ALIGNER

How to use the ALIGNER Fundamental Rights Impact Assessment template

Extract from D4.2	
Work Package	WP4
Dissemination Level	PU
Author	Donatella Casaburo (KUL)
Co-Author	Irina Marsh (CBRNE)
Contributor(s)	Plixavra Vogiatzoglou (KUL)
Reviewed by	Daniel Lückerath (Fraunhofer)

Contact:

info@aligner-h2020.eu
www.aligner-h2020.eu

How to use the ALIGNER Fundamental Rights Impact Assessment template

The **ALIGNER Fundamental Rights Impact Assessment** (AFRIA) is a tool addressed to LEAs who aim to deploy AI systems for the purposes of prevention, investigation, detection or prosecution of criminal offences or execution of criminal penalties (i.e., law enforcement purposes) within the EU. As such, the AFRIA is **not** designed to be used in the following circumstances:

   a. During the development stage of the AI systems, even if carried out by LEAs; and
   b. When deploying AI systems for purposes other than law enforcement ones.

The AFRIA is a reflective exercise, seeking to further enhance the already existing legal and ethical governance systems of LEAs. Hence, the AFRIA has two main functions. First, it helps LEAs identify and mitigate the risks posed by the deployment of a certain AI system in relation to ethical principles and (selected) fundamental rights of individuals. Second, it is a suitable instrument for LEAs to explain and record their decision-making processes. In other words, the AFRIA is a **process aimed to assist LEAs in building and demonstrating compliance with ethical principles and fundamental rights** while deploying AI systems in a law enforcement context, as established by Article 27 of the AI Act.

   a. What the AFRIA addresses: A single AI system deployed for a single law enforcement purpose or a set of connected law enforcement purposes in a pre-determined context of use

An AFRIA addresses a **single AI system** deployed by LEAs. As a consequence, LEAs-users need to perform a separate AFRIA for each AI system they intend to deploy.

LEAs can perform a single AFRIA for an AI system deployed for either **a single law enforcement purpose or a set of connected law enforcement purposes**.[1] The connection between the purposes needs to be evaluated in the particular case by the LEAs-users themselves. Therefore, it is of paramount importance for LEAs to always perform the AFRIA in relation to a **pre-determined context of use**. This may include, for instance, information on the AI system's target group, geographical area and time period of deployment, and trigger conditions.

   b. When a AFRIA should be performed: Prior to the deployment of the AI system

In accordance with Article 27 of the AI Act, in AFRIA should be performed by LEAs **prior to the deployment of the AI system**, to inform the decision-making process on the *if*, *when*, *why* and *how* of the deployment. In case an AI system is already deployed for law enforcement purposes, LEAs are even more encouraged to conduct an AFRIA, unless they are already using a similar instrument to ensure compliance with the AI Act.

Performing an AFRIA is an iterative process. The AFRIA needs to be recorded, reviewed, and updated throughout the whole lifecycle of the AI system to reflect eventual changes in the functioning of the technology and/or its circumstances of deployment.

---

[1] For instance, LEAs can perform a single AFRIA for an AI system deployed for both detection and prosecution of criminal offences.

c. Who is responsible to perform the AFRIA: A dedicated multidisciplinary team

LEAs should establish a diverse and **multidisciplinary team**, responsible for performing the AFRIA. The team should include members of the organisation with legal, operational, and technical expertise. It is also advisable to involve the organisation's data protection officer in the AFRIA process.

If possible, LEAs should engage in discussions with the provider of the AI system assessed to clarify eventual uncertainties on the functioning of the AI system itself.

The AFRIA consists of two different, but connected, templates: the Fundamental Rights Impact Assessment and the AI System Governance.

1. The Fundamental Rights Impact Assessment

The **Fundamental Rights Impact Assessment template** helps LEAs identify and assess the risks that the AI system they wish to deploy may pose to the fundamental rights of individuals.

