TRESSPASS a novelty if compared to current rule-based border checks, the assessment allows a scaled comparison with current border checks systems, as it allows to consider scenarios in which only some of the TRESSPASS components are applied.

Providing a comprehensive assessment of the overall TESSPASS systems is beyond the scope of this document, since it would require to take into account the total expected benefits or disadvantages, including non-ethically specific indicators referring to efficiency and security (such as changes in flow-rates and in security levels). An overall quantitative assessment is thus only possible considering the ethics KPIs in combination with the further KPIs identified for TRESSPASS. Such combination is the aim of the simulations carried out in WP7. A non-quantitative discussion of benefits and costs of the overall TRESPPASS system from an ethics point of view is provided in D9.9 *Ethical guidelines for decision-makers*.

[2] TRESSPASS D9.7, Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts, November, 2019.

1.3 Input / Output to this document

Input to this document is provided by previous T9.2-T9.3 deliverables: D9.6 *Typology of ethical, legal and societal issues of risk based screening concepts* and D9.7 *Framework for assessing direct ethical, legal and societal impact of risk based screening concepts*. It takes up part of the results of these deliverables, but also reworks the framework substantially, in order to integrate the recommendations of the Guidelines for Trustworthy AI published by the High-Level Group on AI in 2019, which were not available at the time in which the TRESSPASS framework was originally set up in D9.6 and D9.7. Integrating the Guidelines into the TRESSPASS framework also accommodates reviewers' suggestions made as part of the second project review.

Further input to this document has been provided by all technical WPs (WP3, 4 and 5), which developed the individual TRESSPASS components assessed in this document.

The results of the acceptability survey presented in WP6 as part of T6.3 "Integration of acceptability data into design criteria" provides additional input to this deliverable.

The present deliverable generates input for D9.9 *Ethical guidelines for decision-makers*. Outside of WP9, the results of the ethics assessment of single technologies will inform T7.5 in WP5 "shared evaluation platform for risk based border control systems", since ethics KPIs are part of the evaluation carried out in the simulations.

This deliverable is public. Its intended audience are BCP designers that considers using the types of functionalities covered by TRESSPASS. The assessment contained in this deliverable is intended to inform their decisions on the kind of ethical risks each component could bring about and how these can be mitigated. Additionally, industry developing these components can use the indications contained in this deliverable for adopting an ethics-by-design approach and proactively mitigating the possible ethical risks. The general public is not an intended target readership, but the deliverable can also provide useful information for the interested general public in order to understand the ethical implications of single technologies used in specific contexts. However, this document is not intended to provide a basis to answer the question whether risk based border management or single components are in general "ethical" or not.

2 ASSESSMENT METHODOLOGY

This section is based on previous work in WP9, especially the output of deliverable D9.6 *Typology of ethical, legal and societal issues of risk based screening* and D9.7 *Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts*. We also implement the suggestions by the reviewers to integrated the *Guidelines for Trustworthy AI* into the ethics framework (see 2.2.).

2.1 Previous results of WP9

In order to meet the goal within the TRESSPASS approach of a risk based border management concept for air, maritime and land border crossing points, WP9 follows an "Ethics and Data Protection by Design (EDPbD) approach that builds on previous ethical research"[3].

As a key method we identified ethical, legal and societal aspects (ELSA) considered as "unintended negative impact of introducing risk based border management". 12 types of such potential, ELSA-related negative impact were specified and grouped in three categories. These categories are:

- ELSA category A: privacy and data protection issues;
- ELSA category B: unfair distribution of impact across different social groups;
- ELSA category C: restrictions of societal freedoms and liberties.[4]

These categories and the underlying concept illustrated the evaluation framework that allows an assessment of "risk based screening concepts for border checks along the identified types of unintended negative impact"[5] which means a better understanding of positive and negative effects of the implementation of specific risk based approaches.

The conceptualization of these impact types enabled us to draft qualitative scales for assessment as part of the ethical evaluation framework. It is also consistent with the concept of operations (CONOPS) framework. This framework (according to the IEEE) proposes a communication document specifically designed to support a process of system development or system change. The document communicates the vision for alteration and shifts from the current system to the envisioned system to all stakeholders. The aim of such a document is to consolidate the "core concept behind this vision for change, in this case the TRESSPASS system for the development of risk based border management"[6]. The TRESSPASS CONOPS is based on Engeström's cultural-historical activity theory (CHAT) which describes a psychological framework for the analysis of interaction between groups or individuals and their sociocultural surroundings.

[3] TRESSPASS D9.7, Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts, November, 2019.

[4] TRESSPASS D9.6, Typology of ethical, legal and societal issues of risk based screening, November, 2018.

[5] TRESSPASS D9.6, Typology of ethical, legal and societal issues of risk based screening, November, 2018.

[6] TRESSPASS D9.7, Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts, November, 2019.

These approaches are meant to be tools for technology designers and decision makers in the context of border control procedures "to evaluate the impact of introducing risk based border checks and make ethically informed and well-balanced design decisions"[7].

2.2 Framework for trustworthy AI

In 2018 the European Commission released its vision for artificial intelligence (AI) which specifically aims to develop "ethical, secure and cutting-edge AI made in Europe"[8]. This vision is based on three key points: 1. increasing public and private investments in AI to boost its uptake, 2. preparing for socio-economic changes, and 3. ensuring an appropriate ethical and legal framework to strengthen European values. To enable the implementation of this vision the *High-Level Expert Group on Artificial Intelligence* (AI HLEG) was established. The AI HLEG is "an independent group mandated with the drafting of two deliverables: (1) AI Ethics Guidelines and (2) Policy and Investment Recommendations"[9].

2.2.1 Definition of AI systems

In addition to these two deliverables, the AI HLEG also suggested the following definition of 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".[10]

The TRESSPASS system overall and its single components do not carry out all functions specified in the definition. Data acquisition and inforamton processing are core functionalities perfomed within the TRESSPASS system. Both symbolic rules and numeric model learning are used in single components. However, none of the single TRESSPASS technologies, nor the system as a whole performs fully autonomous decision-making. The TRESSPASS system generates and presents to operators classification of travellers in risk-categories and the operators can use this classification to decide what actions shall follow. This fuctionality is performed by the DRAS (Dynamic Risk Assessment System) and not ba the single components assessed in this deliverable.

The rationale for using the Guidelines for trustworthy AI to assess the TRESSPASS components is that their framework is useful for assessing the ethical impact also of technologies performing parts, but not all, of the range of functionalities included in the definition.

[7] TRESSPASS D9.6, Typology of ethical, legal and societal issues of risk based screening, November, 2018.
[8] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.
[9] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.
[10] A definition of AI: Main capabilities and scientific disciplines, April 2019.

2.2.2 Guidelines for trustworthy AI

The AI Ethics Guidelines are based on the work of the European Group on Ethics in Science and New Technologies. The foundation of the compiled Ethics Guidelines is that "AI systems need to be human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom"[11]. This is all the more important as AI systems, although they offer great opportunities, can carry certain risks that must be monitored carefully. Trustworthy socio-technical environments in which there are embedded are therefore key: "In a context of rapid technological change, we believe it is essential that trust remains the bedrock of societies, communities, economies and sustainable development. We therefore identify Trustworthy AI as our foundational ambition, since human beings and communities will only be able to have confidence in the technology's development and its applications when a clear and comprehensive framework for achieving its trustworthiness is in place". "Trustworthiness is a prerequisite for people and societies to develop, deploy and use AI systems. Without AI systems – and the human beings behind them– being demonstrably worthy of trust, unwanted consequences may ensue and their uptake might be hindered, preventing the realisation of the potentially vast social and economic benefits that they can bring ".[12]

To establish this concept a measurement scale for trustworthy AI was developed. It describes three components which are crucial throughout the AI system's entire operating duration: "1. it should be lawful, complying with all applicable laws and regulations; 2. it should be ethical, ensuring adherence to ethical principles and values; and 3. it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm"[13]. All of these components must overlap and work interdependent since they are not by themselves sufficient to ensure a Trustworthy AI.

The Guidelines for trustworthy AI are addressed to all groups of stakeholders, including individuals and groups designing and developing, implementing, using or being affected by AI. They are designed not only to provide a list of ethical principles but to give guidance on the operationalization of these principles in socio-technical systems by the stakeholders:

The key guidance provided by the Guidelines can be summarized as follows:

- AI systems should be developed, deployed and used in a way that complies with the ethical principles of: respect (for human autonomy), prevention of harm, fairness and explicability. It is crucial to acknowledge and address the potential tensions between these principles. Situations involving more vulnerable groups (children, persons with disabilities and others that have historically been disadvantaged or are at risk of exclusion[14]), and situations in which asymmetries of power or information are suspected (such as between employers and workers or businesses and consumers) must be carefully monitored and evaluated.
- AI systems imply certain risk and may have a negative impact on individuals or society. This also include impacts which may be difficult to anticipate, identify or measure. It

[11] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.
[12] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.
[13] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.
[14] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.

is therefore important to "adopt adequate measures to mitigate these risks when appropriate, and proportionately to the magnitude of the risk"[15].

