Tailored IoT & BigData Sandboxes and Testbeds for Smart, Autonomous and Personalized Services in the European Finance and Insurance Services Ecosystem

D2.7 – Security and Regulatory Compliance Specifications – I

Lead Beneficiary	FTS
Due Date	2020-05-31
Delivered Date	2020-05-31
Revision Number	2.0
Dissemination Level	Confidential (CO)
Type	Report (R)
Document Status	Draft (Release)
Review Status	Internally Reviewed and Quality Assurance Reviewed
Document Acceptance	WP Leader Accepted and Coordinator Accepted
EC Project Officer	Pierre-Paul Sondag

HORIZON 2020 - ICT-11-2018

This project has received funding from the European Union's horizon 2020 research and innovation programme under grant agreement no 856632500228229-1

Contributing Partners

Acronym	Role[1]	Name Surname[2]
FTS	Lead Beneficiary	Jürgen Neises
DWF	Contributor	Marc Meerkamp, Axel von Goldbeck
BOS	Contributor	Maja Skrjanc
GFT	Contributor Quality Assurance	Marcelo Colomer, Marina Cugurra, Martin Felix de Miguel Lillo, Ernesto Troiano
ATOS	Contributor	Nuria Ituarte Aranda, Ignacio Elicegui
GRAD	Contributor	Lilian Adkinson Orellana
CP	Contributor	Marinos Xynarianos
Pilots (BANKIA, JRC, BOI, PRIVE, LIB, BOC, NBG, FBK, BOS, AKTIF, PI, ATOS, SILO, WEA, GEN AGRO)	Contributors	Massimiliano Aschi, Pablo Carballo, Nikolaos Droukas, Elena Femenía, Eymard Hooper, Klaudija Jurkosek-Seitl, Bruno Lepri, Lukas Linden, Giorgos Marinos, Orkan Metin, Grigoris Mykdakos, Paul O'Connel, Aristodemos Pnevmatikakis, Carlos Alberto Portero, Petra Ristau, Silvio Walser
RB	Internal Reviewer	
BOUN	Internal Reviewer	Can Özturan

Revision History

Version	Date	Partner(s)	Description
0.1	2020-01-15	FTS	ToC Draft
0.2	2020-03-10	FTS	ToC Version updated
0.3	2020-03-17	FTS	Focus Regulations
0.4	2020-04-24	FTS	Input on Focus Regulations from Workshop
0.5	2020-05-04	FTS	Alignment with Ethics requirements
0.6	2020-05-12	DWF	Include Pilot Statements on Ethics
0.7	2020-05-18	DWF	Include Pilot Statements on Ethics
0.8	2020-05-19	BOS, JSI	Add requirements on AI
0.9	2020-05-20	DWF	Preliminary Draft
1.0	2020-05-25	FTS, DWF, BOS, JSI	First Version for Internal Review
1.1	2020-05-27	FTS, RB, BOUN, GFT	Version for Quality Assurance
1.2	2020-05-28	FTS	Version for Submission
2.0	2020-05-31	FTS, GFT	Version approved by Coordinator

[1] Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance

[2] Can be left void

Executive Summary

The goal of task T2.4 is the specification of the standards and regulatory environment of the INFINITECH project. This task will specify the standards and regulations that will drive the INFINITECH developments, including the technology developments of the project.

This deliverable is the first version of a total of two deliverables which are meant to provide the outcome of task T2.4. This version of the document elicits, which **regulations** are in focus of the INFINITECH project with regards to the pilots use cases.

Specific emphasis is put on the **GDPR** due to its high relevance in BigData and analytics scenarios which apply in the INFINITECH's sharp services. Moreover, latest guidelines **on AI and its application in the Financial Sector** are considered as a valuable cross cutting insight for the overall project.

Finally, relevant Financial Sector regulations as PSD II, MiFiD II and 4AML are assessed with respect to the INFINITECH pilot scenarios.

In particular the deliverable contains

- An overview on the emerging AI related requirements.
- A description of high-level requirements for GDPR, PSD II, MiFiD II and 4AML in the INFINITECH project
- A detailed description of GDPR related requirements in the INFINITECH SHARP services.

At this stage of the project the most relevant requirement seems generating the awareness of the pilots regarding personal data and the details on how to deal with the various requirements resulting from the GDPR.

This results in a non-typical development related technical of functional requirements. However, there are certainly general data protection requirements, which can be obtained from this exercise:

- Data shall be checked in detail, if they have to be considered as personal data. This is underestimated in general.
- **Anonymization** and **pseudonymization** are key technologies within the INFINITECH project.
- Technical and organizational measures for data protection shall be considered in the pilots.
- Data Privacy Impact Assessments and Informed Consent will play an important role in the pilots.

Furthermore, the emerging role of AI guidelines and regulations will need attention during the project and the next phase of task T2.4.

The work related to task T2.4 will continue until Month 15, when the 2nd version of this deliverable will be submitted (D2.7), with the updates on these topics.

Table of Contents

List of Figures

List of Tables

Abbreviations

AML	Anti-Money-Laundering
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
GDPR	General Data Protection Regulation
KYC	Know Your Customers
MiFID	Markets in Financial Instruments Directive
MiFIR	Markets in Financial Instruments and Amending Regulation
NDA	Non-Disclosure Agreement
NIS	Network and Information Systems
OES	Operators of Essential Services
PAN	Primary Account Number
PaaS	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
PIA	Privacy Impact Assessment
PSD2	Payment Service Directive 2
PSP	Payment Service Provider
PSU	Payment Service User
P2PP	Peer-to-Peer Payment
RTS	Regulatory Technical Standard
QTSP	Qualified Trust Service Provider
SCA	Strong Customer Authentication
SME	Small and Medium-Sized Enterprises
SA	Supervisory Authority
SECaaS	Security-as-a- Service
TI	Threat Intelligence
3DS	Three-Domain Secure

1 Introduction

The purpose of the deliverable is the initial identification of standards and regulations relevant within the INFINITECH project based on the pilots and its data-based reference scenarios and services. A first set of relevant standards and regulatory compliance requirements for the INFINITECH pilots is assessed, i.e. GDPR, PSD II, MiFiD II, 4AML describing high level requirements for each regulation. The deliverable also includes a brief overview of other relevant standards and regulations.

Special focus is on the GDPR in the context of INFINITECH's approach leveraging BigData for Financial Services' innovation and the role of GDPR in context of INFINITECH's SHARP (Smart, Autonomous, Regulatory compliant, Personalized) services.

With the relation to AI and analytics, latest regulatory approaches towards AI in Finance by the European Commission and the European Banking Authority have been considered. This deliverable includes a first assessment of those.

Finally, this deliverable includes an ETHICS assessment of each pilot with respect to the usage of personal data.

1.1. Objective of the Deliverable

The main objective of the deliverable is obtaining a set of relevant regulatory and standards-based requirements considering the technical developments of the project. The deliverable shall facilitate focusing on the most relevant requirements among the vast amount of standards and regulations related to the pilots.

1.2. Insights from other Tasks and Deliverables

The deliverable is based on the pilot descriptions and user stories described in deliverable D2.1 and the services view outlined in deliverable D2.3. Moreover, especially the pilot contributions on GDPR and Ethics partially resulting from a joint online technical workshop of WP2 and WP7 contributed to this deliverable.

1.3. Structure

The document is structured as follows:

- Section 1 gives a brief overview on scope and objectives of the deliverable.
- In section 2 the methodology obtaining the insights of this deliverable is explained.
- Section 3 gives an overview of standards and regulations in the Financial Sector and determines the focus regulations of the project.
- The next section 4 illustrates the high-level requirements resulting from the pilots' questionnaire.
- The following section 5 provides an overview of the situation of the GDPR in the context of INFINITECH SHARP services.

The appendices include:

Appendix A)	References to literature
Appendix B)	Requirements lists
Appendix C)	Descriptions of the pilots on ethics requirements

2 Methodology

The major source of insights is the INFINITECH pilots and the legal advisory partner DWF.

The first pillar of this deliverable comprises of the pilots descriptions generated in T2.1 and gathered in D2.1. With this in mind, a wide preliminary set of regulations could be identified.

These were addressed in iterative sessions with the pilots to narrow down the focus to a specific set of most relevant regulations related to the INFINITECH project and the foreseen SHARP services.

With a further step, the situation of pilots regarding these regulations was assessed in a questionnaire, which was supplemented by explanations of the legal understanding.

Finally, the specific situation considering personal data was analysed in a round of interviews with pilots utilizing personal data.

The results of those contributions are explained in the following sections.

3 Standards/Regulation Overview

This section starts with an overview of the regulatory landscape. It shows the vast number of standards and regulations, which affect the Financial Services landscape. Clearly, these standards and regulations cannot be considered completely.

Moreover, there are many standards, which are renowned as industry best practice and thus part of the all-day routine of Financial Services providers. Among those ITIL, including an Information Security Management System and the ISO 27000 series are the most prominent ones. Thus, we consider those being regular practice in any Financial Service provider.

The focus of the project shall be on regulations, which influence most the INFINITECH approach of SHARP services. Hence, a dedicated focus is put on AI in view of the latest regulations by the European Commission and the European Banking Authority.

3.1 Overview on Regulations

Standards and regulations that govern INFINITECH services and that drive its developments do stem primarily from financial regulations, as well as from the law on data protection.

As matter of fact, INFINITECH services are subject to a much wider list of standards, regulations and guidelines on a European as well as on a national – including the implementation of European law- and transnational level. A non-exhaustive list of these has been gathered by the consortium and is included below in Table 1.

Table 1 Overview of Regulations in the Financial Sector

Transnational security and compliance Standards
ITIL
ISO/IEC 27000 series
ISO/IEC 27001
ISO/IEC 27005 (Security Risk Management)
ISO/IEC 20000 for Providers
ISO 22301 (Business Continuity)
ISMS implementation guidance developed by SC27 COBIT
SWIFT Customer Security Controls Framework

PCI-DSS
ISAE 3402
ISO/IEC TR 27015 ISMS Guidance for Financial Services
SOX
Payments Accounts Directive (PAD)
FISMA
European security and compliance Standards
Basel II/III/IV
Solvency II
MiFiD II
AMLD4
PSD2
EBA revised guidelines on outsourcing arrangements
GDPR (and national implementations)
NIS
ePrivacy
eIDAS
DIRECTIVE (EU) 2016/943
DIRECTIVE (EU) 2018/843
COUNCIL DIRECTIVE 2006/112/EC (and national implementations)
Regulation EU/847/2015
Regulation (EU) No 575/2013
Regulation (EU) No 1286/2014 (PRIIPs)
Regulation (EU) No 596/2014
Regulation (EU) 2017/1129
Regulation (EU) 2016/1011
Regulation (EU) 600/2014
German example for national security and compliance standards
FAIT 5
MaRisk
HGB AO
UStG
GOB
GOBD
KWG

Bankaufsichtliche Anforderungen an die IT (BAIT)
BaFin - Orientierungshilfe zu Auslagerungen an Cloud-Anbieter
BSI IT-Grundschutz
UStG
TKG, TMG
DSGVO (national Implementation of GDPR)
Bankengeheimnis
GwG
IT Sicherheitsgesetz (national Implementation of NIS)
BGB
Spanish national security and compliance standards
Políticas de seguridad para la pyme (INCIBE)
Slovene national security and compliance standards
Banking Act (Banking Act 2)
The Law on the Bank of Slovenia (ZBS-1)
ZPPDFT-1 (implementation of AMLD4)
ZPlaSSIED (implementation of PSD3)
ZInfV (implementation of NIS)
Law on Protection of Personal Data (implementation of GDPR)
General principles for participation in TARGET2-SLOVENIJA
Other national security and compliance standards
ISO20022
ISO20022 Migration
ECB - Instant Payments
EC - Crypto Assets Consultation - Regulation coming
EC - Ethics Guidelines for Trustworthy AI
ISO TC/307 Blockchain and Distributed Ledger Technologies
Contractual/IP
IP-Rights/Licenses

Considering that the INFINITECH project will innovate the landscape of financial services facilitating the deployment of BigData driven services, it is a valid approach to consider that the standards and guidelines, which apply to the specific business, are applied in the financial industry and thus the financial sector beneficiaries of the project according to the state of the art and with best industry practice.

