Feasibility study on a forecasting and early warning tool for migration based on Artificial Intelligence technology

Written by Ecorys
November 2020

ECORYS

EUROPEAN COMMISSION

Directorate-General for Migration and Home Affairs
Directorate F — Financial Audit, Data Management and Risk Assessment
Unit F.2: Situational Awareness, Resilience and Data Management

Contact: Pawel Busiakiewicz

E-mail: HOME-MIH@ec.europa.eu

European Commission
B-1049 Brussels

Feasibility study on a forecasting and early warning tool for migration based on Artificial Intelligence technology

Directorate-General for Migration and Home Affairs

EUROPE DIRECT is a service to help you find answers to your questions about the European Union

Freephone number (*):
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

LEGAL NOTICE

Luxembourg: Publications Office of the European Union, 2021

PDF	ISBN 978-92-76-29645-4	doi: 10.2837/932912	DR-02-21-127-EN-N

Table of Contents

Abbreviations

AI	Artificial Intelligence
AI HLEG	High-Level Expert Group on Artificial intelligence
AUC	Area Under Curve
BD4M	Big Data for Migration Alliance
CV	Computer Vision
DG CONNECT	Directorate-General for Communications Networks, Content and Technology
DG HOME	Directorate-General for Migration and Home Affairs
DG JUST	Directorate-General for Justice and Consumers
EASO	European Asylum Support Office
FRONTEX	European Border and Coast Guard Agency
EC	European Commission
ECB	European Central Bank
EEAS	European Union External Action
EES	Entry-Exit System
ETIAS	European Travel Information and Authorization System
ETL	Extract Transform and Load
EDPS	European Data Protection Supervisor
EU	European Union
Eurodac	European Dactyloscopy
Eurostat	European Statistical Office
EUROSUR	European Border Surveillance system
ESA	European Space Agency
EPS	Early warning and Preparedness System
ETL	Extract, Transformation and Load
Eu-LISA	European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice
FRA	Fundamental Rights Agency
GDELT	Global Database of Events, Language, and Tone
GDPR	General Data Protection Regulation
IDE	Integrated Development Environment
IOM	International Organisation for Migration
GMDAC (IOM)	Global Migration Data Analysis Centre
IPCR	Integrated Political Crisis Response
ISAA	Integrated Situational Awareness and Analysis
JRC	Joint Research Centre
JHA	Justice and Home Affairs Council
KCMD	Knowledge Centre on Migration and Demography
LDA	Linear Discriminant Analysis
LFS	Labour Force Survey
OSINT	Open-Source Intelligence
OECD	Organisation for Economic Cooperation and Development
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error

NLP	Natural Language Processing
NMF	Non-Negative Matrix Factorization
NN	Neural Networks
PCA	Principal Component Analysis
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
SIS	Schengen Information System
SOX	Sarbanes-Oxley Act
TFEU	Treaty on the Functioning of the European Union
UNHCR	United Nations High Commissioner for Refugees
VIS	Visa Information System
XAI	Explainable Artificial Intelligence

Study glossary

- **Artificial Neural networks (ANN):** An algorithm that endeavours to recognise underlying relationships in a set of data through a process that vaguely mimics the way the human brain operates;
- **Algorithm:** A finite sequence of well-defined, computer-implementable instructions, typically to solve a class of problems or to perform a computation;
- **AI-Tool/platform:** Both terms refer to the hardware and software, including existing solutions and specific code/algorithms;
- **Citizenship:** The particular legal bond between an individual and his or her State, acquired by birth or naturalisation, whether by declaration, choice, marriage or other means according to national legislation;
- **Computer Vision:** A field of computer science that works on enabling computers to see identify, and process images and videos in the same way that human vision does;
- **Data accessibility:** Accessibility refers to open and restricted (e.g. by commercial provider or government) of data sources;
- **Deep Learning:** a subset of machine learning in which artificial neural networks learn from large amounts of data. This approach can be compared to how humans learn from experiences meaning that the deep learning algorithm would perform a task repeatedly, each time altering it to improve the outcome.
- **Explainable Artificial Intelligence (XAI):** Artificial Intelligence technology that is programmed to describe its purpose, rationale and the operation and the outcome (result) in a way that can be understood by key stakeholders of different levels of expertise. In particular, within the frame of the present study, a focus is made on users and operators of the future system;
- **External borders:** The borders between EU Member States and third countries. The borders between Schengen Associated Countries (Norway, Iceland and Switzerland) and third countries are also considered as external borders;
- **ETL (Extract, transform, load):** Is the general procedure of copying data from one or more sources into a destination system that represents the data differently from the source or in a different context than the source;
- **Expert Systems:** A computer system that emulates the decision-making ability of a human expert. They are designed to solve complex problems by reasoning through bodies of knowledge;
- **Fairness:** In machine learning, a given algorithm is said to be fair, or to have fairness if its results are independent of some variables we consider to be sensitive and not related to it (e.g.: gender, ethnicity, sexual orientation, etc.);
- **Forecasting:** Projections that rely on assumptions, such as linearity of previous trends, to produce relatively specific, mostly quantitative estimates on migration across different time horizons. For this study, forecasting will mainly refer to estimates for the short (1-4 weeks) and

medium-term (1-3 months). Forecasting differs from 'predictions', which imply less susceptibility to external shocks.

- **Forecasting models:** Models providing projections by employing data-driven approaches to understand future migration flows in the medium to long term.
- **Foresight methods:** Foresight methods aim to analyse a range of possible trends to create a vision of the future in a few years' or even a few decades, which aims to build uncertainties into policy planning systematically and operational decision-making.
- **Fairness metric:** A quantification of unwanted bias in training data or models;
- **Internal borders:** The borders between two Schengen countries;
- **Irregular migration:** Is the movement of persons that cross the borders of the country by breaching the regulatory norms of the countries of departure, transit or destination;
- **Mixed migration flows:** Complex migratory population movement including refugees, asylum-seekers, economic migrants and/or other types of migrants as opposed to migratory population movements that consist entirely of one category of migrants;
- **Prediction:** Prediction is concerned with estimating the outcomes for unseen data. For this purpose, you fit a model to a training data set, which results in an estimator that can make predictions for new samples.
- **Prediction variables (broad):** Generally speaking, the broader the prediction variable, the more likely that the decisions are strategic, political and over a longer period. Example: assuming the prediction variable is 'increase, decrease or no change in irregular arrivals expected at the EU's external borders six months from now', the variable has a broad geographic scope, so predictions will not help much in deciding where resources could be needed. It would probably be most useful in alerting policy and political decision-makers to changes in the level of attention they should give to external borders and cooperation with neighbouring countries;
- **Prediction variables (narrow):** When the prediction variable is narrow, it is more likely that the following decisions are operational and over a shorter period. Example: if the variable is 'the number of migrants who will arrive at Greece's external borders from visa obliged countries next month without a valid visa and cross', the variable can be said to have a narrow geographic scope and is more directly relevant to specific operational decisions. It would also provide input into cooperation with Turkey. However, it obviously will have limited relevance to decisions in Spain or Italy, since it will not be outputting anything in relation to the Western or Central Mediterranean Routes. Similarly, it would not help much in strategy and diplomacy regarding Libya or countries of the Sahel;
- **Reliability:** The ability to accurately describe an objective to forecast migration and the level of precision in doing this. When speaking of 'data reliability', we refer to reliability in terms of the data source's ability to accurately describe their objective to forecast migration and its precision in doing this.

- **Scenario-building:** The development of scenarios – or 'imagined situations' – equipped to inform strategic 'big picture' thinking than to offer accurate, action-oriented and operational input, as they draw on the insights and imagination of experts that have their own 'cognitive biases'.
- **Secondary movements:** Occur within the Schengen area when asylum seekers and irregular migrants move from the Member State of the first arrival to the other Member States;
- **Structured data:** Highly-organised and formatted in a way so its easily searchable in relational databases;
- **Timeliness:** The extent to which, e.g. data is made available shortly after a reporting period.
- **Machine Learning:** Method of data analysis that automates analytical model building. It is based on the idea that systems can learn from data.
- **Natural Language Processing (NLP):** A subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human languages, how to program computers to process and analyse natural language data;
- **Robotics:** although robotics itself is not exclusively an AI technology, AI helps automate tasks performed, and it introduces flexibility and learning capacity in previously rigid applications;
- **Speech Recognition:** subfield that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers;
- **Unstructured data:** No pre-defined format or organisation, making it much more difficult to collect, process, and analyse.

Executive summary

Overall objective of the study

The **overall objective** of this study was to analyse the feasibility of developing a forecasting and early warning tool based on AI technology (hereafter 'AI-Tool'), which is capable of forecasting and assessing the direction and intensity of irregular migratory flows to and within the EU and to provide early warnings and forecasts on this basis both in the short term (1 to 4 weeks) and in the medium term (1 to 3 months). This AI-Tool should be able to provide reliable estimates to allow the European Commission and the EU Member States to inform their migration management activities, such as the planning and organisation of capacities and other border management aspects accordingly.

Requirements for the feasibility of the AI-Tool

The assessment of the feasibility of a forecasting and early-warning tool takes into consideration various different requirements, which form the baseline of creating an AI-Tool capable of the specifications set out above. These range from an adequate legal basis, sufficiently available data to inform the AI-Tool, the underlying AI architecture itself, to considerations on the host and associated organisational and governance structures. Each of these assessments was assessed for this feasibility study, and the individual outcomes are presented in the figure below.

These requirements are considered to be fulfilled when their respective availability or possibilities of realisation face no legal and operational obstacles and can be implemented without significant additional impediments. Requirements are considered to be partially fulfilled when it is theoretically possible that these can be fulfilled based on a range of alternatives available. However, certain decisions still have to be taken in these cases, whereby the respective outcome might entail additional follow up steps. Requirements are not fulfilled when their entire setup depends on final decisions on several other requirements. As such, these can only be arranged at a later stage of the implementation process.

As such, the requirement of an adequate legal basis being in place was assessed to be partially fulfilled, considering that while the development of the AI Tool is compliant with EU primary legislation, amendments to secondary legislation might be necessary, particularly in terms of working arrangements and/or mandates. In terms of the availability of a variety of relevant qualitative and quantitative data sources to be included in and informing the outputs of the AI-Tool, this was considered to be sufficient. As such, the requirement is considered to be fulfilled. Likewise, the operational assessment indicated that an appropriate AI architecture according to the specifications of the AI-Tool can be developed. Hence, this requirement was

also labelled as fulfilled, considering that the realisation of appropriate AI architecture is possible, albeit not yet conducted. The hosting structure for the AI tool requires additional steps to be taken depending on the choice of a host or hosting structure. This particularly refers to investing in relevant operational capacities, but in some cases, also incorporating legislative changes to mandates and working arrangements. Accordingly, this requirement was considered to be partially fulfilled. Last but not least, the assessment of organisational and governance structures already in place indicates that there is a need for investment into an appropriate coordination mechanism for the AI-Tool, such as additional working agreements, and a central coordination point to facilitate data access and data sharing. Considering that various other steps first need to be taken, this requirement is as of now considered not to be fulfilled.

Figure 0.1 Outcome of the Feasibility Study

Scope of the study

The scope of the AI-Tool was defined as covering mixed migration flows towards the EU, by which 'complex population movements including refugees, asylum seekers, economic migrants, victims of trafficking, smuggled migrants, unaccompanied minors and other migrants' are meant.[1] Here, the focus is on irregular movements, i.e. the various aforementioned types of migrants' attempts to reach the EU outwith regulatory norms governing the entry into or exit from countries of origin, transit or destination. As such, the term 'irregular migration' is used throughout this study.

[1] IOM, *Irregular Migration and Mixed Flows: IOM's approach*.

Practically, the AI-Tool would incorporate and adequately process various data sources on all stages of the process of irregular migration. This includes assessments of situations in third countries in the first place, which could provide early indications of on setting movements of irregular migration. Data on trends from previous movements have to be updated against recent data on the actual size of flows along transit routes, and information on the situation and number of irregular migrants already present in the EU external border countries has to be taken into account to assess and inform in the short term when irregular border crossings can be expected. Likewise, this and other data also inform predictions on the countries to and the timeframe in which secondary movements and asylum applications can be expected.