In ALIGNER D4.1, **four categories of fundamental rights** were identified as the most likely to be impacted by the use of AI systems in the law enforcement domain. These are:

1. Presumption of innocence and right to an effective remedy and to a fair trial;
2. Right to equality and non-discrimination;
3. Freedom of expression and information; and
4. Right to respect for private and family life and right to protection of personal data.

Accordingly, the Fundamental Rights Impact Assessment template is divided in four parts and, in each one of them, a group of fundamental rights is used as **benchmark for the following assessment**. To simplify the assessment process, the template contains an overview of the content of the four selected groups of fundamental rights, as defined by the Charter of Fundamental Rights of the European Union [Figure 1].

1. Presumption of innocence and right to an effective remedy and to a fair trial Everyone charged with a criminal offence must be presumed innocent until proved guilty according to law. Everyone whose rights and freedoms are violated has the right to an effective remedy before a tribunal. Everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal previously established by law, including rights: ❖ to be informed promptly of the nature and cause of the accusation; ❖ to bring their arguments and evidence as well as scrutinise and counteract the evidence presented against them; and to obtain an adequately reasoned and accessible decision.		
Challenge	**Evaluation**	**Estimated risk level**
1.1 The AI system does not communicate that a decision/advice or outcome is the result of an algorithmic decision		
1.2 The AI system does not provide percentages or other indication on the degree of likelihood that the outcome is correct/incorrect, prejudicing the user that there is no possibility of error and therefore that the outcome is undoubtedly incriminating		
1.3 The AI system produces an outcome that forces a reversal of burden of proof upon the suspect, by presenting itself as an absolute truth, practically depriving the defence of any chance to counter it		
1.4 There is no explanation of reasons and criteria behind a certain output of the AI system that the user can understand		
1.5 There is no indication of the extent to which the AI system influences the overall decision-making process		
1.6 There is no set of measures that allow for redress in case of the occurrence of any harm or adverse impact		

Figure 1: Example of Fundamental Rights Impact Assessment template, emphasis added

a. 'Challenge' column

To help and guide LEAs-users in their assessment, the template already lists some '**challenges**'. These are some possible **characteristics embedded in AI systems that may pose risks to the fundamental right** [Figure 2]. The challenges are formulated in a negative form (e.g., "*there is no …*"), so as to reduce the risk of acquiescence biases and stimulate further thought. LEAs may rely on the pre-listed challenges or add additional ones, as required.

1. Presumption of innocence and right to an effective remedy and to a fair trial
Everyone charged with a criminal offence must be presumed innocent until proved guilty according to law. Everyone whose rights and freedoms are violated has the right to an effective remedy before a tribunal. Everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal previously established by law, including rights: ❖ to be informed promptly of the nature and cause of the accusation; ❖ to bring their arguments and evidence as well as scrutinise and counteract the evidence presented against them; and to obtain an adequately reasoned and accessible decision.

Challenge	Evaluation	Estimated risk level
1.1 The AI system does not communicate that a decision/advice or outcome is the result of an algorithmic decision		
1.2 The AI system does not provide percentages or other indication on the degree of likelihood that the outcome is correct/incorrect, prejudicing the user that there is no possibility of error and therefore that the outcome is undoubtedly incriminating		
1.3 The AI system produces an outcome that forces a reversal of burden of proof upon the suspect, by presenting itself as an absolute truth, practically depriving the defence of any chance to counter it		
1.4 There is no explanation of reasons and criteria behind a certain output of the AI system that the user can understand		
1.5 There is no indication of the extent to which the AI system influences the overall decision-making process		
1.6 There is no set of measures that allow for redress in case of the occurrence of any harm or adverse impact		

Figure 2: Example of Fundamental Rights Impact Assessment template, emphasis added

b. 'Evaluation' column

In the '**evaluation**' column, LEAs need to identify **how the listed challenges relate to the assessed AI system**, for the identified law enforcement purposes and in relation to the envisaged context of use. In other words, LEAs need to explain both *whether* and, if so, *to what degree,* the assessed AI system embeds each of the challenges, and *how* it does so [Figure 3].