Having this in mind, this deliverable focusses on the most relevant regulations with respect to BigData and AI driven analytics in the pilots of INFINITECH.

3.2 Focus point AI

Recently, AI technology became mature, widely available, incorporated in various business and social processes and even in our private lives. There seems to be a global consensus on the fact that AI is influencing our personal and societal lives, which in turn raises various ethical issues related to AI. While AI technology is a mystery and, therefore, creates certain discomfort and generates fears for most of the people; in many ways, it is also a mystery whose consequences are not yet fully understood for AI researchers. Therefore, it is necessary to regulate AI technology to remove fears and to make it trustful.

There are several documents addressing common issues from slightly different angles. One of the recent documents [1], addresses ethical issues using a holistic framework of interdependent values, principles, and actions that can guide societies in the research, design, development, deployment, and use of AI systems, referring to human dignity as a compass to deal responsibly with the known and unknown impacts of AI systems in their interactions with human beings and their environment.

EC issued a white paper On Artificial Intelligence - A European approach to excellence and trust [2] in order to set up coordinated European approach on the human and ethical implications of AI as well as a reflection on the better use of big data for innovation. The white paper points up two main objectives: (i) to promote the uptake of AI and (ii) to address the risks associated with certain uses of this new technology. The purpose of that document is to set out policy options on how to achieve these objectives. In pursuing this purpose, the key element to keep in mind is trustworthiness. EBA published a report on the recent trends of BD&AA [3] in the banking sector and on the key considerations in the development, implementation and adoption of BD&AA. In the report, EBA identified four key pillars - data management, technological infrastructure, organization and governance and analytics methodology – as necessary to support the rollout of Advanced Analytics, along with a set of “elements of trust”. EBA supports the rollout of Advanced Analytics, along with the following “elements of trust” that need to be properly and sufficiently addressed:

- Ethics
- Explainability and interpretability
- Fairness and avoidance of bias
- Traceability and auditability
- Data protection
- Data quality
- Security
- Consumer protection

The EBA is of the view that additional efforts are needed to ensure that BD&AA solutions respect and integrate these “elements of trust”, which is well in line with white paper of EC. Towards meeting this objective, a risk-based approach could apply on certain “elements of trust” depending on the impact of each BD&AA application. In the paper EBA suggests that, for example, stricter requirements may apply on the “explainability” element when there is a potential impact on business continuity or a potential harm to the customer.

All the above mentioned papers agree that AI may become an ever more central part of every aspect of people’s lives and therefore needs to be properly addressed. All three papers emphasize that the way to address is to create trust- so that people should be able to trust it. There are minor differences, how other subtopics are defined, for instance fairness and explainability, transparency and traceability and few others.

The main purpose of these activities is to create awareness among different stakeholders and, in particular, to ensure regulators and supervisors are well informed on the developments, in an effort to support technological neutrality across the regulatory and supervisory approaches.

Defining an AI certification framework is a fairly complex task, especially since the AI field and methodologies are quite diverse. Therefore, we narrow down the major AI certification tasks so as to cover several trustworthy principles that can be measured objectively, using standard machine learning metrics.

Our goal is to propose a framework for AI certification, which will focus mainly on two principles of trustworthiness, proposed by EBA:

- Fairness and avoidance of bias
- Traceability and auditability

AI certification should provide answers to two main questions:

- Is the AI System actually doing what its creators are claiming?
  This part of the task consists of analysis of various standard performance measures and robustness, which could be done using a software tool. While the other part of this task require assessment of transparency and audit of methodology, including architecture and integration.

- Is the AI System compatible with a particular value standard?
  This task includes the assessment of compliance with ethical, social and legal norms, regulations, principles, and criteria. The norm and standards are provided by other stakeholders in documents such as UN Human Rights Declaration, OECD AI Principles, EC Ethics guidelines for trustworthy AI and others.

In order to propose an operational AI certification framework, we have to ground it to operational definition of AI system, select appropriate set of technical standards to check scientists' work and define API backdoor for testing and monitoring of AI systems, which should all be set as a standard.

One of the OECD recommendations includes a revised outline document for practical guidance for the Recommendation of the Council on Artificial Intelligence [4], which provides common AI definition, and which we will be using as the basis for describing AI systems, shown in Figure 1.

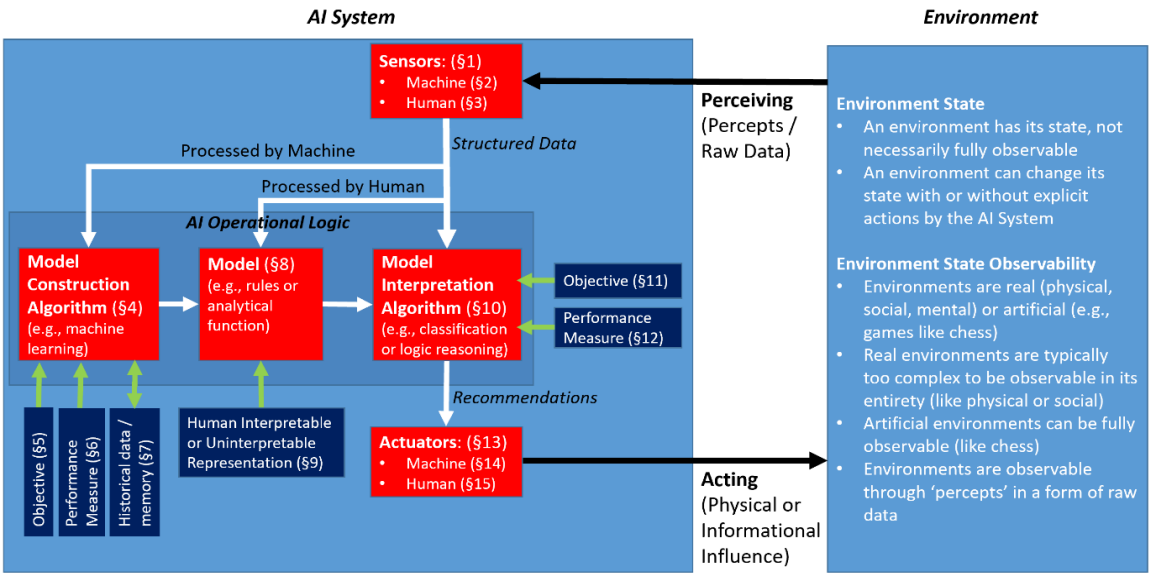

Figure 1 - AI System as defined by the OECD

"An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. It does so by utilising machine and/or human-based inputs to:

i) perceive and/or analyse real and/or virtual environments;
ii) *abstract such perceptions/analyses into models manually or automatically; and*
iii) *use model interpretations to formulate options for outcomes.*"

AI systems are designed to operate with varying levels of autonomy. Some AI systems are fully autonomous, but most of the AI systems can be placed on the large spectrum of semi-autonomy. Level of autonomy depends mostly on the domain and the problem itself, how the system is integrated in the business environment in particular business process and basically, how it interacts with humans. From the technical perspective, it affects the design of the system and the selection of machine learning methodology. Since AI is a wide domain, the role of AI systems can be also very different in terms of expected impact. AI systems can be autonomous, or offer certain source of additional information, or suggest decisions or just provide additional knowledge with extended exploratory functionalities.

Fully autonomous AI systems are systems to act on their own, independent of human intervention. Examples can be found mostly in the fields of robotics (specialized machines) or assistance (chat bots). However, most of the AI systems are not fully autonomous, but need certain degree of human intervention. In Fintech's domain, most of the AI systems or AI models are part of existing business processes and support decision making processes with additional information and/or suggested decisions. An example would be credit scoring model (behaviour and acquisition credit scoring), where a bank can decide to what extend the human intervention is needed. Credit scoring models provide not only information about whether an applicant is going to default, but also an assessment of associated risk. Based on this and business strategy, the bank can adjust the price accordingly. In this particular example, it is crucial how the model is incorporated in the bank's business process. It affects not only how the results of the AI models are used, but also how the models are trained. The other example would be AI models for cross and up-sell services. A client can get different offers (with different prices), based on historical and current behaviour, with the emphasis of the current status and micro-behaviour (for instance, web services), using external and internal information. In the case of flat-rate pricing a bank can use the outputs of the model directly. In case of adjusted pricing, AI models have greater impact and must be reviewed in more detail.

How the AI models are integrated and used in the environment (business or social) determines the impact, it therefore weights the error of the system. For instance, if the weather forecast is not very reliable, the impact for common individual is fairly minimal, while if the case of usage for predicting weather condition data for maximizing crops and therefore adjusted pricing strategy, the impact can be much larger. In another scenario, if recommendation system is suggesting that a set of articles for common reader is not very accurate and is not very crucial, while cross-sell recommendation model predicts client behaviour in wrong way, the client is deprived. These types of possible biases can be described as activation biases, where the outputs of the AI systems measure the impact in the environment: activation biases for non-fully autonomous AI systems are highly dependent on human intervention (when human are actuators).

Activation bias is one of the three major groups of biases, defined in OECD AI definition, as shown in Figure 2. The other two are Perception and Technical bias.