However, various elements beyond the overall objective and practical vision of the AI-Tool need to be taken into consideration before any such AI-Tool may be developed at EU level. This includes an assessment of related risks, the current legal and policy framework, the operational requirements for implementing and running the AI-Tool, technical possibilities, organisational structures relating to hosts and usage rights, and more. Hence, with the specific objective of this study being to assess the feasibility of a forecasting and early warning AI-Tool for irregular migration, all the aforementioned aspects were addressed in a range of designated assessments, as per the paragraphs and table below.

Table 1.2 Assessment objectives

Chapter	Assessment	Conducted by
3 and 4	**General assessment** - Assess the feasibility of developing an AI-based tool for irregular migration forecasting for the European Commission and EU JHA agencies, taking into account similar projects at the EU and national level.	Seefar (section 3.2, 3.3. and 3.4 and chapter 4) Ecorys (section 3.5, 3.6., 3.7)
5	**Operational assessment** - Assess the capabilities and capacities (personnel and financial), including availability, for implementing, running, managing and maintaining the AI-Tool.	Everis (chapter 5,
6	Legislative assessment - Identify gaps and weaknesses (legal and policy framework) at EU level.	Law & Internet Foundation
7	**Organisational assessment** - Analyse possible organisational structures, taking the current organisational structure into account, for the successful incorporation of the AI-Tool.	MPI Europe
8	**Trustworthiness assessment** - Assess the trustworthiness of the AI-based tool (human agency and oversight, technical robustness and	Everis

	safety, transparency, accountability).	
9	**Risk assessment** - Develop a risk assessment related to the use of AI to feed into technical, operational and political considerations based upon standard methodology (ISO 31000 Family).	everis

Study outline and reading guide

The **general assessment** (chapter 3) focusses on the level of development of AI technology in fields related to migration, and situations in third countries in so far as they can lead to irregular migration. Therefore, similar projects already in place at national or European level were taken into consideration. This chapter also addresses specific aspects of migration processes as incorporated by existing AI-Tools, and outlines the types of forecasts the AI-Tool would be able to produce to support the decision-making of policymakers.

Further, an extensive data mapping and assessment exercise was conducted as part of this assessment, which showed that different data sources demonstrate significantly varying characters ranging from qualitative to quantitative data, from historical to current data, as well as timeliness of the data. The data source assessment of this study collected characteristics and provides analyses on 60 data sources, and offers further intensified research into 39 of these data sources. The range of data sources covers a broad spectrum of local and global data, historical and real-time data, statistical, administrative, and innovative data, various data formats, et cetera.

For this reason, the data sources were further assessed on their *relevance for the following three possible forecasting categories*.

- Forecasting category, A focuses on the forecasting of relevant and/or potentially critical situations in third countries and the EU. Therefore, it incorporates data on events at early stages of irregular migratory movements, such as on the development of underlying socio-economic and political factors in a given country and/or sudden events, which could have a joint impact on setting off irregular migratory movements from such countries;
- Forecasting category B focuses on irregular border crossings into the EU. Data informing the forecasting of irregular migration flows would cover the events between irregular migrants' setting off from countries of origin along various routes into the EU, including the routes taken and the emergence of new routes of irregular migration;
- Hence, forecasting category C continues where B leaves off and covers the forecasting of events following the irregular border crossings into the EU. This includes the prediction of levels of (change in and location of) asylum applications lodged in the EU and secondary movements towards destination countries within the EU.

A case study *(chapter 4)* is presented to provide a clear and concrete example of how a forecast would be produced by the AI-Tool and to illustrate the kind

of modelling and design processes that are implied by different purpose-prediction decisions.

The **operational assessment** (chapter 5) provides an operational overview of the proposed AI-Tool from the perspective of the needs to be addressed for its implementation. These needs entail the definition of a High-Level Architecture, which is declined in three different scenarios for implementation. The proposed architecture design of the AI-Tool outlines which functionalities are required to facilitate the operationalisation a solution that responds to the needs expressed by the European Commission. The three scenarios, therefore differ in the functionalities of the AI-Tool, which at the present stage, are defined at high level. Any further Low Level Design will incur in the development of Application Building Blocks (ABBs) and identification of commercial-off-the-shelf (COTS) and/or custom-developed solutions to be defined within the architecture and benchmarked from vendor-specific information in order to establish a reasonable and cost-efficient implementation plan. The proposed scenarios and the definition of the AT technical functionalities also aim at providing guidance for the European Commission on the needs and the scope to be addressed under the next steps in the design and implementation of the AI IT system.

The proposed scenarios are designed to be incremental. Therefore, the functionalities they comprise an increase in complexity across scenarios. For each of the possible designs, the number and profiles of required experts (e.g. data scientists) are also provided and their roles described.

Finally, these scenarios are not mutually exclusive, their incremental nature in functionalities allow to the end-user to start with the low-level ambition scenario and use it as an entry point in order to test available functionalities with the validated data and the irregular migration quantitative models to be developed by experts. From the entry point, the end-user may decide to upgrade the functionalities, therefore moving towards the medium level ambition architecture, and ultimately towards the high-level ambition scenario architecture.

The **legislative assessment** (chapter 6) examined the existing legal framework, including primary (relevant provisions from TEU, TFEU, EU Charter of Fundamental Rights) and secondary legislation (e.g. Dublin, Eurodac, VIS Regulations and others) as well as a few relevant proposals from the New Pact on migration (included in Annex B). Particular attention was paid to the provisions related to the exchange of data. Additionally, the mandates and competences of relevant EU JHA agencies were scrutinised regarding the feasibility of hosting, managing, operating, accessing and feeding information into the AI-Tool. Furthermore, the existing scope of exchange of information between these agencies was addressed through an examination of concluded Working Arrangements (WAs) between them. Further, a legal analysis on a wide range of potential data sources for the AI-Tool was conducted.

The **organisational assessment** (chapter 7) then reviewed which skills, resources, and coordination mechanisms are needed to host the AI-Tool. On this basis, it looked into the most feasible options. Three sets of tasks are outlined, on which the decision to host the AI-Tool hinges, namely on operational capacity, analytical capacity, and possibilities relating to the dissemination capacity.

Furthermore, the **trustworthiness assessment** (chapter 8) provides a framework comprising the techniques and tools that can help address a range of criteria during the design and operation of the AI-Tool for irregular migration forecasting. These are the following: Human agency and oversight, technical robustness and safety, as well as transparency and accountability. To address this set of criteria, the trustworthiness assessment developed a framework by relying on four key dimensions, namely fairness, explainable artificial intelligence (XAI), functional monitoring, and the transversal dimension of governance.

The operational risks associated with the future implementation of the tool have been identified and described in the **risk assessment** (chapter 9); these risks have been categorised in business, technology and organisational risks, and a mitigation strategy has been proposed. Chapter 9 also provides an evaluation of these risks, indicating their likelihood and magnitude of impact.

Finally, chapter 10 summarises the main findings of the study, and chapter 11 contains the recommendations and possible next steps towards the development and implementation of the AI-Tool.

Main findings and recommendations

Current AI forecasting landscape

Some conflict and crisis forecasting systems already exist in the EU or are being developed at prototype level. Some of these were developed by EU JHA agencies and others by Member State institutions. Such conflict and crisis forecasting systems can be considered to be useful for irregular migration forecasting to some extent, as theory and experience suggest that conflicts and crises can influence irregular migration patterns. The closest AI-Tool related to the objectives of this study was held to be EASO's Early Warning and Forecasting System, which forecasts the number of asylum applications that EU Member States can expect by monitoring and forecasting crises in third countries. Similar to the purpose of the AI-Tool covered in this study, the already existing AI-Tools are and were intended to support the operational preparedness of recipients of the systems' outputs.

Stakeholder challenges

Various challenges and necessary factors to the development of a designated AI-Tool to forecast irregular migration movements were identified on the basis of research into already existing AI-Tools and on the basis of stakeholder interviews. Challenges relating to the use of particular data sources were identified upfront by stakeholders. Reporting cycles appear to differ among Member States and EU JHA agencies. This may be down to various reasons but is crucial for expectation management in terms of both the data providers and the receivers of the output of the AI-Tool into which such data is integrated. Additionally, not all Member States report their data at the same points time or within the reporting period itself. This might distort the

underlying analysis of the AI-Tool accordingly. Further, especially in cases of data being provided late, data processing and integration has to be done in a more fragmented manner, rather than all at structurally at once.

Furthermore, *understanding the strengths and limitations of analyses* drawn from complex statistics or machine learning systems might, in some cases, be considered challenging, considering the relative novelty and overall complexity of such systems. Hence, to fully understand the context, input, and underlying analytical steps towards the outputs of such an AI-Tool, training on the interpretation of AI-Tool output for decision-makers will support the analyses and use they can make of the outputs. Hence, for the AI-Tool to be useful, its immediate users, such as data scientists, and its end-users, such as decision-makers, should have multiple opportunities to provide feedback within the scope of their involvement with the AI-Tool.

Last but not least, considering that there are only a few tools equivalent to the AI-Tool which the European Commission envisions, the best guidance in terms of data mapping and analysis activities comes from migration theories, research into available data sources, and expert knowledge about possible influences on prediction variables.

Data source mapping and assessment

In line with the aforementioned, the data source assessment of this study collected characteristics and provides analyses on 60 data sources, and offers further intensified research into data 39 of these data sources. The range of data sources covers a wide spectrum of local and global data, historical and real-time data, statistical, administrative, and innovative data, various data formats, etc. For all three forecasting categories as mentioned above, a wide range of potentially relevant, accessible and operationalizable data sources was identified.

In terms of incorporating particular data sources into the envisioned AI-Tool, it is not possible to quantify the improved accuracy that can be expected from adding new data sources to the AI-Tool prior to building it. Instead, only intuitive estimates can be provided in this regard, as well as estimations on the costs and complexity of adding a data source. Data sources that are already being collected for other purposes may appear cheaper at the current point in time, but are less likely to be optimised for the forecasting purpose and are also more vulnerable to uncertainty and change in future.

Also, no significant legal obstacles were identified regarding the use of most assessed data sources, so long as the respective terms and conditions are complied with. Limitations might potentially arise regarding the terms of service for some data sources. Further, the on-going spread of widespread disinformation and fake news might influence the accuracy of the AI-Tool's outputs. Personal bias might also influence the forecasts.

Legislative framework of the AI-Tool and fundamental rights considerations

Overall, the legislative assessment concluded that the development and use of an AI-Tool for the purpose of this study are compliant with EU primary law. However, amendments to secondary legislation, particularly regulations establishing EU JHA agencies' mandates, might be necessary to ensure specific provisions on the management and related aspects of the AI-Tool are

included. Additionally, more precise Working Arrangements between the relevant EU JHA agencies are necessary to specify elements of the respective EU JHA Agencies' relationships in this regard, with particular emphasis on the exchange of data to and from the concerned AI-Tool.

Further, it is emphasised that fundamental rights (e.g. the right to non-discrimination, right to asylum) should always be taken into consideration in the design of the AI-Tool's functionalities and algorithms, as fundamental rights are the cornerstone of the EU and its values. In this regard, human oversight should be implemented in order to avoid biased conclusions and forecasts.

Location and hosting of the AI-Tool

None of the candidates assessed would currently meet all of the required criteria without additional investment, and in some cases, a change in their mandates. Hence, when selecting a host, the European Commission will therefore have to decide whether an investment should be made towards building the operational or analytical capacity of a single host, or if a joint approach should be developed, whereby the operational and analytical tasks would be shared among multiple agencies. To this end, the organisational assessment also made suggestions on reviewing the governance structure of the AI-Tool, including through further investment into appropriate coordination mechanisms, such as additional working agreements, and a central coordination point to facilitate data access and data sharing.