- AI systems must meet the seven key requirements for Trustworthy AI during development, deployment and use. The seven key requirements are: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal well-being and accountability.
- To implement those requirements technical and non-technical methods should be discussed.
- Research and innovation to help assess AI systems should be fostered, the results and open questions should be disseminated to the wider public, and a new generation of experts must be trained systematically in AI ethics.
- Communication is key: information to stakeholders about the AI system's capabilities and limitations should be passed clearly and in a proactive manner, enabling realistic expectations.
- The traceability and auditability of AI systems is very important, particularly in critical contexts or situations.
- Stakeholders are to be involved throughout the entire AI system's life cycle. Training and education should be offered so that all stakeholders are familiar with and trained in Trustworthy AI.
- There might be fundamental tensions between some principles and requirements. Therefore, it is crucial to identify, evaluate, document and communicate these trade-offs and possible solutions to stakeholders.
- A Trustworthy AI assessment list will can support the process of developing, deploying or using AI systems. Such a list must be adapted to the specific use case.
- Such an assessment list cannot be exhaustive and shall be regularly adapted to keep pace with technological innovation.

These Guidelines illustrate and provide a "horizontal foundation to achieve Trustworthy AI". Nevertheless, it is important to bear in mind that "different situations raise different challenges. It should therefore be explored whether, in addition to this horizontal framework, a sectorial approach is needed, given the context-specificity of AI systems". [16] The Guidelines are not meant to be interpreted as a substitute for policymaking or regulation in the context of AI systems. They are rather a living document that should be updated and pursued following the developments of technology, social environment and scientific knowledge. According to the Guidelines themselves, they can only be seen as the starting point for a profound discussion about the concept of trustworthy AI.

2.2.2 ALTAI

To specify and assist the process of analysis and evaluation in the context of development, deployment or use of AI systems the Guidelines for Trustworthy AI contain an *Assessment List*

[15] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.

[16] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.

for Trustworthy AI (ALTAI) which is "intended for self-evaluation purposes. It provides an initial approach for the evaluation of Trustworthy AI"[17]. The list names seven requirements of Trustworthy AI:

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;

7. Accountability.[18]

The ALTAI is founded on the principle of the protection of fundamental rights. It intends to help stakeholders and organizations to understand what Trustworthy AI is, what kind of risks AI systems can potentially generate, and how to reduce such risks while improving the benefits of AI for all stakeholders. ALTAI illustrates the impact AI can have on the environment and society, especially, for example, children or marginalised groups.

The list is intended to be flexible – stakeholders can add elements to it as needed in the particular process or refer to ALTAI elements. Thus, it enables the involvement of all groups of stakeholders."

[17] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.
[18] Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, April 2019.

3 ETHICS CONSIDERATIONS ON RISK BASED BORDER CONTROLS IN GENERAL

3.1 The basic idea of risk based checks and screening

The basic idea of TRESSPASS (robusT Risk basEd Screening and alert System for PASSengers and luggage) is the development of a risk based border management. "Risk based border management is about using border crossing points (BCPs) as a risk management measure that supports flow-, border- and national security"[19]. Border management is a tool for reducing or mitigating a wide array of potential risks. It can be seen as essential in this regard. The purpose of risk based approaches is to identify and use risk measures suitable for the actual situation and potential threats as well as reducing the remaining risk: "relaxed if possible, more stringent when needed. This implies that for people and goods that pose no significant threat, invasive checks at border crossing points can be limited"[20]. This would allow a faster flow at border control points, less interruptions and time-consuming procedures and less personal data review ("w.r.t. data already collected before arriving at the border crossing point") which might increase the "feeling of trust between public, public servants, commercial operators and travellers"[21].

TRESSPASS is designed to modernize and simplify security checks at border crossing points. The underlying perception is that the way border control has been carried out up until now is not suitable anymore in times of larger traffic volume. The aim of the project is to propose an analytic framework to assess risk and a systematic one to quantify it. It does so "based on a set of indicators that can accurately be measured across all four tiers of the Integrated Border Management"[22]. The individual risk calculation for each traveller is based on "the four-tier risk management approach introduced in TRESSPASS" [23] and rated on data collected from information gained through background checks, applications and sensors.

Based on the estimated risk "the system then adjusts the number and types of security checks required for each traveller, congruent with the associated risk, in order to maintain a desired security level while optimizing the security. Special attention is paid to maintaining a level of mutual trust between the security process and the traveller throughout the entire screening process"[24].

A detailed description of the conceptual framework for risk-based border management is provided in D1.2 "Conceptual Model" (public).

[19] TRESSPASS. The Project. Available at: https://www.tresspass.eu/The-project, October 2021.
[20] TRESSPASS. The Project. Available at: https://www.tresspass.eu/The-project, October 2021.
[21] TRESSPASS. The Project. Available at: https://www.tresspass.eu/The-project, October 2021.
[22] TRESSPASS. The Project. Available at: https://www.tresspass.eu/The-project, October 2021.
[23] TRESSPASS. The Project. Available at: https://www.tresspass.eu/The-project, October 2021.
[24] TRESSPASS. The Project. Available at: https://www.tresspass.eu/The-project, October 2021.

The following table compares risk based with intelligence led and rule based approaches to border management:

Topic	Rule-based	Intelligence-led	Risk-based
Type of checks	Dependent on general characteristics, such as EU/Non-EU)	Dependent on general characteristics, such as EU/Non-EU)	Dependent on risk-profile of the traveller
Fraction of travellers that is checked	All travellers are checked	All travellers are checked	All travellers are checked on identity, outside of that, trusted travellers may not be checked anymore.
Regulations	Regulations stipulate operational rules that all BCPs must follow.	Regulations stipulate rules that all BCPs must follow. Exceptions can be made for busy moments.	Regulations leave room for a local risk-based approach.
Decision-making	Rule-based	Rule-based	Based on risk acceptance
Information needs	General traveller information is required to determine the required type of check (based on EU law)	General traveller information is needed to determine required check (based on EU law), and additional information is needed to gain situational awareness (mostly for planning or instructing personnel)	Information is required on specific risk indicators to determine the type and level of checks for each individual traveller, and information is needed on changes in risks, threats and trends to determine proportionality under EU law
Purpose of screening	N/A	To assess how much capacity is required for this traveller	To assess the risk of a traveller, and thus which checks a traveller needs to undergo
Performance indicators	Flow-rate, efficiency, and the quality of checks	Flow-rate, efficiency, and the quality of checks	Flow-rate, efficiency, quality of checks, and the risk reduction

TABLE 1 COMPARISON BETWEEN BORDER CONTROL APPROACHES[25]

3.2 On different types of risk based border controls

During the project an assessment of risk based border checks was elaborated. It identifies 3 different types of risk based checks and screenings: risk profiling, behavioral analysis and situational checks.

1. Risk profiling: Risk profiling aims to differentiate travellers into risk categories by collecting and analysing background information.
2. Behavioural analytics: this type is based on behavioural data which is collected during the border checks procedures.
3. Situational risk assessment: this approach means that the intensity and the amount of resources used for checks depends on the situation and "contextual factors that do not relate to individual travellers"[26]. In case of a potentially threatening situation authorities would collect information about a specific situation but not about identifiable individuals.

Based on the classification above, TRESSPASS D9.7 *Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts* provided the following recommendation on risk assessment methods (section 3.3.5 of D9.7 Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts):

"Based on the preliminary findings discussed in the previous sections, it becomes clear that the data driven aspects of risk based border management (RBBM) can raise profound ethical issues that need to be addressed as part of defining and choosing risk assessment methods and technologies that enable their implementation. It is important to understand that the impact of introducing RBBM will, in the end, depend on the specifics of the implementation, i.e. the details of the design of the BCP procedures. Nevertheless, based on the preliminary findings, we find it plausible to argue that situational risk based border checks will likely imply less severe negative ethical impact than checks based on risk profiling or behavioral analysis.

While any form of border checks will involve negative ethical impact, situational risk based border checks may offer benefits for a lot travellers that outweigh added negative impact for some. Since traveller differentiation is not reliant on personal or identifying data and since

[25] Source: Table 6 1 "comparison between border control approaches", TRESSPASS D1.2 "Conceptual Model".

[26] TRESSPASS D9.7, Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts, November, 2019.

risk categorizations affect larger cross-sections of society equally (e.g. all travellers on a flight), privacy and data protection impact, as well as discrimination and restrictiveness issues can be limited if implemented accordingly.