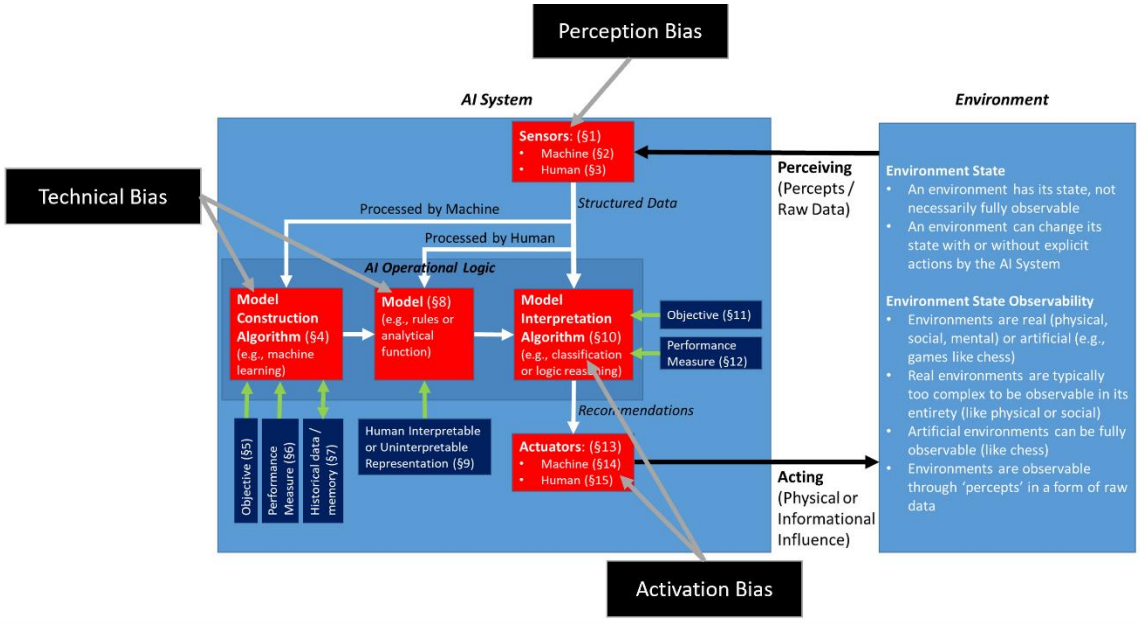

Figure 2 - Types of biases of AI system

Perception biases combine different types of biases, which are initiated by "error" in perception of the environment. AI systems percept environment thru data. And data is never quite accurate. Some errors in the data can be caused by poor data collection and acquisition, some error can be caused by faulty sensors. We can divide perception biases in two main groups:

- data quality issues – which can be managed by using data quality measures and
- conceptual issues, which are more focused on whether the data is "fair".

Fairness of the data cannot be directly measured yet, since there are not valid definitions. However, this issue is starting to raise up in the last few years. The definition of fairness is currently mainly based on human judgement and is not strictly measured. One example would be for instance, if the dataset consists of 65% male and 35% female candidates, one would conclude that the gender distribution of this particular dataset is biased. We can do the same reasoning on some other variables like nationality, geographical data, race and similar features. There is no simple answer to how to deal with different perception biases or even how to detect them, what is "fair" data.

The third group of biases are referred to as Technical biases. They are related to particular properties of machine learning algorithms. Each family of ML algorithms has certain properties, by the design, which can cause certain biases. For instance, decision trees prefer categorical variables with larger number of values.

Another important feature of AI systems, is the way how models are trained. Conventionally trained AI models have two main stages: a training phase and an operation's phase, when a model is put in a production environment. In the last few years, we can also use dynamic AI models, which are constantly learning models. We can describe dynamic models as a conventionally trained AI model, which is constantly retrained. With each new data example, we retrain the model. While for conventionally trained AI models we can use set of standard key performance measures, for dynamic models' performance measures are not straightforward, especially if we want to compare them.

Framework development

We can conclude that AI systems are quite diverse and fairly complex systems, which cannot be observed as standalone systems, but have to be considered as part of the whole environment. For the moment there exist more open questions than answers. A number of international high-level institutions are trying to set up a consistent framework which will enable much needed AI certification.

In the scope of INFINITECH, and in cooperation with other Horizon 2020 projects, i.e. DataBench (GA No. 780966) and FinTech (GA No. 825215), we will contribute to these activities by proposing an AI certification framework that will be tested on real-life examples, including at least one from the Fintech domain.

Defining AI certification framework is a moderately complex task; thus, our development will be focused on two trustworthy principles (proposed by EBA). They could be measured objectively, using standard machine learning metrics:

- Fairness and avoidance of bias
- Traceability and auditability

The proposed framework will include a software tool, which will enable automatic and semi-automatic monitoring of AI systems and provide several standard machine learning metrics, which describes the AI model.

However, since AI certification is a complex process, some of the information needs to be manually reviewed, (using relevant documentation of an AI model). Therefore, the proposed AI certification framework will consist of three different review modes:

- Manual review mode:
  - where the main goal would be detection of possible faults in model construction,
  - manual review is performed by an expert, going through required documentation of an AI model
- Semi-automatic mode where AI certification software can test and monitor actual performance of AI models independently:
  - In Semi-automatic mode, a reviewer would use sw tool in various test scenarios
  - The main goal would be an independent test of performance measures with comparison to the reported performance measures.
- Online monitoring review (automatic mode) for real-life observation to monitor behaviour:
  - In online monitoring mode, a sw tool will be used to monitor incoming data and AI model performance.
  - The main goal would be on-line monitoring of model performance with comparison to reported measures.

In Table 2 below, we listed main activities for certification review of AI systems with respect to review modes and the type of the observed correctness.

Table 2 AI Certification Review – Main Activities

AI model		**Pre-requisites**	**Review method**		
			Manual review	**Semi-automatic (offline)**	**Automatic**
Core	**AI model works correctly (according to the claimed description)**	AI model report ● Methodology ● Technical architecture ● Deployment report ● Performance reports (	Review of ML methodology: ● ML process flow (description of methodological steps, exclusions, filters and outliers) ● Performance report	● Analysis of training, validation and test datasets (distribution analysis of input variables) ● Test Performance measures (selected and reported performance measures)	● Comparison analysis of predicted vs actual performance ● Comparison of distribution of input variables thru time

c t n e s s		Measurable KPIs) Available data sets Available API – observed AI system enables running experiments	• How AI system is integrated in the environment • How the results of the AI system are used (in which part of the business process)	○ Execution of test experiments ○ Test on selected examples • Test for robustness • Analysis of input's distribution vs model's outputs	
	AI model is compatible with Value standards (ethics, fairness)	• Set of value standards, expressed in measurable KPIs	• Assessment of data sets according to data privacy and security standards (GDPS, human rights)	• Analysis of data distributions (detection of various general biases e.g, gender, nationality, age) • Comparison to benchmark dataset	• Analysis of data distribution with respect to value standard

Future work

Future work will include actual development of Framework for AI certification. We will develop a prototype certification system to allow (semi)automatic verification of AI Systems. It will enable monitoring of AI systems and provide several standard machine learning metrics for observed AI model. During the development the following questions will be addressed:

- Which classes of AI Systems are possible to certify?
- Definition and agreement on hard and soft KPIs describing an AI System
- Proposal for standardize backdoor API for testing and monitoring

3.3 Focus in the project

Within the INFINITECH project it shall be considered, that Financial Services providers usually fulfil common standards as the ISO 27000 series, PCI DSS etc. and utilize Information Security Management Systems. These are common technical prerequisites today and should be assumed in typical IT operations.

Therefore, the focus of the project shall be on regulations with respect to the INFINITECH pilots and their scenarios. As there are

- GDPR – The utilization of personal data as part of BigData applications and AI or Analytics based evaluation is core for smart, autonomous and personalized services. Thus, the GDPR will be the absolute focus of this task.
- PSD II – The PSD II opens the payment services market and introduces a collaborative payment landscape of many players beyond the banks. Thus, it influences the INFINITECH project considering BigData services and facilitating SHARP services providers. Therefore, the applicability and high level requirements related to the PSD II have been assessed.
- MiFiD II – This regulation is targeting the protection of investors and harmonizes the application of oversight. It imposes more reporting requirements and tests in order to increase transparency in trading. With several pilots related to investments, this regulation is worth to be in the focus of this task.
- 4AML – Several pilots in INFINITECH are working on Fraud and Financial Crime. Therefore, anti-money-laundering endeavours are one of the core parts of the project and the regulation is highly relevant to be considered.

In the following section 4 the high-level requirements resulting from these four regulations are outlined.

4 High Level Requirements

4.1 GDPR

4.1.1 Background

The GDPR [5] and data protection laws of the Member States aim at protecting the fundamental rights of natural persons in relation to the processing of personal data. Article 8(1) of the Charta of Fundamental Rights of the European Union and Art. 16(1) of the Treaty on the Functioning of the European Union provide that everyone has the right to the protection of personal data concerning him or her [6].

With INFINITECH being a data driven project, several services will make use of personal data in terms of the GDPR.

The legal assessment of the applicability of the GDPR rests on the information lined out in the INFINITECH D2.1 User Stories and Stakeholders' Requirements I, the INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, as well as the information given by the pilots in response to legal guidelines issued to them.

In addition, an ethics assessment focusing on the compliance with requirements of the GDPR was concluded jointly with the pilots. The respective tables are attached as Appendix C.

4.1.2 Applicability of the GDPR

As a prerequisite of the applicability of the GDRP, T2.4 has evaluated if the processing of personal data in terms of Art. 4(1) GDPR is subject to the pilots.

These e "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.*"

The term "*processing means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movement.*"

The scope of this deliverable D2.7 is restricted to the use during the project phase of INFINITECH. Due to choices made by the pilots at the current stage of the project, the GDPR will not apply to most of them. This assessment will, as a matter of fact change with regard to some pilots if and when they decide to process real data or commercially exploit the services developed in INFINITECH.

Against the current background of information provided by the pilots, the GDPR is only applicable to some of the pilots. Thereby, T 2.4 has relied on the information described in section 4.1.1 above.

Most pilots have or will implement anonymization methods at source. Therefore, the information used within the current scope of INFINITECH is not subject to the GDPR. Further information on anonymization under the GDPR are included in section 5.2.1.

4.1.2.1 Pilot #1

Pilot #1 is titled "Invoices Processing Platform for a more Sustainable Banking Industry" and belongs to the "Smart, Reliable and Accurate Risk Assessment" category [7]. It offers a data-intensive system to extract information automatically from notary invoices. It extracts from the invoices tables with amounts and their values from the invoices.

The objective of the POC is to process notarial invoices with artificial intelligence algorithms to extract the table with amounts within the invoice document, extract the concept total amount of the invoice and its value, and extract the concepts simple copy and authorized copy along with the number of copies and amounts of each of the concepts.

These invoices contain the name of the notaries as well as the VAT number. As these are related to the name of the notary as a legal person only, they will not be considered as personal data under the GDPR. Neither will personal data of customers of the bank appear on the invoices. Moreover, the result of the processing activity does not contain personal data.

However, this might change if the invoice does stem from a notary that is not a legal person but a group of individuals and/or the invoice does contain names or contact details of individual persons. This is not within the scope of the project at the moment however.

Following the DPO of the pilot responsible there is no processing of personal data, the only data that appear are those of the notaries who issue the invoices.

4.1.2.2 Pilot #2

Pilot #2 offers "Real-time risk assessment in Investment Banking" and belongs to the "Smart, Reliable and Accurate Risk Assessment**"** category [8]. The pilot provides risk-assessment analytics on the fly for bank traders, risk managers and sales negotiators based upon Value-at-Risk and Expected Shortfall procedures. It estimates market risks and pre-trade risks thus facilitating decision making processes for traders.