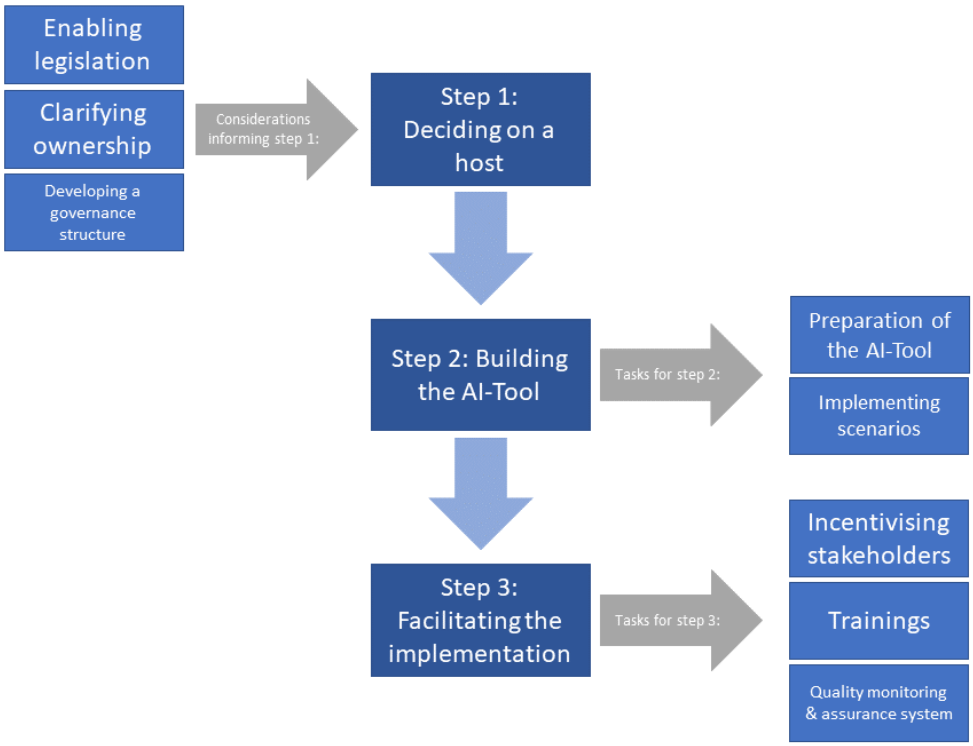
Conclusion of the study

Overall, based on the various feasibility assessments which consisted of an extensive ex-ante data source assessment, desk research, stakeholder interviews as well as the findings of the Closed-doors seminar, it can be concluded that a well-performing forecasting system can be built. However, it will only be possible to precisely assess its reliability ex-post after the implementation of the AI-Tool. This study submits evidence-based recommendations to the European Commission on such steps that should be considered as per each assessed area, and ultimately on how the AI-Tool can be implemented most efficiently and respecting such standards as the EU's Ethics Guidelines on Trustworthy AI.

Next steps

The following section briefly outlines the various next steps that need to be taken. Figure 0.1 below indicates whether these steps can be taken simultaneously, and which steps might be required to be taken before the commencement of others. It should be noted that while the chart indicates that the building of the AI-Tool should follow the decision on the host, these processes can effectively be done in parallel. However, the arrangement of one step following the other was based on the consideration that the eventual host might already have a specific IT architecture in place, which would have to be adapted by the AI-Tool to ensure full functionality. Hence, building an AI-Tool in parallel while a host is chosen would likely result in changes having to be made to it at a later stage to integrate it into the host's environment.

Figure 0.1 Next steps



Step 1: Identification of a host

A first necessary step towards operationalisation is the identification of a host of the AI-Tool. The decision on the host depends on whether the EU has a preference for building the capacity of one EU JHA agency or body to host the AI-Tool, or for pursuing a joint approach where the operational, analytical, and dissemination responsibilities are shared by multiple actors. To inform this decision, we recommend assessing whether it is feasible or desirable to have one stakeholder operate the AI-Tool and another analyse and disseminate its outputs, and what cooperation mechanisms would need to be in place. Further, it needs to be considered how much scope exists for investing in additional operational or analytical capacity.

Considerations informing step 1:

A set of important considerations need to be addressed in the process of making a decision on a host. These are as follows:

- **Enabling legislation:** As outlined in the assessments, there may be a need for enabling legislation and almost certainly a formal definition of the respective roles and responsibilities of EU JHA agencies in relation to the AI-Tool. This framework should define access restrictions to the AI-Tool and information it produces, knowledge dissemination settings, hosting and management rights and coordination mechanisms. This consideration can also encompass looking into the assessments of the mandates of individual

EU JHA agencies with a view to establishing the extent to which changes would have to be made depending on the preferred hosting structure.

- **Clarifying ownership:** Also, the question of ownership of the AI-Tool needs to be clarified since this underlines the point about the governance structure of the AI-Tool. Specifically, this means deciding on who will own the outputs of the AI-Tool, and then reflecting on how this ownership may influence the willingness of stakeholders to, for example, invest resources (e.g. budget, staff capacity) to help design and operate the AI-Tool.
- **Development of the governance structure:** Another important next step consists in the development of the governance structure of the AI-Tool. Synergies between different stakeholders are already in place to some extent, such as some information-exchanges and joint risk analyses, but the frameworks and legal bases for governing an AI-Tool are still missing. For example, additional working arrangements in line with individual mandates are most likely needed to facilitate cooperation on the AI-Tool. It should also be decided on whether the AI-Tool should have a central coordination information point, for example within the European Commission, where the outputs of the AI-Tool can be relayed to different EU JHA agencies, institutions and Member States depending on their needs and preferences.

Step 2: Building the AI-Tool

As indicated above, the building of the AI-Tool can theoretically be already commenced while step 1 is on-going. Nonetheless, it is recommended to wait until a host is decided on, to gain insights on technical specifications on the respective architecture used. In building the AI-Tool, the model itself needs to be prepared. The AI-Tool can be conceptualised as offering the three previously outlined types of forecasting categories (A, B, and C), which reflect time-scale requirements, as well as the reality of the main phases of irregular migration movements. In this regard, forecasting category A addresses the underlying situations and potential shocks in countries of origin, i.e. the drivers of irregular migration. Forecasting category B covers occurrences between the countries of origin from where irregular migrants set off, i.e. the transit phase until including crossing into the EU. As such, this category first assesses shifting flows along routes in a first step, with a view to ultimately predicting irregular border crossings into the EU as output. Forecasting category C covers occurrences within the EU related to irregular migration, namely secondary movements and asylum applications.

Tasks for step 2:

For step 2, the below presented set of tasks is crucial in building the AI-Tool for its envisioned purpose.

- **Preparation of the AI-Tool's data model:** For the AI-Tool to offer the three previously presented types of forecasting categories, preparatory activities have to be conducted to this end, such as obtaining training data for the AI-Tool, and beginning to choose the appropriate architecture and building blocks.

- **Implementing scenarios:** From a technical perspective, the next steps include the implementation of the scenarios proposed under the operational assessment as an incremental process based on the business needs (proof of concept, additional use cases). The proposed architecture design of the AI-Tool outlines which functionalities are required to facilitate the operationalisation a solution that responds to the needs expressed by the European Commission. Relevant steps to follow to ensure such implementation of the solution include:
  - The elicitation and management of requirements;
  - The declination of the proposed architecture in Application Building Blocks (ABBs);
  - The declination of the ABBs into Solution Building Blocks (SBBs);
  - The conduction of a benchmark analysis to select the cost-effective SBBs;
  - The validation of the SBBs by the stakeholders;
  - The development of the Low-level Design;
  - The preparation of the Implementation Roadmap;
  - The preparation of procurement process to initiate the acquisition of licenses and necessary hardware.

Step 3: Facilitate the implementation

Taking decisions on the host and building and providing the AI-Tool to end-users without any additional steps to facilitate the implementation are unlikely to ensure that the AI-Tool will be used to its full potential. To best facilitate the implementation, relevant stakeholders need to be informed on various elements relating to the AI-Tool, including on the data on which the outputs are based, on how outputs are generated in the first place and ultimately how trustworthy they are. This has significant implications on the extent to which the outputs of the AI-Tool will be used and thus can eventually offer added value to EU processes. As such, stakeholders not only need to be informed about the AI-Tool's specificities but also trained on understanding and engaging with its output. This will assist in incentivising full use of the AI-Tool. A similarly important task is to ensure an appropriate quality monitoring and assurance system is in place, which allows for ensuring the AI-Tool continues to provide valuable output. Creating the latter can already be commenced at earlier stages, i.e. while the AI-Tool is built; however, the existence of a functioning quality control system is crucial in the facilitation stage.

Tasks for step 3:

For step 3, the below-presented set of tasks will prove relevant in ensuring the AI-Tool is embraced by stakeholders in accordance with its purpose, and can thus be utilised to its full potential.

- **Identify incentives for stakeholders**: Furthermore, there is a need for identifying the incentives for stakeholders to use the AI-Tool and cooperate on its operation, analysis and dissemination. Relevant policy and operational units should indicate their willingness and ability to use the AI-Tool's outputs for their programming depending on where they see it to be most useful. This may require the European Commission to invest in

political capital and expectation management (e.g. to inform on whether such an AI-Tool can be deemed trustworthy enough to incentivise its use in daily policymaking).

- **Provide trainings:** Next to this, expert knowledge and training of the AI-Tool are required to build the capacity of end-users if they are not analysts. This is related to interpreting the forecasts, including their caveats and limitations, understanding the quality assurance processes (including the caveats and limitations of the data fed to the tool), and how the forecasts are produced, as well as how to translate into actions, measures, policy and decision-making. It is recommended to start early in building the capacity of immediate and end-users. There will likely be three categories of user: technical operators who need to develop and maintain data sources and the actual system itself; analysts who interpret the system's outputs; and decision-makers who need to understand the strengths and limitations of analysis incorporating the AI-Tool's forecasts.
- **Develop a quality monitoring and assurance system** to process the data coming in and out of the AI-Tool in order to produce as accurate as possible forecasts.

1 INTRODUCTION

1.1 Context and relevance of the study

During the so-called migration crisis in 2015-2016, the European Union (EU) and its Member States found themselves insufficiently prepared for the arrival of large numbers of irregular migrants at their sea, land and air borders. While Frontex identified a rise in detected illegal border crossings from 2014 by Syrians and other nationalities, policymakers struggled to analyse these data and translate them into concrete policy actions at the EU and national levels.[2] This experience prompted the EU and Member States to invest in improving their capacity to anticipate and respond to emerging trends of irregular migration.[3] As a first step, these efforts aimed at addressing the lack of availability of comprehensive, real-time data that would allow policymakers to track fast-changing irregular migration flows and routes of irregular migrants and thereby to anticipate the surge in arrivals. The European Union has since invested heavily in data collection and analysis to collate up-to-date, regular information about irregular migration flows in a digestible format to decision-makers, including through the Integrated Situational Awareness and Analysis (ISAA) weekly reports, prepared under the leadership of the European Commission, and EASO's Early warning and Preparedness System (EPS).

As part of these developments, irregular migration forecasting has gained increasing interest among European policymakers who are keen to tap into the mechanism's new analytical possibilities to forecast, manage, and prevent irregular migration flows. The potential role of this AI-Tool in EU migration management centres around its ability to produce more accurate, timely and actionable forecasts of irregular migratory patterns (i.e. flows, routes) that can then support the development of a more coordinated, comprehensive and coherent set of actions by different actors holding migration and border management portfolios.

In order to better understand the risks and benefits of forecasting tools for irregular migration, it is useful to place them in the context of the existing research on different techniques used to anticipate irregular migration trends. Depending on their needs and time horizons, these techniques can help policymakers and operating Agencies to respond to specific events or changing irregular migration flows or routes in the short term (early warning systems) or help them prepare for different eventualities in the longer term (scenarios and other foresight methods). More specifically, early warning systems monitor trends or potential drivers of irregular migration and displacement to warn policymakers and operating Agencies of surges with as much advance warning as possible (although in practice, likely days or weeks

[2] Frontex, *FRAN Quarterly, Quarter 3, July – September 2014* (Warsaw: Frontex, 2014). Elizabeth Collett and Camille Le Coz, *After The Storm: Learning From The EU Response To The Migration Crisis* (Brussels: Migration Policy Institute, 2018).