The nature of risk profiling, on the contrary, involves the processing of personal and identifying information which raises considerable privacy and data protection issues. This is exacerbated by potential discrimination and restrictiveness issues, such as the feasibility of robust biometric identification, false or incomplete external data, transparency and accountability issues as well as the risk of data misuse of collected data. Since the use of such measures tend to yield a high negative impact, enabling technologies should be designed to limit the impact wherever possible. Even then, if implemented, risk profiling should be limited to special cases like second line checks, e.g. to support current forms of interviews that already imply a high privacy impact. Otherwise, it is likely that ethical benefits cannot plausibly outweigh the negative ethical impact on travellers.

Similarly, behavioral analysis raises considerable issues with regard to potential discrimination or other unfair distribution of impact across certain traveller groups. Hence, special attention must be paid to such issues when deciding about implementation in border checks – especially with regard to the feasibility for certain groups of travellers or regarding the fair distribution of false alarms. If such impact can be limited effectively, ethical benefits can still only plausibly outweigh negative impact with regard to restrictiveness, if potential enrolment processes do not effectively increase the temporal costs for travellers and if the transparency and accountability impact can be minimized effectively" [27].

As detailed in the next sections, this assessment still holds. Both risk profiling and behavioral analysis imply serious ethical risks that can be mitigated only partially. Situational risk assessment still remains the approach to risk based border controls for regular border crossing points which potentially minimizes the negative ethical impact.

[27] TRESSPASS D9.7, Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts, November, 2019.

4 ETHICS ASSESSMENT

4.1 Introduction

In this section, we assess selected single components that have an independent ethical impact. We do not apply the ALTAI list to components whose functionality consist in integrating or aggregating the input from other components (such as C2, DFA, DRAS and DMS), since we assume that these technologies give rise to an ethical impact of a different kind of the one that can be assessed by the given criteria.

However, it must be noted that these system components can have a distinct ethical impact deriving from their capacity of aggregating information from different sources. As Daniel Solove has noted, "Aggregation upsets these expectations, because it involves the combination of data in new, potentially unanticipated ways to re-veal facts about a person that are not readily known. People give out bits of information in different settings, only revealing a small part of themselves in each context. Indeed, people selectively spread around small pieces of data throughout most of their daily activities, and they have the expectation that in each disclosure, they are revealing rela-tively little about themselves. When these pieces are consolidated to-gether, however, the aggregator acquires much greater knowledge about the person's life".[28]

Furthermore, DRAS is the system component performing the risk-assessment. As such, it can give rise to ethical issues related to Human Agency and Oversight, a criteria used for the assessment of the single components. For instance, issues relating to over-reliance of operators on the results of the automatic risk assessment or augmented pressure on end-users to follow the indications of the automatic risk assessment can arise. Although these ethical risks are generated by DRAS and not by the single components, they can give rise to a different level of severity depending on the combination with carachteristics of the single components. Moreover, one component, TLTP, does not provide input to DRAS and is therefore not affected by this ethical risk. In order to reproduce these modulations, we have decided to mention the ethical impact related to over-reliance and augmented pressure to follow the DRAS results in the assessment of the single components.

Moreover, we also exclude from the assessment those components, such as TCA and SPA, that only serve the communication of information to travellers or the exchange of information between border personnel and do not collect or process travellers' data on their own.[29]

Measures to mitigate the ethics impact of the TRESSPASS components (such as red teaming for improving accuracy, encryption of data, development of a training concept for operators, data protection impact assessment, monitoring by DPOs and ethics committees …) have already been implemented during the project following an ethics-by-design approach and are documented in the specific deliverables describing the functionalities of the single

[28] Solove, Daniel J., A Taxonomy of Privacy, in: University of Pennsylvania Law Review, 154/3, 2006, 477–564, p. 508.

[29] Regarding SPA, it must be noted, however, that this tool also enables BGs to manually ingest RIs for travellers or their vehicles and to scan passenger's passports. The ethical impact deriving from the first functionality has been taken into account in the single technologies from which the risk indicators manually ingested derive (i.e. LPDS), while passport scanning is considered to be part of rule-based checks procedures and is not assessed in this deliverable.

components and in the WP9 and WP12 deliverables. The following assessment focuses on the **residual** ethical impact, namely the impact still remains after all TRESSPASS-mitigating options have been put in place. It suggests measures to mitigate this residual impact, but it does not discuss whether the realization of such measures is realistic or possible while maintaining the current functionality of the system.

It most also be noted that the TRL level achieved by the single TRESSPASS components assessed here ranges between 5 and 7. This means that these technologies still are in the demonstration phase and are not yet ready for launch and operations.[30] To learn about the ethical implications is part of the TRL progress and the results of this assessment can provide input for further refinements in the next steps of technology development towards TRL7+. The TRL level of each technology at the end of TRESSPASS is mentioned in the respective sections' titles.

The specific impact of the single components depends also on the related risk indicators. For instance, the level of specificity of each indicator and its characterization as weak or strong affect proportionality (since the use of a less specific and weak indicator can be considered to be less proportional – other things remaining equal – than a strong indicator with a high level of specificity) However, given that risk indicators are treated in TRESSPASS as classified information, their assessment cannot be provided here. A specific impact assessment of the risk indicators for the air pilot has been provided as an attachment to D2.2. (classified).

The overall TRESSPASS system is built modularily: it can function with all technologies at once or with only some of them. Moreover, single components can be used in different ways, for instance for screening all passengers or only the ones for which a previous alert has occurred. Whenever ethically relevant, these variations are mentioned in the brief technology descriptions provided in the introductiory subsections of the single assessments. TRESSPASS has also developed two simulators (the iCrowd Simulator and the FHG simulator developed in WP7) that can be used for demonstrating the changing ethical impact along with other performance indicators of different system configurations. A discussion of the ethical aspects of the TRESSPASS risk based concept in general, including the expected advantages and the wieder ethical impact is provided in D9.9 "Ethical guidelines for decision makers".

For each of the remaining components, following the ALTAI recommendations, we first carry out a Fundamental Rights Impact Assessment (FRIA) in advance of the assessment with the specific ALTAI criteria. The evaluation highlights the potential ethical issues identified by answering the questions catalogue contained in the ALTAI and whose full list is available online.[31]

Our assessment considers all 7 criteria contained in the ALTAI tool. However, as far as the category "Technical robustness and safety" is concerned, we restricted the focus to only one of its components, namely "accuracy", the other three components mentioned in the ALTAI ("Resilience to attack and security", "General safety" and "Reliability, Fall-Back Plans and Reproducibility") being addressed specifically in D 5.9 "Report on information security measures".

[30] An exception is LPDS, which has a TRL 9. Hoever, LPDS had this TRL level already at project begin and, additionally, it gives rise to limited ethical concerns (see section 4.7 below).

[31] https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

For the assessment we use the following color code:

no specific impact or entity unknown at the current stage	Moderate ethics impact	High ethics impact	Severe ethics impact

4.2 TRA (achieved TRL: 7)

Traveller Registration Application (TRA) is a pre-registration for travallers at border control points. It is a mobile application for Android systems and also accessible online. TRA corresponds with and provides data to other internet based TRESSPASS work packages.

The process of using TRA is split in two phases: firstly the passenger applies for registration, provides the required personal data and receives a link which must be opended for confirmining registration. After the completed registration, the user is allowed to the next phase of TRA where a new entry for the travel can be created and the related data entered. “Anytime he/she completes the trip creation, the trip and passenger data are sent from TRA to DMS, and then to other components of TRESSPASS system leading to generation of indicators for the actual risk assessment from DRAS”.[32]

4.2.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
A) Non-Discrimination	Risk of discrimination of those groups who do not have access to a digital support for registering or are not able to fill in a digital registration form (potential discrimination based on age, social origin, language, opinion, property, disability).	Provide non-digital alternatives or assistance for registration at the BCP. Additional inclusive design measures specifically addressing age, disability and language.
	The question regarding the possession of health insurance can lead to discrimination based on the economic situation (property).	Delete the related question.
	Risk of discrimination against demographic groups (elderly; digital minimalists etc.) without a social media account if the non-possession of social media accounts is evaluated as a risk indicator.	Do not consider entering no social media account a risk indicator. Increase stakeholder participation by also including representatives of travellers and potentially discriminated groups in the development phase.
B) Rights of the Child		

[32] TRESSPASS D5.7 Integrated and Tested TRESSPASS System. Dynamic Risk Assessment and Alert System, May, 2021.

C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results	The application increases considerably the amount of data collected about travellers, most of which currently do not have to provide such data while planning a travel. Serious issues of necessity and proportionality regarding social media accounts. (The encryption of some data, already integrated in TRA, improves cybersecurity but not the privacy issues related to proportionality and necessity).	Reduce the categories of data collected (data minimization). Exclude social media account from the information required. Regulatory framework ensuring the right to object.
D) Freedom of expr., inform., assembly, assoc.	The request to provide social media accounts data can inhibit these freedoms by inducing self-censorship on social media.	Transparent information about the use of social media information (however, this could be counterproductive depending on the way this information is used for the risk assessment). Do not ask to provide social media account.