Since the pilot is engaged with financial markets data rather than personal data of individuals, the GDPR will not be applicable to the service.

4.1.2.3 Pilot #3

Pilot #3 evaluates how customer, account and transaction data is shared and analysed between banks and Fintechs using APIs to support customer-centric data services. This also involves customer-centric services for data acquisition and collaborative data sharing [9].

As a consequence, the pilot would rely on a wide number of personal (customer) data, whose aggregation, combination and analysis would involve several implications imposed by the GDPR.

However, currently the pilot is only concerned with the sharing of consent declarations between financial institutions using Blockchain technologies. Therefore, the pilot will initially use machine created synthetic data only that do not enable drawing conclusions on any natural persons. Therefore, the GDPR is currently not applicable.

4.1.2.4 Pilot #4

Pilot #4 is titled "Personalized Portfolio Management – Mechanism for AI based Portfolio Construction" and belongs to the "Personalized Retail and Investment Banking Services" [10].

The pilot focuses on the possibilities of AI Based Portfolio construction for Wealth Management and relies mainly on data from public newsfeeds and combines these with investment targets and financial indices which are public. Therefore, although the outcome of the pilot is aimed at customers, the pilot itself does not make use of personal data.

According to the partner Reportbrain the data source is 300+ publicly available newsfeeds in 35 languages. Due to the extremely high cost of cloud/servers, data is only collected upon request of a customer or use case on certain parameters defined by the customer. The same will be done during infinitech Pilots #4 and #6, upon receiving the required financial indices to follow by the Pilot Technical

Leaders. Hence, no previously collected Data will be used, nor any private data are gathered as newsfeeds are by definition public.

4.1.2.5 Pilot #5a

This pilot's description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.1.2.6 Pilot #5b

The pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.1.2.7 Pilot #6

Pilot #6 deals with **"**Personalized Closed-Loop Investment Portfolio Management for Retail Customers" and belongs to the "Personalized Retail and Investment Banking Services" category [13]. The pilot will focus on creating customer profiles as well as enhancing customer-service and portfolio management. As a matter of fact, the pilot will – if real data would be used – process a wide number of personal (customer) data and create customer related profiles. This could be considered as a high-risk activity from a data protection point of view.

However, the pilot will anonymize personal data at this stage of the project. Therefore, the GDPR does currently not apply to the respective INFINITECH service.

For the validation workshop mentioned in the DoA, NBG intends to utilize volunteers from NBG's personnel (as they are already bank's customers and many of them have investor profile) and for this purpose colleagues from several areas of the bank will be invited. All volunteers will provide their consent for the usage of their data, as an extension to the already GDPR terms and conditions already available for all NBG's customers (including also employees). The workshop's results will contribute to the determination of a credible path with a view to the deployment in production.

4.1.2.8 Pilot #7

The pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.1.2.9 Pilot #8

The pilot description is restricted to the consortium. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.1.2.10 Pilot #9

Pilot #9 is titled "Analysing Blockchain Transaction Graphs for Fraudulent Activities" under the "Financial Crime and Fraud Detection category" [16]. The pilot makes use of public Bitcoin and Ethereum Blockchain data in order to assess if the assets issued on the public Blockchain have been acquired using legal means and/or if the addresses on the Blockchain relate to a blacklist. The pilot will also be able to read reports on the analysis carried out by the pilot. Therefore, the pilot relies on the full historical transaction data containing Blockchain addresses.

As long as the addresses on the Blockchain cannot be related to individual persons, the pilot would not use personal data under the GDPR. However, this will depend on the internal structure within the

partners' organisations. Besides, the aim of the pilot in itself is drawing conclusions on the legitimacy of assets of single persons.

The pilot uses randomly chosen Blockchain addresses from publicly available bock explorer websites at the initial stages of the project. It is assumed that the persons behind these addresses are not identifiable and thus only anonymized data are used in the pilot.

4.1.2.11 Pilot #10

Pilot #10 is titled "Real-time cybersecurity analytics on Financial Transactions' BigData" and belongs to the "Predictive Financial Crime and Fraud Detection" category [17]. The outcome will be a tool that monitors in real-time the financial transactions of a domestic mobile banking system using BigData analytics technologies.

It will make use of activities of a customer on their bank account. These include e.g. bank account numbers, and transactions of individuals. Such can in general constitute personal data. Nevertheless, the pilot will only use synthetic data that does not originate from individual persons, but is created by a machine. Therefore, the GDPR does currently not apply to the pilot.

Still, the pilot is preparing for the use of real data taking into account e.g. pseudonymization methods on the momentarily synthetic data.

4.1.2.12 Pilot #11

Pilot #11 develops **"**Personalized insurance products based on IoT connected vehicles" and belongs to the "Personalized Usage-Based Insurance Pilots" category. It focuses on data-based risk assessment and pricing services, including also fraud detection mechanisms [18]. The pilot analyses and takes into account an insured person's daily vehicle driving behaviour taken in order to assess a risk-orientated price of a vehicle insurance. Depending on the result of the analysis the insurance price will rise or fall.

Besides, the pilot will collect information from connected cars in order to analyse traffic incidents and enable fraud detection, as well as roadside assistance. The data used in the pilot are e.g. location data, speed, acceleration forces. Although some of these data contain information on the status of a vehicle, they can be used for analysing the driving behaviour of a natural person and thus constitute personal data under the GDPR.

Besides, within the project phase, the pilot will collect data on a wide number of connected vehicles from their partner's infrastructures in order to create driver profiles that will later serve as a reference for the analysis functions. These will consist of real data gained from IoT devices on the one hand and from simulated smart objects on the other hand.

For the real data gained from connected vehicles the pilot will receive the consent of the respective data subjects in order to address the requirements of the GDPPR. The synthetic data used in the project, however is not subject to the GDPR.

In later stages, or within a commercial operation phase, the applicability of the GDPR will also depend on the relation between the insurance companies and the service offered by the pilots.

4.1.2.13 Pilot #12

Pilot#12 offers "Real World Data for Novel Health-Insurance Products" and belongs to the "Personalized Usage-Based Insurance Pilots**"** category [19]. The Pilot focuses on health- and behaviour-related data collection and analysis to offer to customers of health insurances risk-based personalized insurance offers.

This involves e.g. the collection of vital signs, physical activity and subjective data reposted by voluntary users on their quality of life and nutrition. Therefrom, the pilot will create risk clusters of people that the risk model will be based upon.

The data – that also involves health data under Art. 9(1) GDPR - originates from real persons and are gained from Healthentia data and insurance companies. Within the project, volunteers will provide their (health) data, filling in informed consent forms. These types of data will fall under the GDPR. The risk cluster developed within the project will however most likely be a purely synthetic one provided that enough volunteers provide their input within the project and the model is being pre-trained using synthetic data.

In the project phase, real personal data gained for the purpose of the pilot will be anonymized as lined out in the respective Ethics table. However, the real data sources are being combined before the anonymization process takes place. It will depend on where and how this combination takes place within the infrastructure.

Therefore, at least the information provided by volunteers will constitute personal data under the GDPR irrespective of the methods used for combining and anonymizing later on.

The applicability of the GDPR might be limited to the project phase however. The service offered later on to insurance companies could be designed in a way using anonymized data only, although an insurance company as an end user will still be subject to the GDPR.

4.1.2.14 Pilot #13

Pilot #13 is titled **"**Alternative/automated insurance risk selection- product recommendation for SME" and belongs to the "Configurable and Personalized Insurance Products" category [20]. It focuses on developing an insurance product configuration platform for SMEs. Therefore, it will use data on legal persons and entities only that do not fall under the Scope of Art. 4(1) No. 1 GDPR.

4.1.2.15 Pilot #14

Pilot #14 titled "Configurable and Personalized Insurance Products for SMEs and Agro-Insurance'' is generally aligned with the business process of usage-based insurance (UBI) [21]. It focuses on gathering data and combining information on agricultural fields/farms and thus help evaluating the risks for insurance companies and offers more personalized products. Pilot data involves e.g. geospatial data, weather/seasonal climate and remote earth observation data.

Since these data are related to farms and crops primarily, they do not constitute personal data on first sight. If however, through the insurance company, it can be combined with data of the farm owner as an individual, the risk score could be considered as personal data. Therefore, GDPR will apply within the insurance company as end-user of the pilot.

With regard to the pilot itself however, the pilot receives from the insurance company non-personal data mostly in order to carry out the evaluation. Through the geospatial data received however, the identification of an individual that owns the farm (if not a legal entity) might be possible by conducting research through e.g. Google Maps or researching official registries. Therefore, the GDPR will apply to the pilot nonetheless.

4.2 PSD II

The European Payment Service Directives' purpose is to regulate payment services and the issuance of e-money in order to protect users of such services from abuses by companies rendering such services as well as maintain the integrity of the payment service and e-money markets. Payment service providers as well as e-money issuers receive and may dispose of potentially large funds of their

clients. The risks associated with such business activities require supervisions and compliance with the code of conducts provided for in the national transformations of the Payment Service Directives.

Key terms of the Payment Service Directives are "payment service" and "e-money". The payment services subject to the Payment Service Directives are enumerated and comprise amongst others the receipt and disbursement of cash, the execution of payment transactions with or without lending to account holders, acquiring, and giving crediting to the respective account holders.

Electronic money – or e-money - is any monetary value stored electronically, including magnetically, in the form of a claim on the issuer which is issued against payment of an amount of money in order to facilitate payment transactions and which is accepted by natural or legal persons other than the issuer.

As a rule of thumb, any service rendered with regard to other people's (including companies') money and any form of issuing electronically (magnetically) readable devices storing values for other people (including companies) should be regarded as sensitive from a regulatory point of view.

4.2.1 Pilots with no evidence that PSDII applies

Pilots #1 - #4, #6 - #8-14 do not intend to deal with payment services or transactions.

4.2.2 Pilots with evidence of PSDII applicability and further need to inquire

The related pilot descriptions are confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.2.3 4.2.3 Pilots with clear evidence of PSDII applicability

The related pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.3 MiFiD II

The Market in Financial Instruments directives targets all services of financial instruments, predominantly securities, investment certificates and crypto assets. Such services include advice, brokerage, dealing, storage and financial analysis of financial instruments. The offering of financial instruments to fund own business is not covered by MIFID.

Pilot descriptions of the pilots #5a, #5b, and #7 are confidential, the description of pilot #8 is restricted to the consortium. Hence, these are not listed in the subsequent sections.

4.3.1 Pilots with no evidence that MIFIDII applies

Pilots #1, #3, #4, #9-#14 do not intend to pursue any activities related to financial instruments and/or maintain to be covered by the required license.

4.3.2 Pilots with evidence of MIFID II applicability and further need to inquire

Pilot #2 deals with financial analysis and maintains that it has conflict of interest declaration for each employee in place. Since there are more compliance requirements for distributors of investment research it is unclear whether Pilot #2 is fully compliant or not.

Pilot #4 did not give any information on MIFIDII activities. The scope of the pilot indicates, however, a certain proximity to investment advice. Compliance with MIFIDII should therefore be checked.