[3] Schmidt and Hooper, *Preparing for Future Migration Trends: Using forecasting and scenario-building for forward-looking policies* (working paper, Brussels: Migration Policy Institute Europe, forthcoming).

in advance).[4] These systems use frequently updated data on irregular migration to warn of real-time or imminent irregular migration trends (or shocks) that require pre-emptive action with relatively high accuracy. However, researchers at EASO have found that their success depends heavily on the analytical capacity of experts and policymakers to interpret and respond to these data, including by setting accurate thresholds for early warning.[5]

Scenarios and other foresight methods[6] take a different approach, aiming to analyse a range of possible trends to create a vision of the future in a few years' or even a few decades' time, which aims to systematically build uncertainties into policy planning and operational decision-making.[7]

As highlighted in a recent survey by IOM's GMDAC, scenarios are better equipped to inform strategic 'big picture' thinking than to offer accurate, action-oriented and operational input, as they draw on the insights and imagination of experts that have their own 'cognitive biases'.[8] At the EU level, for example, EASO produced a report in 2019 that presents five possible scenarios for the future of international protection in the European Union, with input from the European Commission, Frontex, Europol, Member States, academia, international organisations, and NGOs.[9]

Forecasts of irregular migration differ from these approaches insofar that they produce relatively specific, mostly quantitative estimates on irregular migration across different time horizons – mainly in the short (1-4 weeks) and medium-term (1-3 months). They typically use historical data on past irregular movements to map movements in the future and, as such, assume a continuity that does not factor in sudden surges in spontaneous arrivals, as well as 'black swan' events like COVID-19. Forecasting methods also comprise

4 Apart from EASO's Early warning and Preparedness (EPS) system, examples of this include IOM's Displacement Tracking Matrix (DTM), which incorporates techniques used in early warning systems, such as tracking and monitoring displaced populations and their movements and needs.

5 EASO, *Quantitative Assessment of asylum- related migration*: *A survey of methodology* (Luxembourg: Publications Office of the European Union, 2016).

6 These can include, for example, horizon scanning and trends analysis, which identify changes in migration trends and draw conclusions about their drivers and implications for policymaking. See: OECD and EASO, *Migration Policy Debates: Can we anticipate future migration flows?* (Paris: OECD).

7 The Joint Research Centre (JRC), for example, has developed 'Demographic Scenarios for the EU' (see: Lutz et al, *Demographic Scenarios for the EU* (Luxembourg: Publications Office of the European Union, 2019).) and a Scenario Exploration System (SES) to engages EU policymakers and other stakeholders in scenario-building exercise that highlight the various outcomes of longer-term migration trends in Europe and beyond. Other recent uses of scenarios methods include 'Many More to Come?' (see: JRC Science Hub, *Many More to Come? Migration From and Within Africa* (Luxembourg: Publications Office of the European Union, 2018)) and 'Demographic and Human Capital Scenarios for the 21st Century' (see: Lutz et al, *Demographic and human capital scenarios for the 21st century* (Luxembourg: Publications Office of the European Union)). EU actors have also produced work around strategic foresight to explore global migration trends, including through the European Strategy and Policy Analysis System (ESPAS) report on Global Trends to 2030 (see: ESPAS, *Global Trends to 2030: Challenges and Choices for Europe* (Luxembourg: Publications Office of the European Union)) and the EPRS' Global Trends Unit initiative on long-term migration in the European Union (see EPRS, *Migration and the EU: A long-term perspective* (Luxembourg: Publications Office of the European Union)).

8 Acostamadiedo & Dag Tjaden, *Forecasting the future of migration—many approaches, one commonality: uncertainty* (Migration Data Portal Blog, 2020).

9 EASO, 2019.

a range of different mathematical models,[10] which vary in terms of their complexity and sophistication. Whereas some statistical models are based purely on numeric input, such as analysing statistics on asylum applications or seasonal changes in the number of irregular border crossings along transit routes, others also incorporate qualitative analyses, for example in the form of expert assessments on the push factors and related events in third countries, or insights from big data analyses on relevant and/or critical events. This improves the addressing of uncertainty in irregular migration patterns and assists in updating forecasts based on past occurrences, and thereby to safeguard policymakers against less expected developments.[11]

When using forecasts, it is important to be transparent about the applicability of these methods and the potential limitations of their results in the context of irregular migration.[12] As already noted, however, unlike scenarios or other more qualitative methods, forecasting is less well equipped to capture the uncertainties of non-selected, spontaneous flows and the range of different drivers of irregular migration that determine migration trends. Researchers at the OECD analysed forecasting tools and found them to be prone to high levels of error, which can be due to a wide range of technical challenges, including finding comparable data sources that rely on similar data collection methodologies (e.g. similar timeframes, reporting periods and underlying mathematical models).

As a result, and given the complexity in irregular migration processes and the diverse range of related and possibly overlapping concepts (mixed migration, circular migration, return migration, etc.), it is necessary to assume a higher degree of uncertainty in forecasting models of irregular migration.

The academic literature, therefore, tends to frame forecasts as 'projections' that rely heavily on assumptions, such as linearity of previous trends, rather than 'predictions' that imply less susceptibility to external shocks. To ensure the envisioned AI-Tool produces projections that apply across the EU context, it is important to draw on existing harmonised EU mechanisms and common definitions for systematic and frequent information-sharing as well as other tools or platforms that can support data analysis related to irregular migration for policy needs. Moreover, the architecture of such an EU-managed system must be based on open-source data, drawing from the intelligence and support of EU JHA agencies (such as eu-LISA), situational awareness systems (such as EUROSUR), and migration data catalogues (such as including KCMD's Dynamic Data Hub).[13]

---

[10] These include, for example, regression, time series, structural and log-linear models. See: Jakub Bijak and Arkadiusz Wisniowski, "Bayesian Forecasting Of Immigration To Selected European Countries By Using Expert Knowledge", *Journal Of The Royal Statistical Society* 173, no. 4 (2010): 775-796.

[11] Disney et al, *Evaluation of existing migration forecasting methods and models* (New York: CPC, 2015).

[12] Outside of the migration context, for example, initiatives such as the Good Judgment Project (see: Good Judgment, *Superforecasting will change the way you think about the future* (Good Judgment, 2020)) have highlighted the potential of forecasting methodology, producing what they call "superforecasts" on geopolitical events that even outperformed intelligence analysts with access to classified data.

As briefly indicated above, researchers and experts have also started to match forecasting and early warning tools with the analysis of 'big data' – large and complex data sets that exceed the processing capacity of traditional software – to better capture the complexity of real-time changes in drivers of irregular migration, such as the detection of relevant events. For example, EASO's Early Warning and Forecasting system use searches by the GDELT Project, which monitors and updates open-source data on global trends and emerging social, political and economic risks worldwide every 15 minutes. Moreover, IOM's GMDAC and KCMD have launched the Big Data for Migration Alliance (BD4M) to advance discussions on how to harness the potential of big data sources for the analysis of irregular migration and its relevance for policymaking, while ensuring the ethical use of data and the protection of individuals' privacy.[14] While outside the scope and aim of this study, other non-EU actors have also explored the use of privately owned big data applications to boost preparedness and response to irregular migration.[15]

Finally, using AI technology can be a powerful way to forecast irregular migration trends and to transform operational (and policy) migration decision-making at EU level. With AI technology already being applied to project future conflicts,[16] there is also growing evidence on the risks and benefits of using AI in forecasting that can be analysed in the context of irregular migration. This study combined these different elements, looking at the feasibility for the European Commission and EU JHA agencies to engage in irregular migration forecasting with the help of AI technology and big data.

1.2 Study description and scope

The **overall objective** of this feasibility study is to analyse the feasibility of developing a forecasting and early warning tool based on AI technology (hereafter 'AI-Tool'), which is capable of forecasting and assessing the direction and intensity of irregular migratory flows to and within the EU and to provide early warnings on this basis both in the short term (1 to 4 weeks) and in the medium term (1 to 3 months). This AI-Tool should be able to provide reliable estimates with a view to allowing the EU and Member States to inform their migration management activities, such as planning and organisation of capacities and border management aspects accordingly. As such, the scope of the AI-Tool would cover mixed migration flows towards the EU, by which 'complex population movements including refugees, asylum seekers, economic migrants, victims of trafficking, smuggled migrants, unaccompanied minors and other migrants' are meant. Here, the focus is on irregular movements, i.e. the various aforementioned types of migrants' attempts to reach the EU out with regulatory norms governing the entry into or exit from

[14] Big Data for Migration Alliance, *Harnessing the potential of new data sources and innovative methodologies for migration* (2019).

[15] For example, ESA has conducted several feasibility studies to this end. See: ESA, Migration Radar 2.0 - Big data applications to boost preparedness and response to migration – Feasibility Study (2018); ESA, *Big data for migration study - Big Data applications to boost mitigation preparedness and response to migration feasibility study* (2019); ESA, *Big data applications to boost preparedness and response to migration* (2018).

[16] For example, see: Ingham, *UN urged to develop global AI platform to predict wars* (Verdict, 2018).

countries of origin, transit or destination. As such, the term 'irregular migration' is used throughout this study.

Practically, the AI-Tool would be required to incorporate and adequately process various data sources on all stages of the process of irregular migration. This includes assessments of situations in third countries in the first place, which could provide early indications of on setting movements of irregular migration. Data on trends from previous movements have to be updated against recent data on the actual size of flows along transit routes, and information on the situation and number of irregular migrants already present in the EU external border countries has to be taken into account to assess and inform in the short term when irregular border crossings can be expected. Likewise, this and other data also informs forecasts on the countries to and the timeframe in which secondary movements and asylum applications can be expected.

However, various elements beyond the overall objective and practical vision of the AI-Tool need to be taken into consideration before any such AI-Tool may be developed at EU level. This includes an assessment of related risks, the current legal and policy framework, the operational requirements for implementing and running the AI-Tool, technical possibilities, organisational structures relating to hosts and usage rights, and more. Hence, with the specific objective of this study being to assess the feasibility of a forecasting and early warning AI-Tool for irregular migration, all the aforementioned aspects were addressed in a range of designated assessments, as per the table below.

Table 1.1 Assessment objectives

Chapter	Assessment
3	**General assessment** - Assess the feasibility of developing an AI-based tool for irregular migration forecasting for the European Commission and EU JHA agencies, taking into account similar projects at the EU and national level.
5	**Operational assessment** - Assess the capabilities and capacities (personnel and financial), including availability, for implementing, running, managing and maintaining the AI-Tool.
6	Legislative assessment - Identify gaps and weaknesses (legal and policy framework) at EU level.
7	**Organisational assessment** - Analyse possible organisational structures, taking the current organisational structure into account, for the successful incorporation of the AI-Tool.
8	**Trustworthiness assessment** - Assess the trustworthiness of the AI-based tool (human agency and oversight, technical robustness and safety, transparency, accountability).
9	**Risk assessment** - Develop a risk assessment related to the use of AI to feed into technical, operational and political considerations based upon standard methodology (ISO 31000 Family).

The key operational purpose of the AI-Tool is to enhance the performance of migration management and security on borders employing forecasting and early warning technologies, so that the EU and its Member States may put in place the necessary preparatory and preventive measures and, if needed, activate mitigating instruments promptly.

Although the above-introduced forecasting categories also include a forecasting category covering relevant events within the EU (forecasting category C), such as the forecasting of secondary movements, asylum applications and overstayers, this was not covered in-depth, nor by a specific case study. The main reason for this decision is that secondary movements concern a fundamentally different type of migration outputs that would need to be modelled separately and responds to different sets of drivers. In addition, the forecasting of secondary movements is generally a difficult exercise. Further, forecasting categories A and B were prioritised with a view to forecasting irregular entry into the EU in the first place, which is crucial data required to inform most elements within forecasting category C. Nonetheless, relevant data sources were identified for this forecasting category, alongside assessments of their relevance.