1. Presumption of innocence and right to an effective remedy and to a fair trial
Everyone charged with a criminal offence must be presumed innocent until proved guilty according to law. Everyone whose rights and freedoms are violated has the right to an effective remedy before a tribunal. Everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal previously established by law, including rights: ❖ to be informed promptly of the nature and cause of the accusation; ❖ to bring their arguments and evidence as well as scrutinise and counteract the evidence presented against them; and to obtain an adequately reasoned and accessible decision.

Challenge	Evaluation	Estimated risk level
1.1 The AI system does not communicate that a decision/advice or outcome is the result of an algorithmic decision	The AI system communicates that the outcome is the result of an algorithmic decision only in case the risk of the assessed natural person offending equals to or overcomes the threshold of 70%, while the communication is omitted in case of risk level below 70%	
1.2 The AI system does not provide percentages or other indication on the degree of likelihood that the outcome is correct/incorrect, prejudicing the user that there is no possibility of error and therefore that the outcome is undoubtedly incriminating	The AI system does not communicate the confidence score of the output and/or of the algorithm and is impossible for the user to establish it	
1.3 The AI system produces an outcome that forces a reversal of burden of proof upon the suspect, by presenting itself as an absolute truth, practically depriving the defence of any chance to counter it	When the AI system flags that the risk of the assessed natural person offending overcomes the threshold of 70%, an investigation against them is immediately started, even in absence of other evidence incriminating them	
1.4 There is no explanation of reasons and criteria behind a certain output of the AI system that the user can understand	The AI system does not communicate neither the criteria nor the data leading to the output and the users cannot understand them with any other means	
1.5 There is no indication of the extent to which the AI system influences the overall decision-making process	The weight of the output of the AI system in the overall decision-making process is not specifically evaluated	
1.6 There is no set of measures that allow for redress in case of the occurrence of any harm or adverse impact	The subjected natural person can seek redress only in court, if and when an official trial starts	

Figure 3: Example of Fundamental Rights Impact Assessment template, emphasis and text added

c. 'Estimated risk' column

In the '**estimated risk**' column, LEAs need to estimate the **level of the risk** the deployment of the AI system may pose to the fundamental right of individuals, due to the already evaluated challenges posed by the AI system's characteristics. In doing so, LEAs need to consider the following factors:

1. Impact, namely severity of the prejudice experienced by the affected individuals, if the risk occurs; and
2. Likelihood, namely the level of probability the risk will occur.

The risk matrix below helps the user estimate and visualize risks.

		Likelihood				
		Rare	**Unlikely**	**Possible**	**Likely**	**Almost Certain**
I m p a c t	**Extreme**	Limited	Serious	Critical	Critical	Critical
	Major	Limited	Serious	Serious	Critical	Critical
	Moderate	Low	Limited	Serious	Serious	Critical
	Minor	Low	Limited	Limited	Serious	Serious
	Insignificant	Low	Low	Low	Limited	Limited

Table 1: Risks matrix

The user should estimate both the impact (in *insignificant*, *minor*, *moderate*, *major*, or *extreme*) and the likelihood (in *rare*, *unlikely, possible*, *likely* or *almost certain*). Based on the estimations, the user finds the impact level (*low*, *limited*, *serious,* or *critical*) in the square where the impact and the likelihood meet.

For instance, in relation to challenge 1.1, if the user estimates the impact as *minor* and the likelihood as *possible*, the risk level will be *limited* [Figure 4].