4.3.3 Pilots with clear evidence of MIFIDII applicability

Pilot #6 clearly is within the MIFIDII scope, but appears to be covered by a financial services license. It refers to its legal and compliance department, so that we assume that regulatory issues are covered in-house.

4.4 4AML

Money laundering and the financing of terrorisms has been one of the predominant objectives of financial regulation in the last two decades. All banking, financial service and insurance activities as well as a defined number of other business activities (the "obliged entities") attractive for money launderers are object of ever stricter regulations to stop money from criminal activities being "washed" through the financial system. At the heart of anti-money-laundering obligations are the identification and the validation of clients of these businesses. In addition, obliged entities need to provide risk management plans and other documentation to competent authorities.

Pilot descriptions of the pilots #5a, #5b, and #7 are confidential, the description of pilot #8 is restricted to the consortium. Hence, these are not listed in the subsequent sections.

4.4.1 Pilots with no KYC implications

Pilots #1 - #4 as well as #10-14 do not intend to pursue any activities related to KYC.

4.4.2 Pilots with KYC implications and further need to inquire

The pilots' descriptions related to this section are confidential. Thus, in case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.4.3 Pilots with clear evidence of KYC implications

Pilot #9 is within the KYC perimeter by design. Although it does not appear to be an "obliged entity" in its own right and thus does not carry any specific obligations familiarity with KYC should be essential. For the purpose of this paper, we assume that the required knowledge exists within the pilot's team.

5 GDPR in the context of SHARP services

Following the description of the pilots and the assessment in section 4.1 above the GDPR is applicable to some of the pilots. The respective processing activities will thus have to comply with the provisions contained therein.

This section 5 gives a first high-level evaluation of compliance. Moreover, it serves to illustrate the information asked for in D10.2 which is part of this deliverable. The deliverable focuses on the provisions of the GDPR only, not taking into account the respective national laws on data protection containing potentially deviant provisions to some extent.

5.1 Ethics Requirements provided for in D10.2.

The Commission expects information on the processing of personal data and therefore the pilots provided the following information in Annex C:

- Applicability of the GDPR vs. anonymization, see Art. 4 Nr. 1 and Recital 26 of the GDPR
- Pseudonymization as a technical organizational measure under Art. 32 GDPR
- Data minimisation principle according to Art. 5(1) lit. c GDPR
- Processing of special categories of personal data in terms of Art 9(1) GDPR
- Principle of purpose limitation, see Art. 5(1) lit. b)

- Transferring of personal data to recipients and/or thirds as a means of processing, see Art. 4 No. 2 GDPR
- Principle of lawfulness of processing pursuant to Art. 5(1) lit. a) GDPR, see also Art. 6 and 9 GDPR
- Principle of integrity and confidentiality according to Art. 5(1) lit. e) GDPR, see also Art. 32 GDPR.
- The principle of accountability as required in Art. 5(2) GDPR.
- Transfer of personal data to third (non-EU) -countries, see Art. 44 – 50 GDPR.

5.2 Specific aspects in the context of SHARP services

5.2.1 Pseudonymization and Anonymization

Most pilots that deal with essentially personal data at the project stage have decided to use means of anonymization or pseudonymization. This serves to minimize the impact of processing on the rights and freedoms of natural persons on the one hand and leads to the (partial) exclusion of the GDPR on the other hand. For fully anonymized data, the GDPR does not apply at all. Therefore, such data will not constitute personal data. On the contrary, pseudonymised data is governed by the GDPR.

These steps are even taken by some pilots that currently use purely synthetic non-personal data but would like to prepare for a subsequent use of real personal data either at later stages of the project or for a commercial operation.

Art. 4 No. 5 GDPR defines **pseudonymization** as

> *"the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."*

The crucial aspect here is that the natural person is in fact identifiable at all, taking into account additional means reasonably available to the data controller or third persons. While still constituting personal data, pseudonymization is a means to comply with adequate technical and organizational measures required by Art. 32 GDPR, if the "missing" information necessary for identification is subject to certain measures preventing the identification.

For **anonymization**, the GDPR neither offers a binding definition, nor a straightforward generally binding approach. Anonymization will, in contrast to mere pseudonymization, only be given where the natural person behind that data cannot be identified at all, subject to means reasonably available to the data controller (pilot) or third parties. This is apparent from Recital 26 of the GDPR stating:

> *"The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymization, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.* ***The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.***

This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes."

While "reasonable means" for identification may not be available as of today, this might change in the future with AI offering more and more opportunity of identification of apparently not connected data.

In any case, assessing reasonable means includes "taking into account all objective factors". A purely hypothetical opportunity to identify the natural person will thus not exclude anonymization. The same applies to means which would practically not be feasible e.g. because they require an inadequate input of time, cost or workforce or are legally not permitted. Although the GDPR does not specify the technical means necessary in order to achieve anonymization, such methods would have to be state of the art.

The question as to whether anonymization methods chosen by the pilots do comply with these requirements of anonymization in a legal sense will be subject to further scrutiny, in particular with regard to T3.5 and T3.6. T3.5 will also provide a list of data governance technologies.

However, given the opportunities of AI and big data within the project as well as the amount of data potentially available to project partners and the fact that they act within the financial sector, this assessment will be a rather technical one.

In particular, with regard to some pilots it will be necessary to anonymize data at source level in order to avoid processing personal data at all.

5.2.2 Assessment of Privacy Impact Assessment methodologies

Within D10.2 the Commission is expecting an assessment of different Data Protection Impact Assessment methodologies and tools that can be later adapted for the specific industrial domains addressed by the project, also considering pilot scenarios where such DPIA reports should be produced.

5.2.2.1 Background

A Data Privacy Impact Assessment (DPIA) is required to the extent that the provisions contained in Art. 35 GDPR do apply to a processing activity. The duty to carry out such a DPIA aims at the controller of personal data in first place. At the current stage of the project this primarily does apply to the respective pilot making the decision to use certain dataset(s) of personal data for the project purpose defined by the pilot. If used commercially, this situation could however be turned the other way around. Then, the pilot's customers will in most cases decide upon the purpose and means of processing personal data. Rendering them the controller in terms of the GDPR, the duty to perform a DPIA is upon them then, whereas the pilot offering his services will in most cases take the role of a mere processor of data, see Art. 28 GDPR.

5.2.2.2 Methodologies

The basic methodology and requirements of conducting a DPIA are lined out in Art. 35 of the GDPR itself. This includes e.g. the consultation of the controller's DPO, the duty to carry out the DPIA before the start of the respective processing activity, as well as the assessment provided for in para. 7 of the provision.

The methodology has been specified in particular by the Article 29 Data Protection Working Party in its WP 248 rev.01 "Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is likely to result in a high risk for the purposes of Regulation 2016/679" adopted on April 4th 2017 and last revised on October 4th 2017.

These guidelines were also used by the CNIL (French Data Protection Authority) for its three guides on conducting a DPIA which are also consistent with risk management international standards. These guides are retrievable via https://www.cnil.fr/en/PIA-privacy-impact-assessment-en. Based upon these guides, the CNIL has made available the so-called "PIA software" that assists in carrying out and documenting a DPIA. Further information on the software can be found via

https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment .

The method propagated by the CNIL is from our impression turning out to become state of the art and should be the preferred one when carrying out a DPIA within INFINITECH. This is especially true as it can be accessed publicly and would also be available to the users of INFINITECH services. The CNIL's methodology is described briefly in the CNIL Infographic [22].

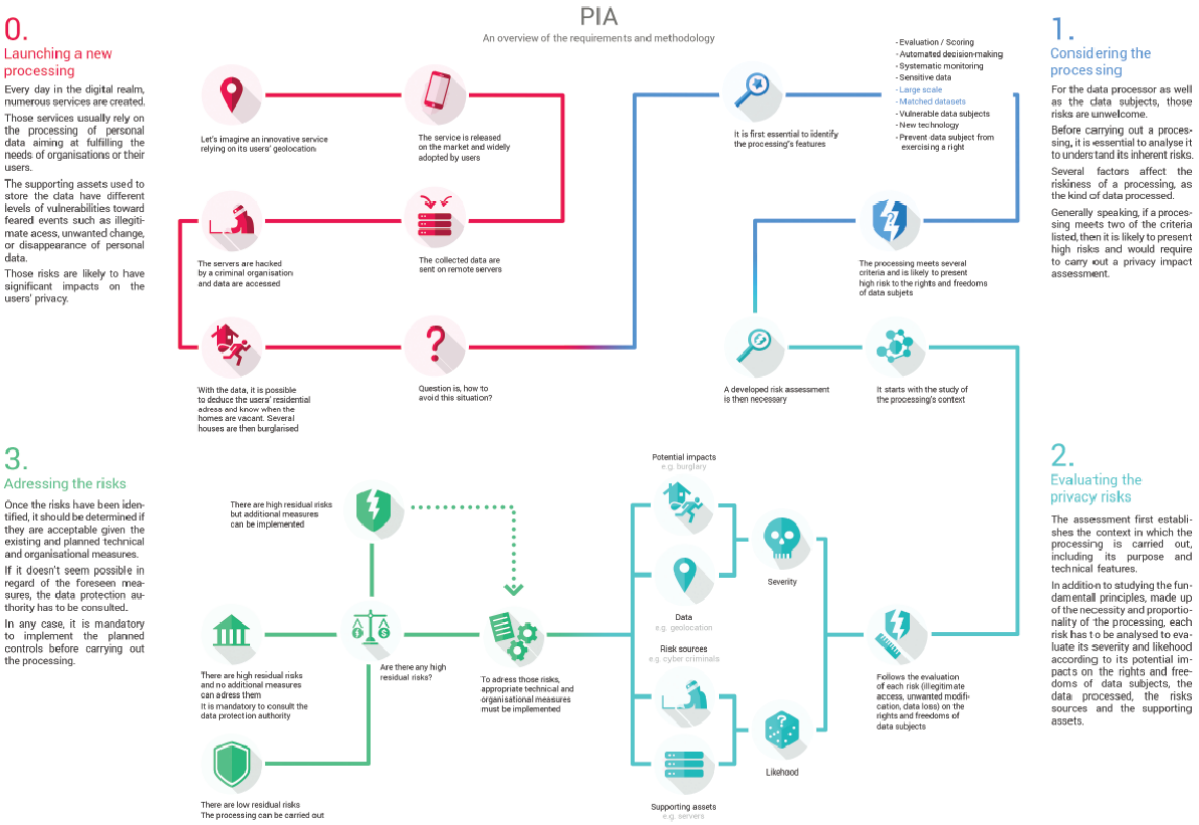

Figure 3 CNIL Methodology

The WP 248 guidelines have also been adopted by the European Data Protection Board, without giving guidance as detailed as done by the CNIL. Besides, several national data protection supervisory authorities have provided explanations, checklists and guidelines. There are also other tools on the market helping in carrying out a DPIA.