2 PROJECT OVERVIEW

2.1 Research activities conducted

During the **inception phase**, desk and literature reviews were conducted in order to shed light on the current academic debate as well as the policy and legal framework. Additionally, a mapping of potentially (big) data sources were provided. An overview of the main findings was presented in the final inception report. In preparation for the various assessments, each sub-team conducted its own desk research. However, when thematically relevant, findings were shared, coordinated and integrated amongst the various partners.

Initial **scoping interviews** were planned and carried with key stakeholders from various EU-agencies and bodies, including the European Commission, Frontex, EASO, to garner an initial overview of their understating and expectations for the AI-Tool in question. For nearly all of the scoping interviews, representatives from Ecorys, Seefar, MPI, LIF and Everis were present. Afterwards, the main trends shared views and differences as well as divergences identified amongst stakeholders were incorporated into the Inception Report and are also presented in the general assessment (chapter 3) of the present report.

Subsequently, following the Desk review, further **expert interviews** were planned and conducted by each individual team in preparation for the various assessments. For example, for the **general assessment,** various semi-structured interviews were planned with key informants, such as experts in AI and IT systems and migration policy at the EU level, to ensure that knowledge-gaps were targeted. For the **operational assessment,** interviews with key stakeholders, including Member State agencies and relevant multilateral organisations, were carried out with the goals of understanding the skills and resources required to operate these AI-Tools, how the results are analysed and disseminated, and consequently what organisational structures and coordination mechanisms they have put in place. In total, 14 stakeholder interviews were conducted for the operational assessment, with representatives from the following stakeholder groups being interviewed: the European Commission, EASO, EEAS, FRONTEX, Eurostat, Europol, JRC, FRA, EDPS Eurostat and eu-LISA.

Similarly, for the **legal assessment, trustworthiness assessment and organisational assessment**, expert interviews were conducted to understand the needs of stakeholders further, identify relevant knowledge gaps and further to validate findings of the desk research and data collection. In the areas where the assessments overlapped thematically, expert interviews were conducted by more than one team in order to coordinate their efforts and share access to given experts.

In order to contextualise and exemplify the study's findings further for the project's final phase, a **case-study** was developed, using a specific and known route of irregular migration, which was chosen in deliberation with the European Commission. Additionally, placing further focus on the availability and potential usability of data sources by the AI-Tool in question, **data analysis and subsequent assessment**, was performed. In doing so, all data sources that stood to be of relevance to the study's thematic focus were listed, analysed and ranked according to a wide range of pre-determined criteria. The outcome of the assessment and analysis was subsequently presented in the form of **recommendations**, which were further incorporated into a **roadmap to action**, serving as a guide in the eventuality that the European Commission wishes to develop and operationalise an AI-Tool for the purpose of forecasting irregular migration.

The study was conducted across eight months. During this time, regular meetings were held between the European Commission and the study team, several preliminary versions of the final report were produced (draft, revised, and final inception and interim reports). Further, several interviews were conducted with stakeholders from across EU JHA agencies, the European Commission and representatives of Member State institutions, and a virtual closed door seminar was held for a wide range of participants. For the latter, representatives of relevant authorities and selected stakeholders were brought together and presented with an overview of the preliminary results of the various assessments of this study, and with insights from already existing AI-Tools in the wider area of irregular migration.

2.2 Structure and outline of the final report

The final report presents the outcomes of each assessment, as well as an executive summary, which outlines the key outcomes of the assessment on the feasibility of an early warning and forecasting AI-Tool for the prediction of irregular migration-related crises. The report is structured as follows:

Table 2.1 Structure final report

Chapter	Chapter title
0.	Executive summary
1.	Introduction
2.	Project overview
3.	General assessment
4.	Modelling considerations
5.	Operational Assessment
6.	Legislative Assessment
7.	Organisational Assessment
8.	Risk Assessment
9..	Trustworthiness Assessment
10.	Main outcomes and conclusion of the study

Chapter	Chapter title
11.	Next steps
Annex	Bibliography

3 GENERAL ASSESSMENT

3.1 Objective and summary of the general assessment

Objective

The objective of the General Assessment is to assess the feasibility of developing an AI-based tool for irregular migration forecasting for the European Commission and relevant EU JHA Agencies. The General Assessment focusses on the level of development of AI technology in the field of migration, and on situations in third countries in so far as they can lead to irregular migration. To this end, experiences from similar projects at EU and national level are taken into account. By analysing, mapping and comparing existing AI-Tools on forecasting and early warning in this area, gaps were identified, as well as limitations and strengths of existing AI-Tools determined, with a view to understanding what type of forecasts would be most relevant for policy and operational purposes. The General Assessment also includes a comprehensive data source mapping and the analysis of the most relevant data sources for various forecasting categories.

Interlinkages with other tasks and approach

The General Assessment serves as a basis for the other assessments, in that it forms both the first and last piece of the assessment sequence, in that preliminary findings were used to inform other assessments, the finalised versions of which in turn informed the general assessment. For example, the General Assessment can identify a specific forecasting ambition, for which the Operational Assessment can then highlight risks and opportunities involved in that ambition.

3.2 Underlying priorities and needs

Priorities and needs identified for the AI-Tool covered various aspects, including the type of output, scope, and its usability.

As for the intended type of output of the AI-Tool, it was noted that priority should be given to forecasting, as opposed to the prediction of irregular migration flows. While both terms can be used interchangeably, their difference relates to forecasting involving a certain timeframe in which predicted events are deemed likely to occur. This was decided against the background of the overall purpose of the AI-Tool being to enable relevant EU JHA agencies and Member States to better inform their organisation of their migration management and the security of EU external borders. As such, on the basis of forecasted information, the EU and Member States may put in place the necessary preparatory and preventive measures and, if needed, activate their respective mitigating instruments in a timely manner.

In terms of the AI-Tool's scope, it was decided that this should encompass several steps. First, it entails the forecasting of critical situations in third countries and countries of origin that could have an effect on setting off irregular migration flows to the EU. As such, it would cover events entailing underlying socio-economic and political trends, as well as sudden events which could individually or jointly push irregular migration movements. Further, it includes the forecasting of irregular migration flows, including levels on established routes and irregular border crossings into the EU. Last but not least, it includes events following irregular border crossings into the EU, such as secondary movements and asylum applications. As such, the timeframe of forecasts to be provided by the AI-Tool range from long term, to medium and short term. While the primary focus of the AI-Tool's output in this regard was placed on irregular arrivals at the EU external borders or irregular migrants in countries neighbouring the EU, it is nonetheless recognised that the earlier step of establishing information on trends and events in countries of origin is crucial in informing the primary focus. Further, on the basis of insights on arrivals to EU external borders, secondary activities within the EU can be forecasted on a short term.

In this regard, looking at the content of output, this should entail timely forecasts of irregular migration trends into the EU, across various points in time. Usability concerns were also addressed, in that it would have to be transparent how exactly the system operates in terms of how results and estimates are generated for policymakers and relevant national authorities to trust and use it.

In terms of types of decision-making to be supported by the AI-Tool, the following were identified:

- Preparations by Member States that receive incoming irregular migrants;
- Preparations by Member States that are affected by secondary movements;
- Resourcing asylum management operations;
- Application of measures related to all components of IBM (in all four tiers of the IBM access control model);
- Contingency planning and allocation of financial resources at EU as well as at national level;
- Operational crisis management; and
- Political and operational international cooperation concerning irregular migration.

3.3 Comparative analysis of other already developed AI-Tools

Some conflict and crisis forecasting systems already exist in the EU or are being developed at a prototype level. Some of these were developed by EU JHA agencies and others by Member State institutions. Such conflict and crisis forecasting systems can be considered to be useful for irregular migration forecasting to some extent, as theory and experience suggest that conflicts and crises can influence irregular migration patterns.

None of the already existing systems and AI-Tools, however, fully cover the objectives set out for the envisioned AI-Tool covered by this study. The closest AI-Tool related to the objectives of this study is EASO's Early Warning and Forecasting System, which forecasts the number of asylum applications which EU Member States can expect up to three weeks in advance. This is done on the basis of monitoring and forecasting crises in third countries, whereby events in third countries are fed into scenarios based on correlations between the occurrence of similar prior events and numbers of asylum applications in the EU. On the basis of such assessments of previous occurrences, alerts are issued once a newly detected, similar event increases.

Similar to the purpose of the AI-Tool covered in this study, the already existing AI-Tools are and were intended to support the operational preparedness of recipients of the systems' outputs. Referring to the example of EASO's Early Warning and Forecasting System again, its outputs in terms of forecasted numbers of asylum applications allow recipients of this information to adjust their overall preparedness if need be, including the planning of resources and other activities within asylum management. However, it is also noted that the outputs of this system would have to be complemented by another more qualitative report to ensure such information can be effectively used for policy decisions.

Stakeholders developing and using these different AI-Tools have reported that the quality, frequency and timing of the data they use, especially those coming from open sources, often present limitations (see Table 3.1 below). To counteract limitations, it might be required to have in place numerous monitoring, review and/or pre-processing steps, which can sometimes prove to be costly and time-consuming. For example, this can refer to the conversion of unstructured data, such as images, weather data, geospatial data, etc. into structured data, whereby all data sources are moulded into the same format to enable analysis. In this regard, it must be noted that few data sources are by default optimised for the use by an AI-Tool as covered by this study, and more so, that as soon as various types of data sources are used, processing steps might be required for a broader range of input data. Thus, the building of an ambitious AI-Tool, i.e. one that would incorporate various forecasts and/or forecasts across several points in time, would require investment in data sources which are either already explicitly structured or processed accordingly, to support it.

The following table provides an overview of relevant forecasting and AI-Tools, which currently exist or are being developed in the EU. It thus excludes any systems which might exist out with the Eu. Considering that no AI-Tools with a specific focus on forecasting irregular migration currently exist within the

EU, this table outlines those AI-Tools, which are closest in scope and similar in intended purpose, namely informing decision-making processes, measures and operations at the EU or Member State level. It was compiled on the basis of desk research and interviews with relevant stakeholders, with a view to enabling a good understanding of AI forecasting tools that present similarities in terms of outputs, data sources it uses, operational and technicalities, et cetera, so that lessons learned from previous experiences can be incorporated into the specific assessments of this studies.

Table 3.1 Overview of strength and weaknesses of AI-Tools which already exist or are under development

Name	Type of predictions/decision-making	Data sources	Strength	Limitations
The Foresight Project	Predict forced displacement of internally displaced people and refugees, with an initial focus on Afghanistan and Myanmar but more recently applied to the Sahel	• Retrieved data from more than 120 open data sources (publicly available); • a machine learning and bayesian network model for predicting total forced displacement from a given country one to three years ahead in which prediction is based on historical data;	• Tool can be adapted to different scenarios, including Covid19 pandemic; • Can differentiate events in certain regions; • Uses publicly available data to generate algorithms; • Strong and wide knowledge of management and dissemination.	• Limited information on the generated data; • Also, generated data is not publicly available, raising transparency issues for a wider public; • Forecasting function only complementary since it does not fully predict irregular migration flows (since based on historical data); • Policies and guidelines needed to provide a foundation for using the model.