1. Presumption of innocence and right to an effective remedy and to a fair trial		
Everyone charged with a criminal offence must be presumed innocent until proved guilty according to law. Everyone whose rights and freedoms are violated has the right to an effective remedy before a tribunal. Everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal previously established by law, including rights: ❖ to be informed promptly of the nature and cause of the accusation; ❖ to bring their arguments and evidence as well as scrutinise and counteract the evidence presented against them; and to obtain an adequately reasoned and accessible decision.		
Challenge	Evaluation	Estimated risk level
1.1 The AI system does not communicate that a decision/advice or outcome is the result of an algorithmic decision	The AI system communicates that the outcome is the result of an algorithmic decision only in case the risk of the assessed natural person offending equals to or overcomes the threshold of 70%, while the communication is omitted in case of risk level below 70%	Limited
1.2 The AI system does not provide percentages or other indication on the degree of likelihood that the outcome is correct/incorrect, prejudicing the user that there is no possibility of error and therefore that the outcome is undoubtedly incriminating	The AI system does not communicate the confidence score of the output and/or of the algorithm and is impossible for the user to establish it	Serious
1.3 The AI system produces an outcome that forces a reversal of burden of proof upon the suspect, by presenting itself as an absolute truth, practically depriving the defence of any chance to counter it	When the AI system flags that the risk of the assessed natural person offending overcomes the threshold of 70%, an investigation against them is immediately started, even in absence of other evidence incriminating them	Critical
1.4 There is no explanation of reasons and criteria behind a certain output of the AI system that the user can understand	The AI system does not communicate neither the criteria nor the data leading to the output and the users cannot understand them with any other means	Critical
1.5 There is no indication of the extent to which the AI system influences the overall decision-making process	The weight of the output of the AI system in the overall decision-making process is not specifically evaluated	Serious
1.6 There is no set of measures that allow for redress in case of the occurrence of any harm or adverse impact	The subjected natural person can seek redress only in court, if and when an official trial starts	Critical

Figure 4: Example of Fundamental Rights Impact Assessment template, emphasis and text added

1.1.1 The AI System Governance

The **AI System Governance** template helps LEAs identify, explain, and record possible measures to mitigate the risks that the deployment of the AI system would pose to the ethical principles and the fundamental rights of individuals.

In 2019, The High-Level Expert Group on Artificial Intelligence set up by the European Commission published its 'Ethics Guidelines for Trustworthy AI'.[2] There, the Group identified seven key requirements that an AI system should fulfil to be considered 'trustworthy', i.e., a lawful, ethical, and robust AI system. These requirements are:

1. Human agency and oversight;
2. Technical robustness and safety;
3. Privacy and data governance;
4. Transparency;
5. Diversity, non-discrimination and fairness;
6. Societal and environmental wellbeing; and
7. Accountability.

Accordingly, the AI system Governance template is divided in seven parts and, in each one of them, a key requirement for trustworthy AI is used as **benchmark for grouping the minimum standards** that an AI system should achieve [Figure 5].

1. Human autonomy									
Component	Minimum standards to be achieved	Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
		Challenge no.	Risk level		Final estimated risk level	Further actions		
Human agency	The task allocation between the AI system and the user allows meaningful interactions	[1.2]							
		[1.5]							
	There are procedures to describe the level of human involvement and the moments for human interventions	[1.5]							
		[2.2]							
		[4.1]							

Figure 4: Example of AI System Governance template, emphasis added

a. 'Component' column

In the '**component**' column, the **building blocks substantiating the considered key requirement** are listed [Figure 6].

1. Human autonomy									
Component	Minimum standards to be achieved	Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
		Challenge no.	Risk level		Final estimated risk level	Further actions		
Human agency	The task allocation between the AI system and the user allows meaningful interactions	[1.2]							
		[1.5]							
	There are procedures to describe the level of human involvement and the moments for human interventions	[1.5]							
		[2.2]							
		[4.1]							

Figure 5: Example of AI System Governance template, emphasis added

b. 'Minimum standards to be achieved' column

To help and guide LEAs-users in their decision-making process, the template already lists some '**minimum standards to be achieved**'. These are some possible **characteristics that an AI system should embed** or possible **governance procedures that the organisation should always implement for the deployment of the AI system to be considered trustworthy** [Figure 7].