5.2.2.3 Duty to perform a DPIA

The controller is obliged to perform a DPIA when the processing activity is likely to result in a high risk to the rights and freedoms of the data subject. This could be the case with regard to pilot #5a performing scoring activities, pilot #6 evaluating the behaviour of individuals, pilot #11 making use of tracking technologies, as well as pilot #14 relying on health data. Since pilot #12 will offer the means for scoring activities to potential customers, the duty to carry out the DPIA is upon the latter. But within the project, pilot #12 will also rely on health data in terms of Art. 9(1) GDPR. Depending on the number of volunteers and the scope of data used, this could result in the duty to carry out a DPIA too.

The duty to perform a DPIA is subject to the requirement of the use of real personal data. This is not necessarily the case at the moment with regard to the pilots mentioned.

5.2.3 Processing activities within INFINITECH – data sources and processing chains, lawfulness of processing activities

As apparent from sec. 5.1 and the corresponding ethics requirements tables the pilot will partially make use of personal data already collected within their organisation, whereas others rely on data provided by third parties or made available to them by the data subjects for the scope of the project.

In case of real personal data being collected for the purpose of INFINITECH initially by the pilots, this should take place based on the consent of data subjects pursuant to Art. 6(1) lit. a) GDPR. Although this consent could potentially contain the opportunity to share such data with other pilots/partners, if shared with other pilots, these should be subject to anonymization or pseudonymization, depending on the adequacy of measures to protect the rights and freedoms of the data subject. Besides the consent itself, the data subjects have to be provided with the information pursuant to Art. 13 GDPR in the form of privacy policies or privacy notices.

Where personal data is being provided to the pilots by third parties, the legal basis for the transferring is initially depending on the boundaries of the purpose of processing and the legal basis made use of by the third party. Where such third party as a controller of personal data decides to transfer the data to INFINITECH pilots for the purpose of the project, it is in general, but not exclusively, advisable to use such data in an anonymized manner. This for the reason that even a consent given by the data subject to the third party is unlikely to involve the transferring data to a pilot who determines by themselves the purposes of processing. Besides, the parties involved would potentially have to conclude a contract on the joint controllership pursuant to Art. 26 GDPR.

Some pilots plan to use personal data available to their respective organisations already. The use for the purposes of INFINITECH and/or the purposes of scientific research will depend on the legal basis and the purpose of processing that the personal data has been collected for by that organisation. If the use for the purpose is not subject to a consent of the data subject, such change of purpose will only be possible in accordance with Art.6(4), 9 in connection with Recital 50 of the GDPR. In most cases, the use of such data should, given the current information provided by the pilots, be subject to an anonymization at source.

If and to the extent that personal data is being shared between pilots/partners in order to carry out specific tasks, this should be done on the basis of a controller-processor contract according to Art. 28 GDPR.

5.2.4 Automated individual decision-making including profiling

Profiling is being defined as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects of concerning that person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements", see Art. 4 No. 4 GDPR.

This can be the case – in particular – with regard to the pilots that were being named within section 5.2.2.3.

Art. 22(1) of the GDPR states that a natural person shall have the right not be subject to a decision based solely on the automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

This is according to (2) however not the case where the exceptions lined out there-in apply.

It should be noticed that the processing of personal data within the current phase of the project is not directed at making a decision on the natural persons. Rather, personal data is being used for the validation of the systems created within INFINITECH. A more detailed assessment will be reserved for subsequent deliverables.

However, if automated decision-making and profiling will be carried out, it will be – depending on the case - upon the end-user or the pilot to ensure

- a consent of the data subject
- that the decision is necessary is for entering into, or the performance of, a contract between the data subject and a data controller or
- the process is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests

Moreover, in the cases of consent and performance of a contract, according to (3) the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, **at least the right to obtain human intervention on the part of the controller**, to express his or her point of view and to contest the decision.

Besides, with reference to the use of special categories of personal data, according to Art. 9(1) GDPR, this will be allowed either on the basis of a consent of the data subject either or in cases of Art. 9(2) lit. g). Suitable measures to safeguard to the data subject's rights and freedoms and legitimate interests have to be in place.

In addition, national data protection regulation is likely to contain additional provisions for profiling and scoring, e.g. section 31 of the German Federal Data Protection Act, containing provisions for the process underlying the scoring.

6 Conclusions

In this document analyses the **regulations**, which are in focus of the INFINITECH project with regards to the pilots use cases.

Specific emphasis is put on the **GDPR** due to its high relevance in BigData and analytics scenarios which apply in the INFINITECH's sharp services. This includes a detailed description of GDPR related requirements in the INFINITECH SHARP services and an ethical assessment per pilot.

Moreover, latest guidelines **on AI and its application in the Financial Sector** are considered as a valuable cross cutting insight for the overall project. Based on this overview on the emerging AI related requirements the emerging role of AI guidelines and regulations will need attention during the project and the next phase of task T2.4.

Relevant Financial Sector regulations as PSD II, MiFiD II and 4AML are assessed with respect to the INFINITECH pilot scenarios. This way an overview of the applicability of these regulations in the pilots and related high-level requirements is outlined.

At this stage of the project the most relevant requirement seems generating the awareness of the pilots regarding personal data and the details on how to deal with the various requirements resulting from the GDPR.

This results in a non- typical development related technical of functional requirements. However, there are certainly general data protection requirements, which can be obtained from this exercise:

- Data shall be checked in detail, if they have to be considered as personal data. This is underestimated in general.
- Anonymization and pseudonymization are key technologies within the INFINITECH project.
- Technical and organizational measures for data protection shall be considered in the pilots.
- Data Privacy Impact Assessments and Informed Consent will play an important role in the pilots.
- Technical requirements considering the new AI related guidelines will need further attention and elaboration during the project.

Appendix A: Literature

[1] UN AHEG (2020), First version of a draft text of a recommendation on the ethics of Artificial Intelligence, SHS/BIO/AHEG-AI/2020/4, limited distribution

[2] European Commission (2020) On Artificial Intelligence - A European approach to excellence and trust (White paper), COM(2020) 65 final, available at https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[3] European Banking Authority (2020), EBA report on Big Data and Advanced Analytics, EBA/REP/2020/01, available at https://eba.europa.eu/sites/default/documents/files/document_library//Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf

[4] OECD (2019), "Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)", *OECD Digital Economy Papers*, No. 291, OECD Publishing, Paris, https://doi.org/10.1787/d62f618a-en

[5] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[6] GDPR Recital (1)

[7] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.1.

[8] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.2.

[9] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.3.

[10] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.4.

[11] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.5.1.

[12] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.5.2.

[13] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.6

[14] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.7.

[15] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.8.

[16] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.9.

[17] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.10

[18] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.11

[19] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.12

[20] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.13

[21] INFINITECH D2.3 – Reference Scenarios and Use Cases – Version I, Sec. 4.14

[22] CNIL, Infographic on DPIA methodology, retrieved on Mai 13th 2020 via https://www.cnil.fr/sites/default/files/atoms/files/171019_fiche_risque_en_cmjk.pdf

Appendix B: Requirements List

At this stage of the project the most relevant requirement seems generating the awareness of the pilots regarding personal data and the details on how to deal with the various requirements resulting from the GDPR.

This results not in typical development related technical of functional requirements. However, there are certainly general data protection requirements, which can be obtained from this exercise:

- Data shall be checked in detail, if they have to be considered as personal data. This is underestimated in general.
- Anonymization and pseudonymization are key technologies within the INFINITECH project.
- Technical and organizational measures for data protection shall be considered in the pilots.
- Data Privacy Impact Assessments and Informed Consent will play an important role in the pilots.

Appendix C: Ethics Assessment – Ethics Tables

C.1 Pilot #1

Do you plan to collect and/or process personal data in your pilot operations?	NO. There is no processing of personal data.

C.2 Pilot #2

Do you plan to collect and/or process personal data in your pilot operations?	The Pilot #2, as it is currently defined, actually does not involve the use of any personal data and no anonymization is needed. Risk evaluation will be based on market price data and the signals from our algorithmic strategy portfolio and NOT on actual client orders or client accounts. On the other hand, on the textual data side, the pilot will rely on public data sources like news feeds and social media that will also need no anonymization.

C.3 Pilot #3

Do you plan to collect and/or process personal data in your pilot operations?	No use of personal data in Pilot 3. The approach is to manually create small data sets against a very clear semantic definition.

C.4 Pilot #4 (answered by RB)

Do you plan to collect and/or process personal data in your pilot operations?	NO. Not applicable. Reportbrain's Knowledge Equity and Sentiment Analysis AI Engine, only gathers publicly available newsfeeds.
Overview of the different non-fully anonymous datasets that will be used in the project	The source is 300+ publicly available newsfeeds in 35 languages. Due to the extremely high cost of cloud/servers, data is only collected upon request of a customer or use case on certain parameters. The same will be done during infinitech Pilots #4 and #6, upon receiving the required financial indices to follow by the Pilot Technical Leaders. Hence, no previously collected Data will be used.

C.5a Pilot #5a

This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

C.5b Pilot#5b

This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

C.6 Pilot 6 (answered by RB)

Do you plan to collect and/or process personal data in your pilot operations?	NO. Not applicable. Reportbrain's Knowledge Equity and Sentiment Analysis AI Engine, only gathers publicly available newsfeeds.
Overview of the different non-fully anonymous datasets that will be used in the project	The source is 300+ publicly available newsfeeds in 35 languages. Due to the extremely high cost of cloud/servers, data is only collected upon request of a customer or use case on certain parameters. The same will be done during infinitech Pilots #4 and #6, upon receiving the required financial indices to follow by the Pilot Technical Leaders. Hence, no previously collected Data will be used.

C.7 Pilot 7

This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

C.8 Pilot 8

This pilot description is restricted to the consortium. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

C.9 Pilot 9

Do you plan to collect and/or process personal data in your pilot operations?	NO
Please explain how all of the personal data you intend to process is relevant and limited to the purposes of the research project (in accordance with the "data minimisation principle")	Personal data will not be used. The whole big public Blockchain data and randomly chosen Blockchain addresses from publicly available block explorer web sites (https://etherscan.io/ and https://www.blockchain.com/explorer) will be used for the purposes of the research project. Addresses will not be linked to any person.
Anonymization of research data	
What research data will be anonymized/pseudonymised?	Anonymous bitcoin and Ethereum Blockchain data will be used.
What research data will not be anonymized/pseudonymised?	None
Please, provide details on the anonymization/ pseudonymization techniques that will be implemented in your pilot activities.	No, anonymization/ pseudonymization need to be carried out since anonymous public Blockchain data will be used.

Overview of the different non-fully anonymous datasets that will be used in the project	
Please, specify the source of the data and clarify if the data collection specifically occurs in relation to INFINITECH research or if these are previously collected data (i.e. collected for a purpose other than INFINITECH project, for instance for another EU project)	N/A
Please, provide information on the type of personal data and if they are common or sensitive data	N/A
Please, provide details on the specific purpose of data collection and processing in your pilot activities in relation to this dataset	N/A
Please, clarify if the intended recipients are INFINITECH partners and/or external entities.	N/A
Please, indicate the legal basis of the processing according to Art. 6 GDPR (see Table 2)	N/A
Please, provide information on the activities that you will implement for legal compliance, such as providing the volunteers with information, collecting informed consent and give them the opportunity to exercise their rights.	N/A
Please, provide information on the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing.	N/A
In case of further processing of previously collected personal data, please,	
Confirm that the beneficiary has lawful basis for the data processing	N/A since anonymous public Blockchain data will be used.
Confirm that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects	N/A since anonymous public Blockchain data will be used.
Clarify if personal data will be transferred to/from non-EU countries	*According to 5.1.6 DoA Part B:* No transfer or Personal Data is planned, neither from Turkey, Switzerland or Israel to the EU nor from EU to Turkey, Switzerland or Israel. In Turkey, the partners will comply with national laws and ethical mandates.