Name	Type of predictions/decision-making	Data sources	Strength	Limitations
		• The model is generic and builds on historical data for 28 countries with a history of displacement; • This informs scenario-building by showing how the various aspects are interrelated.		
Early Warning and Forecasting System (EWFS) (EASO)	Number of asylum applications to plan resources and actions on border and asylum management (preparedness).	• Media-covered events in third countries (GDELT); • Google search data in third countries;	• The AI-Tool can differentiate events in certain regions, not just in a country. This can also include	• Cannot predict, only recognise big events (war, conflict, pandemic) that would have a big impact on irregular migration flows; • Although the system makes the factors that impact irregular migration flows very explicit, it was reported that it does not provide full added value for some stakeholders in their decision/policy-making, as it only

Name	Type of predictions/decision-making	Data sources	Strength	Limitations
		• Frontex open-source data on illegal border crossings; • Data on EU asylum applications and recognition rates; • Use data from back to 2016.	situations in a specific city; • Successful in prediction of the number of asylum applications.	complements situational awareness alert analysis; • Relies on events reported from the media from across the world. Thus, if local/national media in certain countries is controlled by the government and consequently certain events are not reported, a bias results, which, however, might to some extent be remedied by complementary reporting by free media outlets of other countries; • In terms of pull factors, it relies on irregular migrants having access to positive events and information about the destination country; however, this might not be the case for all irregular migrants, considering the often-limited access to accurate information online for prospective irregular migrants or irregular migrants already in transit. In addition, it is often the case that irregular migrants have access to a false perception of how refugee-friendly policies in destination countries are which can influence their decision-making; • The AI-Tool is unable to track internet use of people using VPNs (like in Syria);

Name	Type of predictions/decision-making	Data sources	Strength	Limitations
				• Works with an important assumption that (prospective) irregular migrants from third countries intend to come to the EU, which is not necessarily the case with all irregular migrants.
JRC Research on Migration Modelling, Global Conflict Risk Index and Dynamic Data Hub	Now-casting on irregular migration trends for policymakers on preparedness (not clear what type of decision).	• Past migration data; • Datasets from IOM, GMDAC, EU JHA agencies such as EUROSTAT and Frontex, Facebook (open source for internal use); • Some data sources are restricted to internal use and not shared on	• Strong and wide knowledge management and dissemination.	• Usability of the AI-Tool for policy-makers who need trainings and time to be able to exploit the full potential.

Name	Type of predictions/decision-making	Data sources	Strength	Limitations
		the platform.		
PREVIEW (German government)	There are two AI-Tools: the first is used by the Ministry of Defence, and the goal is to do crisis forecasting to plan resources and actions (preparedness). The second is used by the Federal Foreign Office, however, at this stage, the AI-Tool is not used for irregular migration forecasting but rather provides monitoring and analysis of the current crisis and to some extent, nowcasting. The modelling and predictive analysis are still under development. Outputs can be used for contextualising current crisis and their development and how to prioritise certain countries for action, as well as what country should be prioritised for prevention-oriented actions in the	• Open data from UN agencies, World Bank, academia, think thanks, GDELT, ACLED, mostly structured data, but slowly moving towards unstructured data with news reports for instance; • A policy decision was made on not using social media data.	• Supported some data providers in improving the update frequency of their data.	• Availability, coverage and aggregation of data. Some data are only available from 1-2 years ago (e.g. some of UNHCR data); • It is very expensive to use natural language processes to process unstructured data.

Name	Type of predictions/decision-making	Data sources	Strength	Limitations
	longer-term. Risk of conflicts in the next 1 to 4 years to select countries for conflict-prevention actions for relevant MS/EU.			
EEAS EU Conflict Early Warning System	Risk of conflicts in the next 1 to 4 years to select countries for conflict-prevention actions for relevant MS/EU.	• JRC Global Conflict Risk Index; • EU in-country delegation insights; • Country intelligence; • Open data sources from UNHCR, international organisations and EU JHA agencies.	• Put the spotlight on countries that are maybe not central on conflict angle; • Foster good internal and external coordination and communications.	• The sensitivity of the topic: Resistance from delegations (who are responsible for government relations) to be labelled as a country on the early conflict list; • The model puts together a prediction of conflict at 1 to 4 years ahead, but does not precise if the forecast will be on year 1, 2, 3 or 4; • Major limitation on updating the dataset. The AI-Tool is now using data from 2018. If something has dramatically changed, the AI-Tool is not able to take it into consideration. This, therefore, implies numerous and thorough qualitative assessments and checks.

3.4 Challenges and clarifying purposes

As indicated previously, various challenges and necessary factors to the development of a designated AI-Tool to forecast irregular migration movements can be identified on the basis of research into already existing AI-Tools and on the basis of stakeholder interviews. These will be discussed in turn below, as they enable a better understanding of the coordination of irregular migration data at the EU level and how these challenges could be mitigated with the development of an irregular migration forecasting AI-Tool. These insights also feed into the operational and organisational assessments.

Understanding the purpose of the output while creating it: Challenges are noted in relation to the understanding of the ultimate use of outputs created by current irregular migration data analysis systems at the preparation stage. In this regard, a lack of clarity was indicated by some EU JHA agencies which are tasked to produce irregular migration forecasts or related analyses, on which types of (policy) decisions they are ultimately supporting. This makes analytical processes less efficient and also increases risks that the analysis is targeting the wrong operational, policy or political issue. For a potential future AI-Tool, the lesson that can be drawn from this is that the decision-making context or purposes must be clear.

(Pre-)processing of data to increase usability: As indicated above, unstructured data sources, such as qualitative input, images, weather data, geospatial data, and others, require processing before they can be fully incorporated into an AI-Tool. This might impact the timeliness of the output and associated costs and needs to be factored in when managing expectations about the timing and timeliness of the output of the AI-Tool towards its users.

Collaboration and data integration: Barriers to data sharing among agencies were identified. This can reduce opportunities for informing policy or operational decisions. This often results in a lack of available relevant and accurate data for some of the actors, repetition in the usage of the same data, as well as gaps in accessing current data. Furthermore, it was noted that the EU currently lacks a central data management system. The Central Repository for Reporting and Statistics being developed by eu-LISA might potentially address some issues by making available the data that are in different systems such as Eurodac, VIS, EES and ETIAS. This, however, also depends on any future changes to existing mandates and entailed rights to access certain data. In a scenario based on an assumed broadened data access, the AI-Tool would have direct and immediate access to all its data inputs. If no relevant changes to mandates are made, the AI-Tool will rely on multiple agencies to collect, process and forward data.

Restrictions on data sharing: Formal restrictions on sharing data contribute to data fragmentation. These restrictions were identified on the basis of different mandates of EU JHA agencies. For example, eu-LISA is not able to access data belonging to Member States and EASO does not have access to the EURODAC, data which is key to the analysis (and forecasting) of secondary movements. Again, the Central Repository for Statistics could possibly address this issue. A complementary or alternative approach would

be to invest in anonymisation and hashing methods so that detailed primary data can be shared while minimising privacy or security risks.

Reporting timeframes: Reporting cycles appear to differ among Member States and EU JHA agencies. EUROSTAT, for example, has a reporting period of 2 to 3 months, whereas Frontex uses a two month reporting period. This may be down to various reasons but is crucial for expectation management in terms of both the data providers and the receivers of the output of the AI-Tool into which such data is integrated. Additionally, not all Member States report their data at the same points time or within the reporting period itself. This might distort the underlying analysis of the AI-Tool accordingly. Further, especially in cases of data being provided late, data processing and integration has to be done in a more fragmented manner, rather than all at structurally at once.

Training and understanding: Understanding the strengths and limitations of analyses drawn from complex statistics or machine learning systems might, in some cases, be considered challenging, considering the relative novelty and overall complexity of such systems. Hence, to fully understand the context, input, and underlying analytical steps towards the outputs of such an AI-Tool, training on the interpretation of AI-Tool output for decision-makers will support the analyses and use they can make of the outputs. More details on hosting and coordination of a potential AI-Tool can be found in the Organisational Assessment.

Feedback: For the AI-Tool to be useful, its immediate users, such as data analysts, and its end-users, i.e. recipients of the AI-Tool's output, should have multiple opportunities to provide feedback on the areas of their involvement in relation to the AI-Tool.

3.5 General feasibility of irregular migration forecasting

In line with the academic literature, the overall predictability of irregular migration depends, among others, on i) the **quality of the underlying data available**, ii) on the **complexity of key drivers**, as well as iiI) the **time horizon** considered.[17] The relevance of the data source furthermore depends on the **type of required forecast and policy purpose**. A growing amount of literature is dealing with the question of whether and how irregular migration forecasting can be made (for a comprehensive and state-of-the-art literature overview see **Error! Reference source not found.**). The main challenges, opportunities and findings of the data assessment are briefly outlined in the section below.

Early warning systems have the potential to change decision-making from a reactive to a proactive manner. To have a well-functioning early warning

[17] Carammia & Dumont, 2018.

system, there should be a good understanding of trigger points to minimise false negatives and false positives. Early warning systems, however, do not provide information about the drivers and causes of irregular migration. Moreover, these systems only tell something about the likelihood of irregular migration and do not contribute to the explanation of irregular migration trends.[18]

Forecasting models employ data-driven approaches to understand future irregular migration flows in the medium to long term. These quantitative models tend to be explicit in how assumptions within a model can affect future irregular migration flows. In other words, they are transparent on how underlying principles and guiding theories are used for analysis to estimate future flows of irregular migration.[19] (Forecasting models are less limited in the time span they can cover, as past data trends can be extrapolated for many years in the future.

However, forecasting is a difficult task and can produce rather unreliable results.[20] This firstly has to do with the lack of uniform concepts and definitions for irregular migration and irregular migration-related issues.

The data may also, be unreliable, especially in developing countries where empirical evidence about past irregular migration may be incomplete or even entirely missing.[21] Moreover, by focusing on a limited amount of pre-defined drivers of irregular migration, many potentially relevant drivers, including unpredictable ones, are left out of the analysis. The diversity of motives behind irregular migration flows and the emergence of new types of irregular migration, make it difficult to actually forecast irregular migration.[22]

In addition, the defined drivers of irregular migration are based on assumptions regarding demographic dynamics, the political, environmental and socio-economic changes as well as migration policies, which continuously influence each other and cannot be observed as separate and non-evolving drivers. Models are in this way vulnerable to unpredicted shocks that move away from their assumptions, as the sheer number of push and pull factors and drivers of mobility all interact with each other, thereby making a comprehensive estimate of irregular migration hardly possible.

Foresight models then are useful for decision-makers, as it allows for irregular migration scenario planning. The qualitative scenario creation approach questions assumptions on irregular migration and lets decision-makers imagine, anticipate and prepare for uncertain and unexpected future

[18] EASO, *Quantitative Assessment of asylum- related migration*: *A survey of methodology* (Luxembourg: Publications Office of the European Union, 2016); OECD and EASO, *Migration Policy Debates: Can we anticipate future migration flows?* (Paris: OECD).
[19] Sardoschau, 2020.
[20] Bijak, 2016.
[21] Buettner and Muenz, 2016.
[22] Bijak, 2016.

irregular migration trends. [23] Moreover, it allows for long-term and future-oriented thinking. By involving experts in a systematic, iterative and participatory way, short-sighted policies can be avoided and stimulate long-term strategic thinking among decision-makers, rather than merely providing operational input.[24]

However, when consulting experts, it often appears that there is large disagreement among these experts on how drivers of irregular migration will affect future irregular migration flows. In addition, the migration scenario reports sometimes produce ambiguous results on the impact of future drivers of irregular migration on irregular migration flows.[25] In other words, the theoretical model in qualitative scenarios tends to be less explicitly expressed in contrast to quantitative forecasts.[26] This combined makes it hard to design concrete and actionable policies based on the strategic scenarios, especially in the short- and medium-term. Moreover, just like for forecasting models, the world that we live in is rather complex and uncertain, and the longer one plans in the future, the more uncertainty is involved.

Complexity of key drivers of irregular migration and real-world dynamics of irregular migration

All approaches hinge on the underlying data that they use, such as experts, administrative data, surveys or censuses. Forecasts in the field of irregular migration appear particularly difficult given the complexity and diversity of migration processes, the limited availability and quality of data, and the limited understanding of the drivers of irregular migration.[27] Due to the complexity and volatility of international irregular migration flows, they are difficult to anticipate. Still, some types of movements are more stable or more regulated than others, and therefore more predictable. The most difficult migration types to forecast are forced and irregular movements.[28] Modelling is possible when rich numerical data are available, for example, on past inflows and outflows, policy changes, as well as various other irregular migration factors and drivers.[29] There is relative consensus in the literature on some drivers and relative divergence on others.[30] Relative consensus consists of the relevance of socio-economic factors and political factors such as violence, insecurity, instability and fundamental rights abuses. Relative divergence consists in the relevance of demographic factors, historical cultural and geographic proximity, the impact of networks, environment, climate change and natural disasters as well as migration policy and economic pull factors. [31]

---

23 Vezzoli et al., 2017.
24 Acostamadiedo et al., 2020.
25 Sohst et al. 2020.
26 Sardoschau, 2020.
27 Acostamadiedo & Dag Tjaden, *Forecasting the future of migration—many approaches, one commonality: uncertainty* (Migration Data Portal Blog, 2020).
28 Carammia & Dumont, 2018.
29 Sohst et al., 2020; Bijak, 2016.
30 Bijak, J., Forster, J. and Hilton, J., *Quantitative assessment of asylum-related migration: A survey of methodology* (EASO, 2017).
31 EASO, 2016.