1. Human autonomy								
Component	Minimum standards to be achieved	Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
		Challenge no.	Risk level		Final estimated risk level	Further actions		
Human agency	The task allocation between the AI system and the user allows meaningful interactions	[1.2]						
		[1.5]						
	There are procedures to describe the level of human involvement and the moments for human interventions	[1.5]						
		[2.2]						
		[4.1]						

Figure 6: Example of AI System Governance template, emphasis added

c. 'Initial risk estimate column

To further help and guide LEAs-users in their decision-making process, in the '**initial risk estimate** column, the template already connects the **minimum standard with** (at least) **one previously estimated challenge and risk level**, as that was already estimated in the Fundamental Rights Impact Assessment template. The link between the minimum standard and the estimated risk is highlighted where the minimum standards are suitable to mitigate possible risks that the deployment of the AI system would pose to the fundamental rights of the individuals. The numbers (e.g., 1.2, 1.5, and so on) correspond to the 'challenges' listed in the Fundamental Rights Impact Assessment template. For each of the challenges, the column automatically reports the risk level (i.e., *low*, *limited*, *serious,* or *critical*), as it was already estimated in the Fundamental Rights Impact Assessment template [Figure 8].

1. Human autonomy								
Component	Minimum standards to be achieved	Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
		Challenge no.	Risk level		Final estimated risk level	Further actions		
Human agency	The task allocation between the AI system and the user allows meaningful interactions	[1.2]	Serious					
		[1.5]	Serious					
	There are procedures to describe the level of human involvement and the moments for human interventions	[1.5]	Serious					
		[2.2]	Serious					
		[4.1]	Limited					

Figure 7: Example of AI System Governance template, emphasis and text added

Where the minimum standards are not suitable to mitigate possible risks that the deployment of the AI system would pose to the fundamental rights of the individuals, the '**initial risk estimate**' column is **left blank** [Figure 9].

2. Transparency								
		Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
Component	Minimum standards to be achieved	Challenge no.	Risk level		Final estimated risk level	Further actions		
Traceability	There are mechanisms to ensure the traceability of the input data used by the AI system and its							

Figure 8: Example of AI System Governance template, emphasis added

d. 'Additional mitigation measures implemented' column

Whenever an initial risked is linked to a minimum standard, in the '**additional mitigation measures implemented**' column, LEAs need to state:

- *if* and *how* the minimum standard **is (foreseen to be) implemented** in the AI system and/or within the organisation; and
- *how* the minimum standard **is suitable to mitigate the connected previously estimated risk**, by paying particular attention to how the standard is reducing the impact and/or the likelihood of the risk [Figure 10].

1. Human autonomy								
		Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
Component	Minimum standards to be achieved	Challenge no.	Risk level		Final estimated risk level	Further actions		
Human agency	The task allocation between the AI system and the user allows meaningful interactions	[1.2]	0	The AI system communicates the confidence score of the output, so that an informed decision on the follow-up actions is possible				
		[1.5]	0	The deployer can play an active role in the decision-making process, by modifying the parameters informing the decision of the AI system				
	There are procedures to describe the level of human involvement and the moments for human interventions	[1.5]	0	The weight of the output of the AI system in the decision-making processes of the organisation is concretely evaluated. The results are made known to the deployers, who are tasked to take an informed decision on the follow-up actions				
		[2.2]	Serious	Deployers can flag the poor performance of the AI system and this triggers an evaluation from the technical department				
		[4.1]	Limited	A data protection impact assessment is performed yearly				

Figure 9: Example of AI System Governance template, emphasis and text added

Whenever an initial risk is not linked to a minimum standard, and thereby left blank, in the '**additional mitigation measures implemented**' column, LEAs need to state:

- *if* and *how* the minimum standard **is (foreseen to be) implemented** in the AI system and/or within the organisation [Figure 11].