How will this comply with applicable data protection rules at EU and national level?	*According to 5.1.6 DoA Part B:* The process of managing INFINITECH data by partners outside the EU will be done in ways compliant to laws of at least one EU member states. Compliance for data protection given by INFINITECH in Europe will be extended to the whole financial network if using INFINITECH outside Europe.

C.10 Pilot 10

Do you plan to collect and/or process personal data in your pilot operations?	NO, because the datasets we will work on are fully synthetic ones.
Please explain how all of the personal data you intend to process is relevant and limited to the purposes of the research project (in accordance with the "data minimisation principle")	The datasets will be used to detect signs of fraudulent activities by analysing financial transactions. Anyway, transactions will be fully synthetic, therefore, non-real identities and non-real personal data will be managed.
Anonymization of research data	
What research data will be anonymized/pseudonymised?	Even though the datasets will be fully synthetic, the pilot will have the technical capabilities to deal with personal data. Therefore Pilot 10 will integrate and make use of INFINITECH pseudonymization technologies capable to protect bank customers' identities (e.g. name, surname, social security number, fiscal code, etc.) and details related to bank accounts. Anyway NO real personal data will ever be treated in Pilot 10.
What research data will not be anonymized/pseudonymised? Please, explain why.	Bank account holders' personal information and information related to their bank accounts and transactions (e.g. IBAN, Account number, etc.). The Pilot will be able to pseudonymize data in order to avoid them to be abused by Fraud Analysts (data minimisation) until it is really necessary to reveal customers' identity (potential fraud detected damaging the customer). These functionalities will be integrated in Pilot 10 but not used during INFINITECH project lifetime, as only fully synthetic dataset will be used.
Please, provide details on the anonymization/ pseudonymization techniques that will be implemented in your pilot activities.	Pseudonymization tools brought by INFINITECH's partners will be integrated and used in Pilot 10
Overview of the different non-fully anonymous datasets that will be used in the project	
Please specify the source of the data and clarify if the data collection specifically occurs in relation to INFINITECH research or if these are previously collected data (i.e. collected for a purpose other than INFINITECH project, for instance for another EU project)	Data will be created fully synthetic for the specific INFINITECH research activity
Please, provide information on the type of personal data and if they are common or sensitive data	Common data, non-sensitive. Dataset contains personal data of bank account holders and info related to their bank accounts and transactions, but the entire dataset is fully synthetic therefore there will not be personal info from real persons and accounts.

Please, provide details on the specific purpose of data collection and processing in your pilot activities in relation to this dataset	Pilot aims at identifying potential frauds in banking transactions. Once a potential fraud will be identified, it will be necessary to investigate further. The identity of the potential fraud's victims need to be known both for investigation purposes and for notification to customer and authorities
Please, clarify if the intended recipients are INFINITECH partners and/or external entities.	INFINITECH Partners
Please, indicate the legal basis of the processing according to Art. 6 GDPR	legitimate usage for fraud prevention and compliance (mandatory notifications to authorities when requested)
Please, provide information on the activities that you will implement for legal compliance, such as providing the volunteers with information, collecting informed consent and give them the opportunity to exercise their rights.	data pseudonymization
Please provide information on the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing.	data encryption, accountability
In case of further processing of previously collected personal data, please,	
Confirm that the beneficiary has lawful basis for the data processing	The beneficiary controlling the data in this pilot confirms that it has lawful basis for the data processing.
Confirm that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects	The beneficiary controlling the data in this pilot confirms that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.
Clarify if personal data will be transferred to/from non-EU countries	*According to 5.1.6 DoA Part B:* No transfer or Personal Data is planned, neither from Turkey, Switzerland or Israel to the EU nor from EU to Turkey, Switzerland or Israel. In Turkey, the partners will comply with national laws and ethical mandates.
How will this comply with applicable data protection rules at EU and national level?	*According to 5.1.6 DoA Part B:* The process of managing INFINITECH data by partners outside the EU will be done in ways compliant to laws of at least one EU member states. Compliance for data protection given by INFINITECH in Europe will be extended to the whole financial network if using INFINITECH outside Europe.

C.11 Pilot 11

Do you plan to collect and/or process personal data in your pilot operations?	YES

Please explain how all of the personal data you intend to process is relevant and limited to the purposes of the research project (in accordance with the "data minimisation principle")	Use Case 1 [Driving profiles] are based on the data collected from the connected cars (speed, acceleration, fuel consumption etc.) combined with the location of these vehicles. Vehicle's tracking to define routes is also essential, combined with drivers' profiles to classify the drivers as normal/aggressive (among others). One of the outcomes of the project's research process includes de identification of the vehicle's technical data subset that better defines the driving profiles. At this stage of the project, we will capture all available vehicle's technical data (current speed, average speed, acceleration, fuel consumption, emissions, etc.) that may vary, depending on the vehicle's manufacturer and the On Board Unit (OBU) built in that captures it, to find out the relevant correlations that defines the driving profiles. Once this subset is identified, the data capturing process will be restricted to it. Use Case 2 [Driver's Classification] requires location data to identify the area (e.g. city) where a driver drives according to an identified driving profile. Routes (location data plus the vehicle's technical dataset required to match a driving profile) from concrete drivers will be used to obtain the corresponding driving classification. This information (driving classification with no personal data) would be managed only by the insurance company and would be anonymized data or synthetic data.
Anonymization of research data	
What research data will be anonymized/pseudonymised?	The Pilot #11 research personal data (Use Case 1) includes all data collected from the (real) connected cars and captured by their corresponding On Board Unit (OBU) plus the location data provided by their GPS installed module. In this sense and particularly, the ID of the OBU/Connected Car will be pseudonymised. In Use Case 2, as only a rough vehicle's location is needed, this will be anonymized. DYN Data used to classify a given driver (data coming from the driver's car or from other source linked to the driver) will be completely anonymized or would be synthetic data
What research data will not be anonymized/pseudonymised? Please, explain why.	Technical data collected from the (real) connected cars, such speed, acceleration, fuel consumption, CO emissions, etc. won't be anonymized nor pseudonymised. This information will be related to a unique and concrete route, linked to a given and pseudonymised vehicle/OBU ID that changes from a route to another, hindering the matching between this technical dataset and a real driver. A route starts when the car engine is started and ends when this engine is stopped. Other research information (Weather, traffic data and synthetic connected cars won't be anonymized/pseudonymised as they're not considered as sensitive information.
Please, provide details on the anonymization/ pseudonymization techniques that will be implemented in your pilot activities.	Personal data from the pilot will be anonymized following a risk-based approach that will allow to determine automatically which techniques (anonymization operations) should be applied to each of the variables of the dataset. The anonymization process will be performed in several steps: 1) All the possible anonymization configurations (i.e. different combinations of generalization, randomization and deletion operations) for the dataset will be calculated.2) Taking into account the privacy requirements of the pilot and how the data will be used later on, a suitable anonymization configuration will be applied to the data. The selection of this configuration will be done along with the data controller. 3) A set of privacy metrics will be calculated in order to measure the risk of re-identification of the anonymized data. In case this risk was above a certain threshold (established by the data controller), the data would be anonymized again from scratch (going back to step 2) by applying

	a more restrictive anonymization configuration. 4) The anonymized data would be ready to be processed within the pilot
Overview of the different non-fully anonymous datasets that will be used in the project	**Please, fill Table 2 for each data set, summary below**
Please specify the source of the data and clarify if the data collection specifically occurs in relation to INFINITECH research or if these are previously collected data (i.e. collected for a purpose other than INFINITECH project, for instance for another EU project)	Within use case 1 [Driving profile] connected car's historical datasets come from other past projects and initiatives but are authorised to be used within INFINITECH. Specific informed consent signed by participants allow its usage within INFINITECH. The rest of the datasets will be collected within the INFINITECH project framework. In use case 2 [Driver's classification] data required from the driver (driver's car technical data and location) will be captured and provided by the insurance company (but won't be injected in the pilot's repositories). All these data, when may refer to personal (or sensitive) data, will be collected under the corresponding informed consent. Currently, the only way to capture data is by client's statement, unless the insurance company uses CTAG's OBU, or other equivalent device.
Please, provide information on the type of personal data and if they are common or sensitive data	Location of vehicles (current GPS coordinates of car's position gathered from the vehicle's GPS module. Common personal data). ID of vehicle/OBU (anonymized ID of the vehicle reporting technical data. Common personal data. This ID is renewed for each new route and identifies also a given route). Technical vehicle's data (Speed, acceleration, CO emissions, fuel consumption, etc. reported by the vehicle's OBU. Common personal data)
Please, provide details on the specific purpose of data collection and processing in your pilot activities in relation to this dataset	An anonymized vehicle's ID plus the set of locations (not anonymized) and related timestamps reported defines a "route". The captured vehicle's technical data linked to each "route" will be used to define and identify different driving profiles. In turn, these stored "routes'' will be used also to define and train the corresponding AI models to classify the driving styles. These driving profiles, combined with other context information such as weather forecast or anonymized location will help to classify a driver as bad/good or passive/aggressive. This information may assist also in fraud detection, identifying the profile of the driver when an accident happens.
Please, clarify if the intended recipients are INFINITECH partners and/or external entities.	All recipients considered within P#11 are INFINITECH partners. No info nor dataset would be given to third or external parties.
Please, indicate the legal basis of the processing according to Art. 6 GDPR	Use Case 1 participants (Data subjects) will sign an Informed Consent. This consent will be specific (clearly distinguished from other questions, in a form that is understandable and easily accessible and set out in clear and simple terms). Data subjects will have the right to withdraw their consent at any time, and will be expressly informed of that right. Data controllers will be able to show that the data subjects have given their consent to the processing of their personal data. In case that personal data (instead of synthetic data) would be used in Use