The effect of individual drivers varies over time and space (so across irregular migration flows and within the same flow over time), which is what complicates irregular migration modelling and forecasting.

Time and space

It will likely vary across **time and space** (same flows at different points in time, or different flows).

In the area of irregular migration, short-term early warning and alert systems can have medium-to-high predictability (although in practice, more likely medium). In contrast, medium-term models and risk analysis have low-to-medium predictability. The interviewees during the Inception Phase also indicated their interest in longer-term forecasting, which has low predictability. Still, if matched with scenario analyses, can offer some insight into future irregular migration trends that can inform strategic (rather than operational) decision-making. Scenarios are long-term for 1 to 5 years or even longer (policy cycles) and are necessary for capability development (see Art. 9 EBCG).

Definition of output and prediction variables

Another point which needs to be taken into consideration is how exactly the output is defined (e.g. irregular migration, asylum applications, secondary movements, etc.) and what the prediction variables are. Generally speaking, the **broader the prediction variable**, the more likely that the decisions are strategic, political and over a longer period. The **narrower the variable**, the more likely that the decisions are operational and over a shorter period. For example, assuming the prediction variable is: increase, decrease or no change in irregular arrivals expected at the EU's external borders six months from now. This variable has a broad geographic scope, so predictions will not help much in deciding where resources could be needed. It would probably be most useful in alerting policy and political decision-makers to changes in the level of attention they should give to external borders and cooperation with neighbouring countries.

By contrast, let us imagine the prediction variable is: The number of irregular migrants who will arrive at Greece's external borders from visa obliged countries next month without a valid visa and cross. The variable has a narrower geographic scope and is more directly relevant to specific operational decisions. It would also provide input into cooperation with Turkey. However, it obviously will have limited relevance to decisions in Spain or Italy, since it will not be outputting anything in relation to the Western or Central Mediterranean Routes. Similarly, it would not help much in strategy and diplomacy regarding Libya or countries of the Sahel.

Predictability of events

No matter how accurate and well-performing a model (or algorithm) is, some events are just unpredictable. The effect of individual drivers varies over time and space (so across irregular migration flows and within the same flow over time), which is what complicates irregular migration modelling and

forecasting. Hence, no approach will be able to predict them, and the longer the horizon, the more difficult to predict those types of potentially disruptive events. There are three main types of uncertainty in migration that challenge migration modelling. First, uncertainty inherent to any forecasts; secondly the uncertainty specific to migration and thirdly the uncertainty in migration data.[32] However, models of high quality are able to approximate those events, especially when they are close in time; or, adapt to them – in the sense of catching up early with those changes in the system and adjusting the forecasts early on. In academic literature, forecasting of conflict events is found to be rather difficult, due to both technical and practical reasons.

When predicting political risk, there is too much uncertainty involved, as different political actors have different hypotheses about the future and different strategies. Political actors also do not want their strategies to be too onefold, as this enhances the predictability of its strategy for adversaries. Predictions can therefore turn out to be false, as actors change their behaviour, thereby undermining the reliability of risk forecasts. Small sparks can turn into an explosion, but it is usually easier to trace back the spark than to predict the explosion. In other words, the predictive validity of a model may take years or even decades to develop and to be evaluated. This why databases of conflict events usually score 'low' or 'medium' (and not 'high) in this assessment.[33]

3.6 Data source assessment

For the data source assessment, approximately 60 potentially relevant data sources were **initially identified and mapped** through desk research, as well as the input received during the stakeholder interviews and the closed-door seminar. In a second step, this was narrowed down to a selection to 39 data sources deemed as **particularly relevant** and further analysed them based on their **content, scope, legal feasibility, relevance and other criteria** as presented in Table 3.2 below. These data sources and analyses relating to the above-mentioned criteria are presented in boxes in Annex C. An overview of the characteristics of these data sources is provided in chapter 3.6.1.

In chapter 3.6.2, the data sources from Annex C were sorted into the **three identified forecasting categories A, B, and C**, each of which seeks to forecast different events in different locations, at different points in time along with irregular migratory movements from the country of origin to the destination country. As such, **forecasting category A** begins with forecasts of events at the earliest stages of irregular migratory movements, namely the forecasts of potentially critical situations in third countries and the EU, e.g. development of underlying socio-economic and political factors in a given

[32] Jakub Bijak et al., "Assessing Time Series Models For Forecasting International Migration: Lessons From The United Kingdom", *Journal Of Forecasting* 38, no. 5 (2020): 470-487; Jakub Bijak and Mathias Czaika, *Assessing Uncertain Migration Futures – A Typology Of The Unknown*, Deliverable D1.1: Quantmig Project (Southampton: Krems: University of Southampton, Danube University Krems, 2020).

[33] Chadefaux, 2017; Gartzke, 1999; Peteranderl, 2019.

country and/or sudden events, which (jointly) could have an impact on setting off irregular migratory movements from such countries. As such, this forecasting category offers output on the drivers of irregular migration towards the EU.

The next **forecasting category B** begins where forecasting category A leaves off, and as such first incorporates data informing the forecasting of irregular migration flows, including which routes are taken and the emergence of new irregular migration routes, also using outputs of forecasting category A to this end. On the basis of this, irregular border crossings into the EU are ultimately forecast within this category. Hence, forecasting category B covers the events between irregular migrants' setting off from countries of origin along routes into the EU. It thereby incorporates a mix of drivers and outputs, but ultimately offers output on levels of (change in) irregular migration arrivals to EU border countries.

Eventually, **forecasting category C** continues where B leaves off and covers the forecasting of events within the EU. For one, this covers the events following irregular border crossings into the EU, but it also includes the phenomenon of overstayers, whereby persons enter the EU legally but overstay the period of their visa or residence permit. As such, forecasting category C includes the prediction of levels of (change in and location of) asylum applications lodged in the EU, secondary movements towards destination countries within the EU, and deviations from regular migration patterns. Forecasting category C necessarily builds on the outputs of forecasting categories A and B, but also on information on previous trends and other data on actual detections and applications provided by the EU Member States.

The various forecasts based on the above categories across different points in time of irregular migrants' movements enable the EU Member States and relevant EU JHA agencies to plan their capacities accordingly. This can entail adjusting reception capacities, but also more focused detection activities.

After establishing which data sources are relevant for which forecasting category, the following chapter 3.7 focuses solely on forecasting categories and data sources which are required for the **forecasting of irregular border crossings**, as an intermediary step towards the case study in chapter 1, which narrows done further and focuses on irregular border crossings at the Greek sea and land borders in particular. As such, forecasting category A is featured, as it informs forecasting category B, which ultimately covers the actual forecasts of irregular border crossings. Meanwhile, forecasting category C is not covered, as it only forecasts events taking place after irregular border crossings have occurred. In this section, the data sources relevant for forecasting categories A and B are sorted into high, medium, and low relevance for the forecasting of irregular border crossings. Practically, this means that data sources offering supporting evidence on trends may feature at a lower relevance, than data sources offering insights on activities in the EU external border region, or along routes. However, the ultimate usability and relevance of each data source can only be determined during technical testing of the AI-Tool, while at this stage the relevance of data sources is only assessed theoretically in an ex-ante setting. Relevance is measured by the potential degree of usefulness and contribution to a forecasting tool.

As a final step, in chapter 4, the data sources were narrowed down to those relevant to **Case Study 1, i.e. the forecasting of the number of irregular**

migrants who will arrive at Greece's external borders from visa obliged countries next month without a valid visa and (attempt to) cross either the land or sea border. This case study serves to provide a concrete example of how a forecast would be produced by the AI-Tool. It further illustrates modelling and design processes. The findings of the case study and the elaborations of the final report will support and guide the European Commission in their decision on the next steps of the development of the AI-Tool.

To begin with the first step of sorting the list of data sources from Annex C into the three forecasting categories, the table below first shows the criteria chosen for assessing the data sources within Annex C.

Table 3.2 Criteria for assessing the data sources

Criteria	Description
Accessibility	Accessibility refers to open and restricted (e.g. by commercial provider or government) data sources.
Cost of accessing the data	Purchase Price indication (free/costs).
Data Reliability	Reliability in terms of the data source's ability to accurately describe their objective to forecast irregular migration and its precision in doing this.
Data availability/accessibility per stakeholder (including Frontex, EASO, eu-LISA, Europol)	Accessibility of data source per stakeholder.
Frequency of being updated	Frequency refers to how often the data is updated and what often the time scale is (daily, weekly, monthly, yearly, etc).
Fundamental rights implications	• To be aligned with the EU Charter of Fundamental Rights as it is a primary source of law as provided by Art.6 (1) TEU. In particular, the following ones; • The right to non-discrimination; • The right to asylum; • The right to respect for private and family life; • The right to protection of personal data; • The right to good administration.
Indicators (non-exhaustive)	• Socio-economic factors in countries of origin (economic development of the country, GDP per capita, demographic change, unemployment levels, level of education, etc.); • (Economic) situation in destination countries (labour market conditions for irregular migrants, level of state support, level of welcoming, etc.); • Size/presence of diaspora in destination

Criteria	Description
	countries and associated networks and family links with persons in countries of origin; • Political risk, tensions, or levels of stability in countries of origin, transit or destination country; • Historical/trend data on (irregular) migration and net migration; • Annual remittances from members of a country's diaspora to persons in their country of origin; • Inflows and outflows of foreign population by nationality; • Expression of migration intention by persons in countries of origin (to be seen in combination with enabling socio-economic factors, such as GDP per capita, unemployment levels, demographic change, etc.); • Extreme weather conditions (drought, flooding); • Number of alerts from EU Immigration Liaison Officers (ILOs) and Frontex staff deployed in third countries; • Schengen visa applications and refusals (short-term) in the Schengen States and visas issued by other EU states (Bulgaria, Romania, etc.); • Activities and patterns on identified irregular migratory routes; • Number of smuggler boats on the Mediterranean; • Number of detected irregular migrants in territorial waters of third countries; • Number of SAR incidents; • Number of illegal border crossings; • Vulnerability of border areas along countries of origin, transit (incl. EU external border) and destination in terms of border crossing points' level of surveillance; • Number of Eurodac registrations for both irregular border crossings and lodging of asylum applications; • Number of irregular migrants in reception centres; • Number of asylum applications; • Refusals of national visa or resident permit applications; • Number of refusals of Schengen visa extension applications; • Number of overstayers (after EES implemented); • Increased use of certain modi operandi, e.g. car rental contracts with large mileage numbers to

Criteria	Description
	facilitate transportation from and to destination countries; • Number of irregular migrants and refugees settled in third countries close to the EU external border; • Number of irregular migrants in refugee and/or reception camps in EU countries; • Number of irregular border crossings from Non-Schengen to Schengen country.
Structured/unstructured	**Structured data**: Highly-organised and formatted in a way so its easily searchable in relational databases. **Unstructured data:** No pre-defined format or organization, making it much more difficult to collect, process, and analyse.
Timeliness	Timeliness indicates that data are made available shortly after the reporting period.
Type of data source[34]	**Statistical** (Data for the creation of official statistics (e.g. Censuses, Household surveys). **Administrative** (Data primarily to support administrative processes rather than to produce official statistics, e.g. Visa, residence-, work-permit, and often provided by official entities). **Innovative** (Emerging sources of migration data, e.g. big data or IOM's Displacement Tracking Matrix).
Forecasting category	**A** is forecasting potential critical situations in third countries and in the EU, which could have an impact on setting off migratory movements (i.e. drivers); **B** is forecasting two subsequent steps, namely 1) irregular migration flows, including levels on established routes and the emergence of new migration routes, and 2) irregular border crossings into the EU. Hence, events between setting off from countries of origin into the EU are covered in this category; **C** is forecasting events following irregular border crossings into the EU, such as asylum applications lodged in the EU and secondary movements.
Operationalisation of the data source	Practical usability from a technical perspective.
Overall assessment of relevance	Relevance of the data source for the purpose of the feasibility study.