2. Transparency								
		Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
Component	Minimum standards to be achieved	Challenge no.	Risk level		Final estimated risk level	Further actions		
Traceability	There are mechanisms to ensure the traceability of the input data used by the AI system and its			Both the input data and the outcomes are recorded and accessible to the user				

Figure 10: Example of AI System Governance template, emphasis and text added

e. 'Final assessment' column

Whenever an initial risk is **linked** to a minimum standard, in the '**final assessment**' column, LEAs need to:

- Use the risk matrix seen above [Table 1], to estimate the **final risk level** to fundamental rights that the deployment of the AI system may pose, despite the implementation of additional mitigation measures; and
- if any, list **further actions** suitable to improve the implementation of the minimum standard and further mitigate the final risk to fundamental rights, for instance in case where the mitigation measures are not considered sufficient in relation to the estimated risk [Figure 12].

1. Human autonomy								
Component	Minimum standards to be achieved	Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
		Challenge no.	Risk level		Final estimated risk level	Further actions		
Human agency	The task allocation between the AI system and the user allows meaningful interactions	[1.2]	0	The AI system communicates the confidence score of the output, so that an informed decision on the follow-up actions is possible	Low	N/A		
		[1.5]	0	The deployer can play an active role in the decision-making process, by modifying the parameters informing the decision of the AI system	Limited	Implementing a mechanism to allow the user to add new parameters informing the decision of the AI system		
	There are procedures to describe the level of human involvement and the moments for human interventions	[1.5]	0	The weight of the output of the AI system in the decision-making processes of the organisation is concretely evaluated. The results are made known to the deployers, who are tasked to take an informed decision on the follow-up actions	Low	N/A		
		[2.2]	Serious	Deployers can flag the poor performance of the AI system and this triggers an evaluation from the technical department	Limited	Establishing a framework for periodical performance evaluation		
		[4.1]	Limited	A data protection impact assessment is performed yearly	Low	N/A		

Figure 11: Example of AI System Governance template, emphasis and text added

Whenever an initial risk is **not linked** to a minimum standard, in the '**final assessment**' column, LEAs need to:

- list, if any, **further actions** suitable to improve the implementation of the minimum standard [Figure 13].

2. Transparency								
Component	Minimum standards to be achieved	Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
		Challenge no.	Risk level		Final estimated risk level	Further actions		
Traceability	There are mechanisms to ensure the traceability of the input data used by the AI system and its outcomes			Both the input data and the outcomes are recorded and accessible to the user		Embedding retrieval of criteria leading to the decision		

Figure 12: Example of AI System Governance template, emphasis and text added

f. 'Responsible department' and 'timeline' columns

In the '**responsible department**' and '**timeline**' column, LEAs need to specify the department of their organisation responsible for the implementation of the mitigation measures foreseen, and their (estimated) timeline of adoption [Figure 14].

1. Human autonomy								
		Initial risk estimate		Additional mitigation measures implemented	Final assessment		Responsible department	Timeline
Component	Minimum standards to be achieved	Challenge no.	Risk level		Final estimated risk level	Further actions		
Human agency	The task allocation between the AI system and the user allows meaningful interactions	[1.2]	0	The AI system communicates the confidence score of the output, so that an informed decision on the follow-up actions is possible	Low	N/A	Technical department	dec/24
		[1.5]	0	The deployer can play an active role in the decision-making process, by modifying the parameters informing the decision of the AI system	Limited	Implementing a mechanism to allow the user to add new parameters informing the decision of the AI system	Legal and technical departments	mar/25
	There are procedures to describe the level of human involvement and the moments for human interventions	[1.5]	0	The weight of the output of the AI system in the decision-making processes of the organisation is concretely evaluated. The results are made known to the deployers, who are tasked to take an informed decision on the follow-up actions	Low	N/A	Legal and technical departments	mar/25
		[2.2]	Serious	Deployers can flag the poor performance of the AI system and this triggers an evaluation from the technical department	Limited	Establishing a framework for periodical performance evaluation	Technical department	dec/24
		[4.1]	Limited	A data protection impact assessment is performed yearly	Low	N/A	Legal and technical departments	dec/24

Figure 13: Example of AI System Governance template, emphasis and text added