4.2.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The data collected by TRA flow into the automatic risk assessment which is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision (i.e. to conduct second-line checks) is the result of a completely human or of an automatically-supported decision (also considering that the time laps between the TRA registration and the border checks).	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
	Over-reliance of operators on the results of the automatic risk assessment performed by DRAS is possible.	Provide training to operators which also includes effective methods to counteract over-reliance.
	Augmented pressure on end-users to follow the indications of the automatic risk assessment performed by DRAS (need to justify deviations from the automatic risk assessment).	Regulative framework appropriately addressing liability issues. Appropriate training.
	Limited human oversight for this component: human-in-command aspects	Strengthen human oversight as the possibility for human operators to

	are foreseen as the possibility for end-users not to act upon the outcome of the whole risk assessment, but not to sort out the contribution of this single component to the overall risk assessment. No fallback option to a rule-based system foreseen in case of dysfunctionalities (see D 5.9 "Report on information security measures").	monitor the system's operation (human-on-the-loop). Introduce detection and response mechanisms for undesirable adverse effects of the system (such reverse engineering and black-box methods to detect unintentional discriminations) for subjects. Introduce fallback options (human-in-command).
2) Technical Robustness and Safety **(limited to Accuracy)**	The input to the TRA consists in non-verified information (self-registration provided by the travellers themselves).	Introduce verification mechanisms (this, however, could increase the negative impact in other respects, especially regarding privacy and data protection, introducing a verification spiral).
	So far no mechanisms to document (by means of statistics) the false-positive and false-negative rates tha can be detected by operators or ad-hoc functionalities.	Introduce monitoring and documentation of false-positives and false-negatives, i.e. via mandatory statistics.
3) Privacy and Data Governance	See FRIA, section C) above.	See FRIA, section C) above.
4) Transparency	Traceability[33]: So far no mechanisms for tracing back which information lead to a given risk assessment.	Introduce mechanisms that enable to trace back a given output to a given input.
	Confidentiality of the way this component contributes to risk indicators negatively affects the explainability of the results to the travellers.	Make risk indicators transparent. Explore other ways for explaining the risk assessment results to travellers.
	Since this component has not yet been deployed in real word in comparable contexts and as part of a risk-based border management Communication issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers. Appropriate training.
5) Diversity, Non-discrimination and Fairness	See FRIA, section A) above.	See FRIA, section A) above.
6) Societal and Env. Well-being	Possible increase in energy used and carbon emissions	Evaluate environmental impact and if necessary introduce measures to reduce it.

[33] Here and whenever this aspect is assessed in the following pages, "traceability" is understood as the internal traceability within each component. By contrast, within the overall TRESSPASS system it is possible to trace back which input came from which component.This information can be displayed to end-users operating the system.

	Possible impact of democracy related to the use of social media information – see FRIA, section D) above.	see FRIA, section D) above.
7) Accountability	Since the system has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

4.3 LSI (achieved TRL: 6)

At first sight, LSI seems to respond to a rule-based logic, by checking the travellers' data against existing databases (such as SIS, VIS, EES, PNR). However, LSI merges the rule-based model with a risk based logic. It relies, among others, on the PNR databases, which partially follow a risk based logic themselves and use non-criminalistic information (such as itinerary, baggage details, mode of payment etc.) for providing a risk assessment. The use of information that has no close link to illegal conduct and can lead to discrimination on the basis of social status, property and cultural background. The impact of the LSI depends a lot on the risk indicators used and on the information they rely on. Since TRESSPASS risk indicators are classified, the assessment conducted here cannot make direct reference to them. An ethics impact assessment of the risk indicators used in the air pilot are provided as attachment to D2.2 "Risk indicators" (classified).

During the TRESSPASS pilots simulated ingestion databases were used. These include additional, not yet existing databases and simulations of the existing ones, whose architecture and functionality however in part deviated from the existing ones (see TRESSPASS D3.3 *Interfaces to external systems* and D3.7 *Interfaces to external systems*).

The assessment conducted here, however, is done under the hypothesis that the LSI relies on the ingestion data of only existing databases and provides information relevant for the risk calculation. The following impact assessment is therefore not to be considered as an assessment of the components as they were tested in the TRESSPASS pilots, but as an assessment of the impact of LSI if it were used with ingestion of existing databases.

Given the fact that the specifications of LSI in TRESSPASS and in real life are/will be provided by end-users (LEAs and BGs), the involvement of the former is particularly crucial for the success of mitigation activities.

The compatibility of the PNR Directive with fundamental rights is currently under review by the EUCJ. The EUCJ's decision will thus provide further guidance on the possible ethical and fundamental rights impact deriving from the use of the PNR database by LSI.

4.3.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS

A) Non-Discrimination	Information ingested into the LSI from databases such as PNR (baggage, kind of itinerary chosen, mode of payment) etc., if used for risk assessment, can lead to discrimination based on property (choice of cheapest route, payment in cash, …) and cultural background (country of origin/destination).	Not use the related information for risk assessment (for further details see ethics assessment of risk indicators in D2.2).
B) Rights of the Child		
C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results	The datasets used for ingestion into the LSI contain biometric data.	Ensure that the use of biometric data complies with strict human rights standards.
D) Freedom of expr., inform., assembly, assoc.		

4.3.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The data processed by LSI flow into the automatic risk assessment which is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision (i.e. to conduct second-line checks) is the result of a completely human or of an automatically-supported decision.	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
	Over-reliance of operators on the results of the automatic risk assessment performed by DRAS is possible.	Provide training to the end-users which also includes effective methods to counteract over-reliance.
	Augmented pressure for end-users to follow the indications of the automatic risk assessment performed by DRAS (need to justify deviations from the automatic risk assessment).	Regulative framework appropriately addressing liability issues. Appropriate training.
	Limited human oversight for this component: human-in-command aspects are foreseen as the possibility for end-users not to act upon the outcome of the whole risk assessment, but not to sort out the contribution of this single component	Strengthen human oversight as the possibility for human operators to monitor the system's operation (human-on-the-loop).

	to the overall risk assessment. No fallback option to a rule-based system foreseen in case of dysfunctionalities (see D 5.9 "Report on information security measures")..	Introduce detection and response mechanisms for undesirable adverse effects of the system (such reverse engineering and black-box methods to detect unintentional discriminations) for subjects. Introduce fallback options (human-in-command).
2) Technical Robustness and Safety	So far no mechanisms to document (by means of statistics) the false-positive and false-negative rates tha can be detected by operators or ad-hoc functionalities.[34]	Introduce monitoring and documentation of false-positives and false-negatives, i.e. via mandatory statistics.
3) Privacy and Data Governance	The datasets used for ingestion into the LSI contain biometric data.	Ensure that the use of biometric data complies with ethical standards by: - Making sure that the data used are strictly necessary for the risk assessment; - Assess proportionality; Adopt encryption, log records etc.
4) Transparency	So far no mechanisms for tracing back which information lead to a given risk assessment.	Introduce mechanisms to trace output back to specific input.
	Confidentiality of the way this component contributes to risk indicators negatively affects the explainability of the results to the travellers.	Make risk indicators transparent. Alternatively: Explore other ways for explaining the risk assessment results to travellers.
	Since this specific component has not yet been deployed in real word as part of the TRESSPASS risk-based border management Transparency issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers. Appropriate training.
5) Diversity, Non-discrimination and Fairness	See FRIA, section A) above.	See FRIA, section A) above.
6) Societal and Env. Well-being	Possible increase in energy used and carbon emissions due to the increased information processing.	Evaluate environmental impact and, if necessary, introduce measures to reduce it.

[34] Even if we refer to similar systems deployed in real-world, such as the PNR framework, the false-positive and false negative rates are neither part of the compulsory information that member states have to provide to the Commission for its evaluation statistics, nor provided by the Commission itself it its evaluation reports. See Art. 20 of the Directive (EU) 2016/681 on the use of passenger name record data and COM(2020) 305 final, 24.07.2020.

7) Accountability	Since the system has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

4.4 WI (achieved TRL: 6)

The Web Intelligence (WI) module gathers information from open sources (Internet and Darknet) to add to the risk indicators that are used to calculate the potential risk per traveller. The purpose is to support border guard authorities using the risk based border management approach and enable a more robust classification of potential risk.

The rationale for using the WI components relies on the assumption that some types of mala fide travellers or victims of mala fide travellers (human trafficking) publish information online that is relevant for border checks. WI calculates predefined attributes of travellers based on public social media information which by themselves or when combined with other attributes contribute to risk indicators. It does not allow border guards to look themselves for other information. WI would typically only be deployed after hits on other risk indicators.