	Case 2, the participants (Data subjects) will sign a similar Informed Consent, managed by the insurance company
Please, provide information on the activities that you will implement for legal compliance, such as providing the volunteers with information, collecting informed consent and give them the opportunity to exercise their rights.	All the Use Case 1 participants (real vehicle drivers) will be volunteers, related to CTAG (partner providing the real connected cars infrastructure). All of them will be properly notified about the scope of the pilot, the datasets to be collected and the way this information will be managed, as well as their corresponding rights about their provided data, according to current GDPR. After this, an informed consent properly signed will be collected from each participant willing to get involved. A template about current informed consent managed by CTAG will be shared to help on an INFINITECH project consent form, covering all other issues related to the project and its framework. This final form would be the one to be signed by participants. In the same line, any possible Use Case 2 participant, related to the insurance company will be properly notified about the scope of the pilot, the datasets to be collected and the way this information will be managed, as well as their corresponding rights about their provided data, according current GDPR. After this, an informed consent properly signed will be collected.
Please provide information on the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing.	A Security framework (based on OAUTH protocol) will provide Identification and Authentication mechanisms to access the datasets through the proper APIs and access points defined for this purpose. The access (Create/Read/Update/Delete -CRUD- actions) to the data stored (REST APIS) on the P11 platform will be protected by this framework. Each potential external user, either physical user or component (IoT Agent, Data injectors, etc.) executing CRUD actions will require from a valid OAUTH token provided by this security framework. This means that any potential user should have first requested for an ID and a role, linked to a set of access policies. The registrations process, the ID and roles' management plus the policies access are controlled within this framework. All CRUD accesses to the data storage and management platform will be properly identified, filtered, authorised and registered. TLS/SSL (HTTPS) mechanisms will be used to encrypt communications through the REST APIs. Data within the P11 framework is not kept encrypted. Direct access to the stored data (understood as the access through the DBMS that supports the data storage and not through the INFINITECH defined REST APIs) is restricted ONLY to data controller/processor (depending on the final implementation) and isolated (within private VPN and VM infrastructures) from external accesses. Internal user/password mechanisms, provided by the DBMS and/or the VPN provider will be used to manage these accesses. The access to the hardware/software platform that supports the different components of the P#11 framework is managed by the Testbed provider (UNINOVA). This partner will implement the mechanisms to avoid unauthorised access to data and/or components.
In case of further processing of previously collected personal data, please,	

Confirm that the beneficiary has lawful basis for the data processing	The beneficiary controlling the data in this pilot confirms that it has a lawful basis for the data processing.
Confirm that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects	The beneficiary controlling the data in this pilot confirms that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.
Clarify if personal data will be transferred to/from non-EU countries	*According to 5.1.6 DoA Part B:* No transfer or Personal Data is planned, neither from Turkey, Switzerland or Israel to the EU nor from EU to Turkey, Switzerland or Israel. In Turkey, the partners will comply with national laws and ethical mandates.
How will this comply with applicable data protection rules at EU and national level?	*According to 5.1.6 DoA Part B:* The process of managing INFINITECH data by partners outside the EU will be done in ways compliant to laws of at least one EU member state. Compliance for data protection given by INFINITECH in Europe will be extended to the whole financial network if using INFINITECH outside Europe.

C.12 Pilot 12

Do you plan to collect and/or process personal data in your pilot operations?	Yes
Please explain how all of the personal data you intend to process is relevant and limited to the purposes of the research project (in accordance with the "data minimisation principle")	The end user of the risk assessment tool (health insurance company) identifies what data they need for the assessment.
Anonymization of research data	
What research data will be anonymized/pseudonymised?	The pilot will collect measured data from sensors and subjective answers to questionnaires. Both categories of data will be anonymized.
What research data will not be anonymized/pseudonymised? Please, explain why.	N/A

Please, provide details on the anonymization/ pseudonymization techniques that will be implemented in your pilot activities.	The sensitive data from the pilot will be anonymized following a risk-based approach that will allow to determine automatically which techniques (anonymization operations) should be applied to each of the variables of the dataset. The anonymization process will be performed in several steps: 1) All the possible anonymization configurations (i.e. different combinations of generalization, randomization and deletion operations) for the dataset will be calculated. 2) Taking into account the privacy requirements of the pilot and how the data will be used later on, a suitable anonymization configuration will be applied to the data. The selection of this configuration will be done along with the data controller. 3) A set of privacy metrics will be calculated in order to measure the risk of re-identification of the anonymized data. In case this risk was above a certain threshold (established by the data controller), the data would be anonymized again from scratch (going back to step 2) by applying a more restrictive anonymization configuration. 4) The anonymized data would be ready to be processed within the pilot. Furthermore, to prevent identification of individuals form the trained models, due to the lack of large numbers of people participating in the pilot, the training is done in two stages: Initial training with synthetic data generated without taking into account the actual behaviour of the trial users ensures a broadly pre-trained model. Final training of the pre-trained models using the anonymized data from the actual users, enhanced by synthetic data of similar characteristics.
Overview of the different non-fully anonymous datasets that will be used in the project.	**Please, fill Table 2 for each data set, summary below**
Please specify the source of the data and clarify if the data collection specifically occurs in relation to INFINITECH research or if these are previously collected data (i.e. collected for a purpose other than INFINITECH project, for instance for another EU project)	Real world data collected in the project for project use, or - in case specific consent is given - for broader research purposes outside of the scope of INFINITECH.
Please, provide information on the type of personal data and if they are common or sensitive data	Sensitive data : Physical activity data, medical info, subjective answers for symptoms, habits or self-assessments
Please, provide details on the specific purpose of data collection and processing in your pilot activities in relation to this dataset	Assess lifestyle and health of the users to derive a risk factor for premium health insurance contracts, and use of lifestyle and health parameters to provide feedback (insight) and advice (coaching) to the user in order to improve his/her behaviour.
Please, clarify if the intended recipients are INFINITECH partners and/or external entities.	Data is used only by pilot partners. Risk assessment tool is to be used by any insurance company, without any access to the collected data. To use the tool that an external company has to be feeding it data from their clients.
Please, indicate the legal basis of the processing according to Art. 6 GDPR	
Please, provide information on the activities that you will implement for legal compliance, such as providing the volunteers with information, collecting	Before participation in any data collection activity, each participant fills out an informed consent form in which detailed information is provided on the design of the study (data collection procedure), as well as the collected data and how this data is stored and processed. Prior to providing informed consent, participants will be given the

informed consent and give them the opportunity to exercise their rights.	opportunity to ask questions that may arise after reading the provided information on the studies (information leaflet in simple language). As part of the informed consent procedure, participants are made aware of their rights to drop out of the study (e.g. stop collecting data), without given reason. Informed consent may only be provided by participants of legal age (i.e. 18 years or older). No additional inclusion or exclusion criteria are envisioned at this point in time.
Please provide information on the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing.	ISO27001 certified (procedures & good practices for information management), GDPR compliance
In case of further processing of previously collected personal data, please,	
Confirm that the beneficiary has lawful basis for the data processing	The beneficiary controlling the data in this pilot confirms that it has a lawful basis for the data processing.
Confirm that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects	The beneficiary controlling the data in this pilot confirms that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.
Clarify if personal data will be transferred to/from non-EU countries	*According to 5.1.6 DoA Part B:* No transfer or Personal Data is planned, neither from Turkey, Switzerland or Israel to the EU nor from EU to Turkey, Switzerland or Israel. In Turkey, the partners will comply with national laws and ethical mandates.
How will this comply with applicable data protection rules at EU and national level?	*According to 5.1.6 DoA Part B:* The process of managing INFINITECH data by partners outside the EU will be done in ways compliant to laws of at least one EU member state. Compliance for data protection given by INFINITECH in Europe will be extended to the whole financial network if using INFINITECH outside Europe.

C.13 Pilot 13

Do you plan to collect and/or process personal data in your pilot operations?	Pilot 13 will not use personal data.

C.14 Pilot 14

Do you plan to collect and/or process personal data in your pilot operations?	There will be two different datasets for the pilot. The first one will be the dataset created by the insurance companies for the scope of the pilot implementation and evaluation purpose. This dataset will contain personal data. The second one will be the anonymized dataset that will be provided to the service providers in order to develop, calibrate and validate the services required for the pilot implementation. Despite the fact that the dataset will be anonymized, the data will be characterised as personal data and will be handled under the restrictions of GDPR.
Please explain how all of the personal data you intend to process is relevant and limited to the purposes of the research project (in accordance with the "data minimisation principle")	In order to develop the required services for the pilot implementation, the geospatial information, crop type, sowing and harvest days, yield and actual estimation are required. The dataset that will be provided to the service providers will be anonymized and no personal data are required (name, surname, address, etc.). However, these data will be considered as personal data since indirectly the service providers will be able to identify the owners of the fields.
Anonymization of research data	
What research data will be anonymized/pseudonymised?	The personal data of the fields' owners can be provided anonymized.
What research data will not be anonymized/pseudonymised? Please, explain why.	The field's information (crop type, sowing/ harvest dates, yield and actual estimation), as well as indices derived from the sentinel images processing and from the copernicus weather data, because they are crucial information for the services' outcomes and for insurance companies risk assessment process.
Please, provide details on the anonymization/ pseudonymization techniques that will be implemented in your pilot activities.	The data can be provided to the processor anonymized by the insurance companies since no personal data are required in order to deploy the pilot. Each field (the geospatial information along with the aforementioned details) can be accompanied by a specific id (e.g. 1, 2, 3, etc.) and no personal data can be provided.
Overview of the different non-fully anonymous datasets that will be used in the project	**Please, fill Table 2 for each data set, summary below**
Please specify the source of the data and clarify if the data collection specifically occurs in relation to INFINITECH research or if these are previously collected data (i.e. collected for a purpose other than INFINITECH project, for instance for another EU project)	Data will be collected from the insurance companies' clients' database and will occur in relation to INFINITECH research. Furthermore, more data will be used from the copernicus hub (sentinel images and weather data).
Please, provide information on the type of personal data and if they are common or sensitive data	No sensitive data are required for the pilot implementation.
Please, provide details on the specific purpose of data collection and processing in your pilot activities in relation to this dataset	The data collection will be used in order to monitor the fields with regards to the crop monitoring and damage assessment
Please, clarify if the intended recipients are INFINITECH partners and/or external entities.	INFINITECH partners

Please, indicate the legal basis of the processing according to Art. 6 GDPR (see Table 2)	NDA between insurance companies and Technology/ service providers
Please, provide information on the activities that you will implement for legal compliance, such as providing the volunteers with information, collecting informed consent and give them the opportunity to exercise their rights.	NDA between insurance companies and Technology/ service providers
Please provide information on the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing.	Possible Data encryption software could be installed/implemented at the Server side of Technology/service provider.
In case of further processing of previously collected personal data, please,	
Confirm that the beneficiary has lawful basis for the data processing	The beneficiary controlling the data in this pilot confirms that it has lawful basis for the data processing.
Confirm that appropriate technical and organisational measures are in place to safeguard the rights of the data subjects	The beneficiary controlling the data in this pilot confirms that appropriate technical and organisational measures are in place to safeguard the rights ot the data subjects.
Clarify if personal data will be transferred to/from non-EU countries	*According to 5.1.6 DoA Part B:* No transfer or Personal Data is planned, neither from Turkey, Switzerland or Israel to the EU nor from EU to Turkey, Switzerland or Israel. In Turkey, the partners will comply with national laws and ethical mandates.
How will this comply with applicable data protection rules at EU and national level?	*According to 5.1.6 DoA Part B:* The process of managing INFINITECH data by partners outside the EU will be done in ways compliant to laws of at least one EU member states. Compliance for data protection given by INFINITECH in Europe will be extended to the whole financial network if using INFINITECH outside Europe.