WI is interconnected with other components within TRESSPASS and is used in various stages of border control procedures: "Web Intelligence is connected to the project's message bus, the Distributed Messaging System (DMS), which is responsible for triggering the various system components by distributing the various messages as generated by the TRESSPASS C2 application. Via the DMS, the WI communicates with the other TRESSPASS components and exchanges information. The main component where the WI shares its outcome is DFA, where the combination of the individually computed RIs will be performed"[35].

4.4.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
A) Non-Discrimination	Web intelligence and social media analysis can provide insight into interest in specific topics and political opinions, as well as on employment status. If risk indicators based on these data are used, then discrimination based on sensitive attributes such as "political opinion", "ethnic and social origin", "membership of a national minority" and "property" are likely.	Not use risk indicators based on the information mentioned. Introduce methods of reverse engineering to check whether discrimination takes place.

[35] TRESSPASS D5.7 Integrated and Tested TRESSPASS System. Dynamic Risk Assessment and Alert System, May, 2021.

	Web intelligence and social media analysis can reveal social graphs and exposes contacts with specific groups of people. If risk indicators based on these data are used, then discrimination based on sensitive attributes such as "political opinion", "ethnic and social origin", "membership of a national minority" and "property" are likely.	Not use risk indicators based on the information mentioned.
B) Rights of the Child	Discrimination of minorities and socially marginal groups can have an increased negative impact on children rights.	
C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results	Web analysis can draw attention to people interested in certain topics for reasons not related to the intended targeted group. Such reasons include professional interest (journalism, social workers…), political activism, research activity (i.e reasons having nothing to do with the aimed target group). This non-specificity impacts negatively on necessity and proportionality.	
D) Freedom of expr., inform., assembly, assoc.	Web intelligence can impact negatively on freedom of association and assembly: people can refrain from activities that are legal and indeed crucial for the flourishing of democracy in order not to impact negatively on their risk level (chilling effect).	

4.4.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The WI output flows into the automatic risk assessment which is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision (i.e. to conduct second-line checks) is the result of a completely human or of an automatically-supported decision (also considering that the time laps between the provision of data used for the pre-travel screening through WI and the border checks).	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
	Over-reliance of operators on the results of the automatic risk assessment performed by DRAS is possible.	Provide training to the end-users which also includes effective methods to counteract over-reliance.

	Augmented pressure for end-users to follow the indications of the automatic risk assessment performed by DRAS (need to justify deviations from the automatic risk assessment).	Regulative framework appropriately addressing liability issues. Appropriate training.
	not to act upon the outcome of the whole risk assessment, but not to sort out the contribution of this single component to the overall risk assessment. No fallback option to a rule-based system foreseen in case of dysfunctionalities (see D 5.9 "Report on information security measures").	Strengthen human oversight as the possibility for human operators to monitor the system's operation (human-on-the-loop). Introduce detection and response mechanisms for undesirable adverse effects of the system (such reverse engineering and black-box methods to detect unintentional discriminations) for subjects. Introduce fallback options (human-in-command).
2) Technical Robustness and Safety	Among others, discard mechanisms for how non-reliable results (see D4.2 and D4.9) are incorporated into the system. However, so far no mechanisms to document (by means of statistics) the false-positive and false-negative rates tha can be detected by operators or ad-hoc functionalities.	Introduce monitoring and documentation of false-positives and false-negatives, i.e. via mandatory statistics.
3) Privacy and Data Governance	Web intelligence (both dark and open web) can provide a wide range of very sensitive information regarding subjects directly targeted and their contacts.	Reduce to the minimum the categories of information explicitly targeted Reduce at the minimum the information processed as "collateral" findings.
	Proportionality issues related to the triggering of the WI and depending on whether it is activated for all travellers or only for the ones having received a hit on legacy databases.	Only activate the search in specific, selected cases with a clear motivation for doing so.
4) Transparency	So far no mechanisms for tracing back which information lead to a given risk assessment.	Introduce mechanisms to trace output back to specific input.
	Confidentiality of the way this component contributes to risk indicators negatively affects the explainability of the results to the travellers.	Make risk indicators transparent. Alternatively: Explore other ways for explaining the risk assessment results to travellers.

	Since the system has not yet been deployed in real word Communication issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers. Appropriate training.
5) Diversity, Non-discrimination and Fairness	See FRIA, section A) above.	See FRIA, section A) above.
6) Societal and Env. Well-being	Possible negative impact on democracy and societal atmosphere due to chilling effect (see FRIA, section D) above).	Do not process any information which can provide inferences on opinions, religion and the exercise of the right of association.
	Possible increase in energy used and carbon emissions due to the increased information processing.	Evaluate environmental impact and, if necessary, introduce measures to reduce it.
7) Accountability	Since the system has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

4.5 VTC and RTBA (achieved TRL: both 5)

The idea behind the Video Tracking Component (VTC) is to establish a camera system to not only film the border control point area but to track a person through the area and and analyse their behavior. Real-Time Behavioural Analytics (RTBA) system is designed to generate "latent representations for micro, short and long term movement patterns for the identification of intent subgroups"[36].

The hypothesis behind the use of these components is that some physical behaviour at BCPs can be relevant for risk assessments and that the related information can be captured by VTC and RTBA. This information can be then attributed to specific travellers by the DFA, by combining it with information obtained through identity authentication systems such as travel document scanners or biometric systems.

An extensive description of the functionalities is provided in D3.1, D.3.5 and D4.10.

4.5.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
A) Non-Discrimination	Risk of discrimination based on disability/age (elderly people with restricted mobility,	Train the system using diverse and representative data or add this

[36] TRESSPASS D9.3, Project baseline for research ethics, August 2019, updated December 2020.

	children and their accompanying persons) or other physical states influencing mobility.	information as a feature to be analysed in RTBA. Introduce methods of reverse engineering to check whether discrimination takes place.
B) Rights of the Child		
C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results	Tracking is considered (among others by the HLEG on AI) to be a particularly sensitive application. This negatively impacts on proportionality, especially considering that movement patterns can only give weak indications of criminal intent or illegal conduct.	Delete the VTC tracking data, after sending these to the RTBA component. Do not store movement patterns, store only risk indicators derived from these patterns.
D) Freedom of expr., inform., assembly, assoc.	Behavior and movement patterns at border crossing areas can be influenced by character, disposition towards crowdy places, attitude towards surveillance systems, opinions etc. Risk of stigmatization of any kind of behavior which deviates from what is set as "normal" and which can influence movement patterns at border crossing areas. Interest in specific topics can be dependent on other factors such as scientific/research/information interest: possible impact on freedom of information and expression.	

4.5.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The VTC/RTBA output flows into the automatic risk assessment which is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision (i.e. to conduct second-line checks) is the result of a completely human or of an automatically-supported decision.	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
	Over-reliance of operators on the results of the automatic risk assessment performed by DRAS is possible.	Provide training to the end-users which also includes effective methods to counteract over-reliance.
	Augmented pressure for end-users to follow the indications of the automatic risk assessment performed by DRAS (need to	Regulative framework appropriately addressing liability issues. Appropriate training.

	justify deviations from the automatic risk assessment).	
	Limited human oversight for this component: human-in-command aspects are foreseen as the possibility for end-users not to act upon the outcome of the whole risk assessment, but not to sort out the contribution of this single component to the overall risk assessment. No fallback option to a rule-based system foreseen in case of dysfunctionalities (see D 5.9 "Report on information security measures").	Strengthen human oversight as the possibility for human operators to monitor the system's operation (human-on-the-loop). Introduce detection and response mechanisms for undesirable adverse effects of the system (such reverse engineering and black-box methods to detect unintentional discriminations) for subjects.
		Introduce fallback options (human-in-command).
2) Technical Robustness and Safety	Loose link between movement patterns and criminal intention impacts negatively on accuracy.	Introduce monitoring and documentation of false-positives and false-negatives, i.e. via mandatory statistics on performance.
	Small and non-diverse datasets used to train the system (RTBA must be trained specifically for each border crossing area and its layout).	Use more representative training datasets.
3) Privacy and Data Governance	Tracking is considered (among others by the HLEG on AI) to be a particularly sensitive application. This negatively impacts on proportionality, especially considering that movement patterns can be influenced by many factors that are not related to criminal intent or illegal conduct.	Reduce to the minimum the categories of information explicitly targeted Reduce at the minimum the information processed as "collateral" findings
	Proportionality issues related to the triggering of the VTC and depending on whether it is activated for all travellers or only for the ones having received a hit on legacy databases.	Only activate the tracking in specific, selected cases with a clear motivation for doing so.
	Second-line checks on the basis of "abnormal" movement patterns can lead to disclosure of sensitive health-related information (i.e. protheses otherwise concealed).	Distinguish among different kinds of "abnormal" movement. E.g. use time for reaching the BCP as indicator for having a prothesis.

4) Transparency	So far, no mechanisms for tracing back which input data lead to a given risk assessment.	Introduce mechanisms to trace output back to specific input.
	Confidentiality of the way this component contributes to risk indicators negatively affects the explainability of the results to the travellers.	Make risk indicators transparent. . Alternatively: Explore other ways for explaining the risk assessment results to travellers.
	Since the system has not yet been deployed in real word Communication issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers. Appropriate training.
5) Diversity, Non-discrimination and Fairness	See FRIA, section A) above.	See FRIA, section A) above.
6) Societal and Env. Well-being	Possible negative impact on democracy and societal climate due to pressure to conform (see FRIA, section D) above).	Only use the VTC in very restricted areas.
	Possible increase in energy used and carbon emissions due to the increased information processing.	Evaluate environmental impact and, if necessary, introduce measures to reduce it.
7) Accountability	Since the system has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

4.6 TCSS (achieved TRL: 7)

The Thermal counter spoofing sensor (TCSS) is a sensor used for detection of presentation attacks (i.e. wearing a mask to conceal one's face). "The sensor provides probability score that the subject's face (a traveller) is a real person or a presentation attack"[37] based on the thermal emissions registered. In TRESSPASS, no images of subjects being screened were stored. All the data were captured in a buffer which is considered a random access memory, and was cleaned once read or used.

The assessment is based on the assumption that TCSS works without face recognition features and without generating facial pictures, corresponding to its original functionality. If face-recognition features are integrated into the system, the negative ethics impact is expected to increase considerably and the following ethics assessment must be updated accordingly.

[37] TRESSPASS D9.3, Project baseline for research ethics, August 2019, updated December 2020.

4.6.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
A) Non-Discrimination	Possibility of detecting anomalies in thermal emissions related to physical conditions or health conditions that can produce, for instance, increased sweating.	Ensure accuracy of results and exclude false positives due to physical or health conditions, for instance based on non-uniformity of sweating as an indicator of non-spoofing.
B) Rights of the Child		
C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results		
D) Freedom of expr., inform., assembly, assoc.	Uncertainty on whether biometric identification is done could give rise to chilling effects.	Clear communication of functionality.

4.6.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The TCSS output flows into the automatic risk assessment which is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision (i.e. to conduct second-line checks) is the result of a completely human or of an automatically-supported decision.	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
	Over-reliance of operators on the results of the automatic risk assessment is possible.	Provide training to the end-users which also includes effective methods to counteract over-reliance
	Augmented pressure for end-users to follow the indications of the automatic risk assessment (need to justify deviations from the automatic risk assessment).	Regulative framework appropriately addressing liability issues. Appropriate training.
2) Technical Robustness and Safety		
3) Privacy and Data Governance	The technology consists in screening an indiscriminate number of travellers to take decisions that may affect them.	Carefully assess the benefits (accuracy of results, existence of alternatives to detect presentation attacks…) as a

	Proportionality issues can thus arise depending on the effective added-value of the technology.	prerequisite to clarify proportionality issues.
	Second-line checks on the basis of "abnormal" thermal emissions can lead to disclosure of sensitive health-related information.	Ensure accuracy of results (minimize false positive).
4) Transparency	Confidentiality of the way this component contributes to risk indicators negatively affects the explainability of the results to the travellers.	Make risk indicators transparent. Alternatively: Explore other ways for explaining the risk assessment results to travellers.
	The use of the technology can be associate with face recognition.	Clearly inform on the functionality and the exclusion of face recognition.
	Since the system has not yet been deployed in real word Communication issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers. Appropriate training.
5) Diversity, Non-discrimination and Fairness	See FRIA, section A) above.	See FRIA, section A) above.
6) Societal and Env. Well-being		
7) Accountability	Since TCSS has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

4.7 LPDS (achieved TRL: 9)

The Legacy Possession detection systems (LPDS), as used in the TRESSPASS pilots, aims to detect hidden contraband goods by scanning large objects (trucks). It uses X-ray imaging and the output is interpreted by a trained human.

In D3.1 there is an extensive general description of LPDS. The only LPDS system in use in TRESSPASS pilots is the Rapiscan M450, designed to screen trucks, and used for detection of hidden contraband (specifically cigarettes). The description is based on this example. The system can screen large objects, more specifically trucks up to a certain size. It uses transmission X-ray imaging. A trained human image interpreter is an integral part of it. Although the component is integrated in the whole TRESSPASS risk based system, the ingestion of a binary risk assessment into the overall systems is decided and operated manually by a human.

An extensive description of the functionalities is provided in D3.1 *Sensors* and D5.7 *Integrated and Tested TRESSPASS System*.

The following assessment is carried out in accordance with the TRESSPASS scenarios in which this component is used to detect prohibited items only. It is possible to ingest the output into the overall system manually. Should the technology be used for detecting humans the ethics implications are likely to be more severe and the following assessment should be updated.

4.7.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
A) Non-Discrimination		
B) Rights of the Child		
C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results		
D) Freedom of expr., inform., assembly, assoc.		

4.7.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The LPDS output flows into the automatic risk assessment which is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision (i.e. to conduct second-line checks) is the result of a completely human or of an automatically-supported decision.	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
2) Technical Robustness and Safety		
3) Privacy and Data Governance		
4) Transparency		
	Since the system has not yet been deployed in real word Communication issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers.
5) Diversity, Non-discrimination and Fairness		

6) Societal and Env. Well-being		
7) Accountability	Since the system has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

4.8 TLTP (achieved TRL: 7)

The Travellers and luggage tracking sensor platform (TLTP) is a tool based on RFID tracking technology that is used at border checkpoints, such as airports, for luggage and passenger tracking.

Through the RFID tracking component luggage and passengers are tagged with RFID tags that are scanned by specialized RFID readers connected with the computer system. The system sends the gathered (location) information to a server so that luggage and passenger information can be matched and analyzed.

An extensive description of the functionalities is provided in D3.1 *Sensors* and D5.7 *Integrated and Tested TRESSPASS System*.

Most of the ethics impact described below is expected to originate from the component's functionality tracking passengers. The ethics impact would be significantly reduced if the functionality would be limited to baggage tracking. Due to the fact that TLTP involves tracking and behavioural anomalies detection like VTC, the ethical impact of both technologies is similar. However, TLTP does not involve recognition of subjects and does not track the walking patterns of travellers to the same level of granularity than VTC. Therefore, in some cases its ethics impact, although similar in quality, can be considered of lower intensity.

4.8.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
A) Non-Discrimination	Risk of discrimination based on mental or physical states than can lead to unusual movement patterns.	Introduce methods of reverse engineering to check whether discrimination takes place.
B) Rights of the Child		
C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results	Tracking is considered (among others by the HLEG on AI) to be a particularly sensitive application. This negatively impacts on proportionality, especially considering that movement patterns can be influenced by many factors non related to criminal intent or illegal conduct.	Limit the use of TLTP to baggage tracking.

D) Freedom of expr., inform., assembly, assoc.	Behavior and movement patterns at border crossing areas can be influenced by character, disposition towards crowdy places, attitude towards surveillance systems, opinions etc. Risk of stigmatizing any kind of behavior which deviates from what is set as "normal" and which can influence movement patterns at border crossing areas and resulting pressure to conform to normal movement patterns.	

4.8.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The output of TLTP flows into the automatic risk assessment which is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision (i.e. to conduct second-line checks) is the result of a completely human or of an automatically-supported decision.	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
	Over-reliance of operators on the results of the automatic risk assessment performed by DRAS is possible.	Provide training to the end-users which also includes effective methods to counteract over-reliance.
	Augmented pressure for end-users to follow the indications of the automatic risk assessment performed by DRAS (need to justify deviations from the automatic risk assessment).	Regulative framework appropriately addressing liability issues. Appropriate training.
	Limited human oversight for this component: human-in-command aspects are foreseen as the possibility for end-users not to act upon the outcome of the whole risk assessment, but not to sort out the contribution of this single component to the overall risk assessment. No fallback option to a rule-based system foreseen in case of dysfunctionalities (see D 5.9 "Report on information security measures").	Strengthen human oversight as the possibility for human operators to monitor the system's operation (human-on-the-loop). Introduce detection and response mechanisms for undesirable adverse effects of the system (such reverse engineering and black-box methods to detect unintentional discriminations) for subjects. Introduce fallback options (human-in-command).

2) Technical Robustness and Safety	Loose link between movement patterns and criminal intention impact negatively on accuracy.	Introduce monitoring and documentation of false-positives and false-negatives, i.e. via mandatory statistics on performance.
3) Privacy and Data Governance	Tracking is considered (among others by the HLEG on AI) to be a particularly sensitive application. This negatively impact on proportionality, especially considering that movement patterns can only give weak indications of criminal intent or illegal conduct.	Reduce to the minimum the categories of information explicitly targeted Reduce at the minimum the information processed as "collateral" findings.
	Proportionality issues related to use of TLTP for all passengers without specific reasons.	Only activate the tracking in specific, selected cases with a clear motivation for doing so.
	Second-line checks on the basis of "abnormal" movement patterns can lead to disclosure of sensitive health-related information.	
4) Transparency	So far no mechanisms for tracing back which input data lead to a given risk assessment.	Introduce mechanisms to trace output back to specific input.
	Confidentiality of the way this component contributes to risk indicators negatively affects the explainability of the results to the travelers.	Make risk indicators transparent. Alternatively: Explore other ways for explaining the risk assessment results to travellers.
	Since the system has not yet been deployed in real word Communication issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers. Appropriate training.
5) Diversity, Non-discrimination and Fairness	See FRIA, section A) above.	See FRIA, section A) above.
6) Societal and Env. Well-being	Possible negative impact on democracy and societal climate due to pressure to conform (see FRIA, section D) above).	Only use the TLTP in very restricted areas.
	Possible increase in energy used and carbon emissions due to the increased information processing.	Evaluate environmental impact and, if necessary, introduce measures to reduce it.
7) Accountability	Since the system has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

4.9 MMCAT (achieved TRL: 6)

The Multi Modal Communication Analysis Tool is an interview support system, providing technology for a security/authority officer ('Interviewer') taking an interview. Its typical application is in interviews with travellers with an elevated risk level, which currently are perfomed by human operators as part of the second-line checks. It is meant to be used in line with the concept of operations and training that border guards have for conducting interviews. It would typically be used in conjunction with a view on the information that caused the traveller to be led to the interview in order to enable operators to consider alternative explanations for the assessment that has conducted to the interview.

The Interviewer uses the information provided by MMCAT to choose lines of questioning and as a support for the assessment if a traveller is telling the truth. A difference from the other TRESSPASS components is that MMCAT is not connected to the DRAS, so it does not contribute to the classification travellers or their behaviours into threats.

MMCAT records the interview with a camera aimed at the traveller. The software is developed to detect facial expressions and to assess the pose of the traveller. Based on these expressions and the traveller's pose the border guard gets real time information about indicators of the traveller's emotions. The visualization of these indicators on a dashboard supports the border guard to choose the line of questioning.[38]

An extensive description of the functionalities is provided in D3.1 *Sensors (Interim)* and D3.5 *Sensors (Final)*.

4.9.1 FRIA

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
A) Non-Discrimination	Risk of discrimination based on physical and psychological status concerning health, emotionality etc.	Train the system using diverse and representative data. Introduce methods of reverse engineering to check whether discrimination takes place.
B) Rights of the Child		Do not use the system to interview children.
C) Personal Data (necessity and proportionality), especially considering the amount of data collected/processed and the expected results	The scientific soundness of technologies aiming at distinguishing veracity from deception on the basis physiological states is questioned by psychophysiologists and psychologists. This negatively impacts on proportionality, especially considering that	

[38] MMCAT Component Infor for pilots, January 2020.

	physiological reactions can only give weak indications of veracity.	
D) Freedom of expr., inform., assembly, assoc.	Physiological reactions during an interview with border guards can be influenced by character, disposition towards law-enforcement authorities, attitude towards surveillance systems, opinions etc. Risk of stigmatization of reactions that deviate from what is set as "normal".	

4.9.2 Ethics assessment

IMPACT CATEGORY	SPECIFIC IMPACT OF THE EXAMINED COMPONENT	MITIGATION OPTIONS
1) Human Agency and Oversight	The MMCAT output is used to support border guards in decision-making. The system can generate confusion for travellers on whether the decision is the result of a completely human or of an automatically-supported decision.	Inform the travellers about the way the system works and the fact that the decision is the result of an automatically-supported decision.
	Over-reliance of operators on the results of the automatic risk assessment performed by DRAS is possible.	Provide training to the end-users which also includes effective methods to counteract over-reliance.
	Augmented pressure for end-users to follow the indications of the automatic risk assessment performed by DRAS (need to justify deviations from the automatic risk assessment).	Regulative framework appropriately addressing liability issues. Appropriate training.
	Limited human oversight for this component: human-in-command aspects are foreseen as the possibility for end-users to ignore the outcome of the whole risk analysis, but not of the single stages.	Strengthen human oversight as the possibility for human operators to monitor the system's operation (human-on-the-loop). Introduce detection and response mechanisms for undesirable adverse effects of the system (such reverse engineering and black-box methods to detect unintentional discriminations) for subjects.
2) Technical Robustness and Safety	Loose link between psychophysiological reactions and criminal intention impact negatively on accuracy.	Introduce monitoring and documentation of false-positives and false-negatives, i.e. via mandatory statistics on performance.

3) Privacy and Data Governance	Technologies aiming at detecting deception from psychophysiological reactions are particularly sensitive and controversial applications. This negatively impact on proportionality, especially considering that such reactions can only give weak indications of criminal intent or illegal conduct.	
	The detection of psychophysiological reactions considered to be abnormal can lead to disclosure of sensitive health-related information (heart problems, blood pressure, anxiety, psychological illnesses etc.).	MMCAT does and should establish a personal base line of what is normal.
4) Transparency	So far, no mechanisms for tracing back which input data lead to a given risk assessment.	Introduce mechanisms to trace output back to specific input.
	Confidentiality of the way this component contributes to risk indicators negatively affects the explainability of the results to the travellers.	Make risk indicators transparent. Alternatively: Explore other ways for explaining the risk assessment results to travellers.
	Since the system has not yet been deployed in real word Communication issues are still open.	Openly communicate benefits, technical limitations (error rates/level of accuracy) to users and travellers. Appropriate training.
5) Diversity, Non-discrimination and Fairness	See FRIA, section A) above.	See FRIA, section A) above.
6) Societal and Env. Well-being	See FRIA, section D) above.	See FRIA, section D) above.
7) Accountability	Since the system has not yet been deployed in real word Auditability issues are still open.	Establish mechanisms to facilitate the auditability of the component by third parties. Enable ethical oversee by independent bodies.

5 INTEGRATION WITH WP6 RESULTS ON ACCEPTABILITY

The acceptability study presented in WP6 refers to the acceptability of no-gate crossing point technology solutions in general. Its results, therefore, cannot be directly and unequivocally associated to single TRESSPASS technologies. Nevertheless, they provide useful insight on travellers attitudes that are likely to influence the acceptance of TRESSPASS solutions as well.

The key findings of the study can be summarized as following:

- overall, travellers are more inclined to accept sharing data with authorities if these data items are clearly related to the border security process (such as identity details). Levels of acceptance are lower when the data to be shared have no obvious meaning for border security, such as economic data related to income status, persons paying for the travel and social media account.
- Similarly, perceived effectiveness is higher (and perceived intrusiveness lower) regarding data categories whose meaningfulness for security purposes is most evident (such as travel history) and lower for information whose security relevance is less obvious (such as social media account), whereas perceived intrusiveness is higher for this kind of data.
- Levels of acceptance decrease significantly when the purpose of data processing is to contrast irregular migration as compared to combating terrorism.
- With regard to behavioral data, the greatest concern of participants regarded the possibility that their behavior would be misinterpreted by border guards. Also, most participants expected that their social environment would require to share these data.

In general, these findings correspond with some of the results of the ethics assessment carried out in this deliverable, especially regarding the sensitiveness of individual profiling, of the use of social media accounts and of risk of normalization effects related to the use of behavioral data for individual risk assessment.

This alignment in the results support the notion that ethics and social acceptance are interrelated and complementary concepts. “For example, when some people would not accept specific border control technologies and protest them, that might affect or change others’ people views on whether these technologies are ethically accepted. The close relevance of these concepts suggests a collaborative approach when exploring them. That is, social acceptance results can feed into the ethics assessment and refine the list of the ethical issues. Vice versa, criteria and issues included in the ethics assessment can provide ideas for the development of the acceptance assessment, such as questionnaires.

Finally, public changing attitudes and acceptance towards border technologies need to be monitored, accounting for the changing landscape in border control processes and technologies. The assessment tools that will be used to monitor acceptance should be treated as living documents which are updated following the developments of technology, social environment, and knowledge.” [39]

[39] D.6.6 *Evolving CONOPS Framework (Final)*.

6 CONCLUSION

Based on the previous research, the *Guidelines for Trustworthy AI* and the *Assessment List for Trustworthy AI* (ALTAI), this report has presented a description of a framework for impact assessment on risk based border control tools and methods. As shown in chapter 4, the framework allows an assessment of different components that can be used modularly at border crossing points and allows therefore a comparison between different checkpoints designs. The assessment aims at a better understanding of the potential risks, concerns and trade-offs involved in introducing risk based border checks as part of future border management strategies.

The assessment refers to selected single system components that have an independent ethical impact. Components that are meant to integrate or combine the input from other components (e.g. C2, DFA, DRAS, DMS) were not assessed, since it was assumed that these technologies do not produce an independent ethical impact that is not captured through the assessment of the other components. Components that only serve the purpose of communication of information to travellers or between border personnel and do not collect or process travellers' data on their own were also not assessed (e.g. TCA and SPA).

Following the ALTAI recommendations, a Fundamental Rights Impact Assessment (FRIA) was first carried out for each of the assessed components. This was followed by an ethics assessment that illustrates the potential ethical issues that could be identified around the seven criteria mentioned in the ALTAI List. These are Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-discrimination and Fairness, Societal and Environmental Well-being, Accountability.

To visualize the assessment outcome a colour code was used. The category "no specific impact or entity unknown at the current stage" was marked neutral, the category "Moderate ethics impact" was marked yellow, "High ethics impact" was coloured orange and "Severe ethics impact" was highlighted red.

Some of the identified ethics issues are cross-cutting and refer to most or all of the components, since they depend on the context (the overall risk based concept) they are included in. In particular:

1) An impact on **Autonomy and Human Oversight** was identified which depends on the fact that the single components are connected into an automatic risk assessment. This can: a) generate confusion for travellers about the human-made or automated nature of the decision; lead to over-reliance of the operators on the automated results and pressure to follow them; c) lead to limited human oversight. The related mitigating options include transparent information for travellers on the decisional process, appropriate training, the development of a regulative framework addressing liability issues and technical and governance mechanismes increasing human oversight.
2) **Explainability** of results is negatively affected by the confidentiality of the way each component contributes to risk indicators. Possible mitigating options consist in making risk indicators transparent or (given that for end-users this option may not desirable for strategic reasons) exploring alternative ways for explaining the risk assessment results to travellers.
3) **Communication and Auditability** issues are in most cases still open because the systems have not been deployed in real-world. In the implementation phase, it would be therefore important to openly communicate benefits, technical limitations (error

rates/level of accuracy) to users and travellers, to provide appropriate training to operators, to stablish mechanisms to facilitate the auditability of the component by third parties and to enable ethical oversee by independent bodies.

The assessment, moreover, showed that seven components still present a *severe ethics impact* in one or more of the ALTAI criteria:

1) Several components (TRA, LSI, WI, VTC and RTBA, TLTP and MMCAT) present a severe ethics impact in relation to **Non-Discrimination**. In the case of TRA, LSI and WI, this impact depends on the kind of data used for the risk assessment; in the case of VTC and RTBA, TLTP and MMCAT this impact is dependent on the intrinsic logic of detecting "abnormal" behaviour. Accordingly, mitigation options range from excluding certain categories of data from the data collection process to not using selected risk indicators based on behavioural and, finally, to introducing methods of reverse engineering to check whether discrimination takes place.
2) Regarding **Accuracy** (as part of Technical Robustness and Safety), TRA, LSI, VTC and RTBA, TLTP and MMCAT present a severe ethics impact mostly related to the fact the the input information consists in non-verified information, that in some cases the link between the information used for the risk assessment and criminal intention is loose and that so far no mechanisms to document false positives and false-negatives rates have been introduced. The corresponding suggested mitigation options consist in introducing verification mechanisms (which, however, could increase the negative impact on privacy and data protection by introducing a verification spiral) and monitoring false-positives and false-negatives rates by means of statistics on system's performance. These statistics should be carried out by default and their results independently verifiable.
3) With the exception of LPDS, all components present a severe impact related to **Transparency** dependent on the confidentiality of risk indicators (for mitigation options, see the point about "Explainability" above).
4) A severe impact on **Privacy** has been assessed for TRA, WI, VTC and RTBA, TLTP and MMCAT. This impact refers to; a) the increase of the amount of data collected about travellers; b) issues of necessity and proportionality regarding the use of certain kinds of sources (such as social media) to retrieve information and the related possibility to collect sensitive information, the use of tracking technologies and the scientific soundness of technologies aiming at distinguishing between deception and veracity on the basis of physiological states. Mitigation options could be identified for some but not all of these impact categories. The suggested options range from reducing the categories of data collected, the introduction of a regulatory framework ensuring the right to object to provide certain data.
5) Four technologies, namely TRA, WI, VTC and RTBA and MMCAT present a serious impact on **Societal and Environmental Well-being**, especially concerning democratic participation, the freedom of association and assembly, the freedom of expression and information and non-conformistic behaviour. Mitigation options were available for some of these aspects and include not to process any information which can provide inferences on opinions, religion and the exercise of the right of association.
6) One technology (WI) present an impact on the **Rights of the Child**, since discrimination of minorities and socially marginal groups can have an increased negative impact on children rights.

The presented mitigating options involve technical, legal, operational and policy measures and can be most effectively implemented if all these domains are involved in the mitigation

process. Beyond the ethical impact of the single components, moreover, also the wider ethical impact shall play a role in the decisional process about the implementation of risk based border management. This wider general impact is presented in D9.9 “Guidelines for decision-makers”.

7 DOCUMENT VERIFICATION AND PLAGIARISM CHECK

This section provides information on document verification including how to complete a plagiarism check. A plagiarism check is mandatory before the submission of a document to the European Commission.

7.1 Verification and Quality Assurance

Before the document's submission to the European Commission, it will be up to the WP leader, Task Leader and the Coordinator to review internally and verify the quality of the content of the deliverable, with careful consideration of the document's dissemination level.

7.2 Plagiarism check and risk mitigation

The document's main author, with the assistance of the WP leader and the Coordinator, is responsible for performing a plagiarism check. Plagiarism is an important issue that can have serious consequences (up to termination of the project by the European Commission) if not addressed. The European Code of Conduct for Research Integrity (published by the European Science Foundation ESF and All European Academies, ALLEA) along with the H2020 reference documents state that no kind of plagiarism is tolerated.

The following kinds can immediately be identified:

- Uncredited verbatim copying of individual elements (sentences, paragraphs and illustrations) of great or small extent,
- Uncredited improper paraphrasing of pages or paragraphs,
- Credited verbatim copying of major/minor portions of text without clear delineation[40]
- Uncredited verbatim copying of one's own work is considered self-plagiarism.

The WP leader and Coordinator will assure quality of the document and verify that the relevant checks have been completed.

Please note that **although there exists a large number of free online tools for plagiarism checking, it is recommended that they be avoided for all security-sensitive deliverables.** Manual checking when necessary should be sufficient.

[40] with proper credits and text delineated (for example by using quotes), it can be considered quoting and not plagiarism. Still, quoting should be used to a small extent, i.e. a few sentences not entire paragraphs. Delineation means that it must be immediately apparent in the text that it has been copied from another source, using quotes or a box surrounding the text and that it is differentiated by the deliverable text.

REFERENCES

EU Commission 2020, Report from the Commission to the European Parliament and the Council on the review of Directive 2016/681 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime, COM(2020) 305 final, 24.07.2020.

Independent High-Level Expert Group on Artificial Intelligence 2019, A definition of AI: Main capabilities and scientific disciplines, April 2019.

Independent High-Level Expert Group on Artificial Intelligence 2019, Ethics Guidelines for Trustworthy AI, European Commission, Brussels.

Independent High-Level Expert Group on Artificial Intelligence 2020, The Assessment List for Trustwothy Artificial Intelligence (ALTAI), European Commission, Brussels.

MMCAT Component Infor for pilots, January 2020 (TRESSPASS internal document).

Solove, Daniel J., A Taxonomy of Privacy, in: University of Pennsylvania Law Review, 154/3, 2006, 477–564.

TRESSPASS 2021, The Project, Available at: https://www.tresspass.eu/The-project, October 2021.

TRESSPASS 2021, D5.7 Integrated and Tested TRESSPASS System. Dynamic Risk Assessment and Alert System, May, 2021.

TRESSPASS 2020, D9.3, Project baseline for research ethics, Available at: https://www.tresspass.eu/sites/default/files/publications/public/TRESSPASS%20D9.3%20Periodic%20ethical%20report%201%20WP9%2013%20-%20ALU-FR%20Report%20Public%2030.pdf, November 2021.

TRESSPASS 2019, D9.7 Framework for assessing direct ethical, legal and societal impact of risk based border screening concepts, Available at: https://www.tresspass.eu/sites/default/files/publications/public/TRESSPASS%20D9.7%20Framework%20for%20assessing%20direct%20ethical,%20legal%20and%20societal%20impact.pdf. November 2021.

TRESSPASS 2019, D1.2 Conceptual Model.

TRESSPASS 2018, D9.6 Typology of ethical, legal and societal issues of risk based screening, Available at: https://www.tresspass.eu/sites/default/files/publications/public/TRESSPASS%20D9.6%20Typology%20of%20ethical,%20legal%20and%20societal%20issues%20of%20risk%20based.pdf, November, 2021.