[34] For further details on the different type of migration sources, please see: IOM, *Global Migration Indicators* (GMDAC, 2018).

Criteria	Description
Related AI technique	Type of AI model which can be applied (e.g. Natural Language Processing, machine learning, etc.).
Ranking	Sorted into high, medium, and low relevance, including a traffic light system on associated potential risks.
Relevance for forecasting time-lines (short-, mid- and long-term)	Early warning (short term – weeks) – covered by forecasting categories B and C; Forecasting (medium-term – months) – covered by forecasting categories B and C; Foresight (long term – years) – covered by forecasting category A.
Location of the data source server	Inside of the EU; Outside of the EU.

As outlined above, the following section first presents an overview of the distribution of the relevant criteria across the dataset of relevant sources as per table 3.2 above. This serves to present a picture on the types of data sources available in the first place, their level of accessibility, financial implications, reliability, scope, and more.

3.6.1 Data assessment findings: general overview of data source characteristics

Beginning with the general overview of the analysis of available data sources, it is first and foremost noted that the different data sources have significantly varying characteristics. These range from qualitative to quantitative data, from historical to current data, as well as from less frequently to frequently updated data, and more. As will be shown further below, these variations all inform different purposes of the AI-Tool, in that not all data sources are useful for every type of forecast or time horizon.

Indicators

Beginning with the question of indicators, which will be informed by the various data sources, the table above outlines a range of several types. For one, these include a non-exhaustive selection of changes to various conditions, including socio-economic factors, such as income levels and demographic change, and political situations, which are among the main **drivers** identified in irregular migration-related literature. Likewise, this also includes data on the diaspora and network effects, whereby irregular migrants particularly aim to reach destination countries in which they already have personal networks or in which groups of their own nationality are already present.

There is wide agreement on some factors being of key relevance in terms of push and pull factors, such as the level of conflict, political instability, socio-economic crises, insecurity, and fundamental rights abuses. Less agreement exists on the relevance of other factors for forecasting irregular migratory movements, such as demographics, climate change, or policy changes, though they are regularly brought forward as drivers.

Besides these common drivers, the list of indicators further encompasses **direct indicators**, such as migration alerts issued on the ground.

Additionally, a range of indicators relate to changes to **facilitating circumstances and activities throughout irregular migration flows**, such as activities of smugglers. Further, several indicators on **on-going irregular migration flows** are listed, which offer updated, and primarily quantitative data, such as on the numbers of incidents, settlements near borders, detected irregular border crossings, and more.

This mix of indicators and data types goes in line with findings which can be derived from already existing AI-Tools, which also incorporate a variety of data types, ranging along the whole spectrum of historical data to frequently updated radio satellite data, to generate algorithms. The rationale behind this is that the data sources used by the AI-Tool must cover the content of such indicators whose changes are closely linked to setting off irregular migratory movements, and which are relevant in monitoring and forecasting their direction. If relevant indicators are monitored closely, and changes are noted, estimates can be generated on the likelihood of (further) irregular migratory movement, and warnings can be issued accordingly. Indicators not only apply individually but also in combination with related ones. Unfortunately, migration literature offers an only limited understanding of interdependencies between different indicators, although drivers are well-identified. In determining routes of irregular migration, historical and geographical aspects play a major role, however, so do short-term changes in policies (e.g. sudden border closures leading to re-directions of routes) and possibilities (e.g. availability of new or non-conventional route through facilitators), which are challenging to forecast and may only be factored in after their announcement, or (detection of) occurrence. Hence, the mix of drivers and indicators along irregular migratory routes are of particular relevance for the development of an AI-Tool.

Qualitative and quantitative data

In terms of **qualitative data**, inputs of experts and interviews on irregular migration flows and routes serve as adequate qualitative data sources in order to understand the actual motives and reasons of irregular migrants to emigrate to different locations. However, due to the sensitive and highly politicised nature of the topic, such interview data is often restricted. Nonetheless, such data is collected by relevant EU JHA agencies, such as Frontex and EASO. Hence, such data sources being restricted does not necessarily present a challenge in terms of access to them.

The analysis of qualitative data and the generated risk assessments based on the inputs, however, might be prone to interpretation bias since the reports are based on subjective interpretations of statements. **Quantitative data** is presented in a mix of historical data (see more about this below) and data monitoring on-going and directly migration-related factors, such as border crossings. The combination of qualitative and quantitative data is important for the AI-Tool, considering that qualitative data can inform the context of quantified data, and thereby make the datasets more informative. For example, in terms of the context of irregular migration movements, qualitative data could provide insights on the motives and reasons for irregular migration, which in turn serves to sharpen the indicators regarded in the situation in question.

Output of data sources

In terms of output of data sources, nearly half of the data sources assessed for this report are **statistical sources**, while the other half of sources range from **innovative types** (one-sixth) to **administrative sources** (one quarter), and to **mixed sources** (one sixth). Considering the prevalence of statistical output, further insight can be offered on these types of data sources. Here, particularly so-called social statistics offer insights on conflict zones, displacements, irregular migration flows, political violence and protests across regions. They rely themselves on a variety of data sets, ranging from qualitative conflict assessments to numerical assessments of dates, fatalities and frequency of armed conflicts, as well as categorical data in which protest types events are regularly assessed. Likewise, innovative sources refer to real-time news updates and already existing niche predictions on relevant aspects, such as the occurrence of conflict in a given region. Clustered among administrative sources are those provided by official institutions and entities, such as reports by embassies and delegations, and other data sources and outputs provided by Member States and EU JHA agencies.

Providers of data sources

As for providers of data sources, nearly three-quarters of the assessed data sources are put at disposal by either non-profit or public sector organisations, by international or regional agencies and organisations, or by the EU Member States. Around a quarter of the assessed data sources is also offered by private sector providers.

The nature of the provider of data sources influences the availability of data sources to some extent. In this regard, most of the sources stemming from the public sector are also publicly available (open source) via databases or other forms of dissemination. Meanwhile, relevant data provided by entities such as the EU JHA agencies or Member State institutions is necessarily of restricted nature by not being available to the general public due to their relating to sensitive information. This includes both data from law enforcement agencies, but also outputs of other relevant and already existing similar projects and systems to the envisioned AI-Tool, such as the EASO Early Warning and Forecasting System and Germany's PRISM. However, as the question of data availability must be regarded from the perspective of the European Commission, general levels of restriction by EU JHA agency data providers may either not apply, or may be subject to further discussions on dissemination policies, i.e. seeking data sharing agreements and/or expansions of access to better inform the AI-Tool. In this regard, it can also be differentiated between products offered by EU JHA agencies, such as qualitative reports based on, e.g. SIENA data, and the data sources on which such outputs are based. Data disseminated by EU Member States to particular EU JHA agencies or other recipients for particular purposes may also be subject to discussions on respective dissemination policies for the purpose of informing the AI-Tool.

Besides restrictions in availability based on the nature of the data in question and access restrictions in terms of mandates and policies, other data restrictions, particularly by private providers, refer to restrictions in terms of

financial implications. As to the latter, this is not always transparent for all data sources. Nonetheless, most of all data sources are free of charge. In some cases, cost implications depend on the level of access: in such cases, basic access is free, but significant costs occur towards full level access. Further, in a few cases, access to data may be free of charge, but due to necessary data treatment steps required before the data can be utilised, extra costs might apply.

Timeliness

In terms of timeliness, the assessed data sources vary significantly across a range of updates every ten minutes to updates every two years. In many cases, the level of timeliness is unclear or varying depending on particular data sources offered by the same platform. Nonetheless, it can be noted that a large number of the data sources relies on historical data, which is updated on **an annual or bi-annual basis** only, which make them inadequate for short- and medium-term predictions, although they can inform trends for the longer term. Against this background, historical data is of prevalent use in existing predictive algorithms and patterns, in that each data source referring to predictive analytics makes use of statistical analysis of case data to calculate the probability and patterns of irregular migration movements. The known correlations are represented by data sets that firstly have been collected in the past and secondly are also suitable for forecasts. The core assumption relating to the use of historical data is that the same conditions must produce the same result, i.e. that a person behaves identically under identical conditions.

Some **real-time data** are available from databases, while other real-time data depends on the dissemination by agencies and institutions. Real-time data provided by private providers are oriented to current global events and is dependent on steady changing environments. Such real-time databases are regularly updated, often several times per day, and are thus prone to changes over a short or long period of time. It becomes apparent that real/time data is often dependent on **numerical data** in which frequencies of events are counted, or satellite pictures capturing movements and phone signals of individuals. Others, on the other hand, rely on fast-paced news analysis.

Reliability and relevance

Judging the reliability and relevance of data sources is not always straightforward, in that the assessment of the reliability of the content is either proclaimed by data providers themselves, or can often only be based on experiences by previous and long-term users of such data, or own trial and use. Nonetheless, the selection of data sources sought to include only authoritative and established data sources in the first place, which is reflected by the range of well-respected data providers entailed in the selection of data sources.

Another impact on the reliability of data is any potential time lag in which data is reported to or collected by a resource before being published on a publicly available database or in another format or platform. In this regard, some data sources are of much less immediate relevance, since frequent delays of up to several weeks or months can be noted in their reporting, as is the case for example with Member State statistics. Other data sources, particularly in terms of real-time news reporting, are reliable in terms of frequency but may be less so in terms of their content. As such, this is part of a trade-off that

may be faced between timeliness and reliability of data sources. Here, data provided on a fast and frequent basis does not undergo as much scrutiny as data reviewed and published after a certain amount of time. Nonetheless, data available on short notice, even if not fully reliable, may still serve as a crucial indicators for forecasts of the occurrence of events. An example in this regard can be provided involving the two data providers ACLED (Armed Conflict Location & Event Data Project) and GDELT. Here, previous versions of ACLED were of high reliability but scored low in terms of timeliness. Meanwhile, GDELT data was of high timeliness, but also relatively lower reliability.

In terms of relevance, the geographical scope can also be taken into account, considering that it is important to have data available and updated for all relevant countries, including countries of origin, transit countries and destination countries. The data source assessment shows that around 80% of data sources have a worldwide scope, while the rest is restricted to the EU and surrounding relevant countries.

Operationalisation of data sources

Concerning the operationalisation of the assessed data sources, most of the data indicated by organisations are considered to be **completely usable** since the data sets are comprehensive, wide-ranging and are expected to be reliable. In terms of the data formats in which sources are offered, wide variations are noted. First, some data sources offer various formats, while others only offer outputs in one format. Hence, visualisations such as dashboards, charts, graphs, and (interactive) maps, are available by several providers. **Text format** is very prevalent through reports, links, and CSV files. **Excel format** and **numerical datasets** are likewise often available, to an equal extent as access to data via APIs. Looking further at the division between structured and unstructured data, it can be noted that, nearly two-thirds of data providers offer structured data and are as such easily usable by the AI-Tool. The remaining data sources relate to primarily unstructured data, but also some mixed sources. The above sections indicate as such that numerous data sources would be accessible and operational for the purpose of forecasting short- and medium-term forecasting of irregular migration trends using AI.

From an **organisational perspective**, decisions about which data sources the AI-Tool will need to access, on what timeframe, and whether any of these data sources are listed as restricted will also influence considerations of where to host the AI-Tool. The use of confidential data sources would lead to different levels of access for different stakeholders unless these data are anonymised before they are entered into the AI-Tool. In turn, sharing the outputs of the AI-Tool with different audiences will require technical knowledge about data-sharing regulations, as well as a broader awareness of what information (and in what format) is relevant to different stakeholders (e.g. policy teams vs operational units. The preliminary data source assessment shows that in terms of accessibility per stakeholder, three-quarters of the data sources for which this is known provide full access to all stakeholders, while a quarter holds restrictions depending on mandates and levels of clearance.

Legal assessment

From a **legal point of view**, the majority of data sources identified have been deemed feasible. Some may have limitations, related to their terms of service (ICPSR, Google Trends Index-GTI), or the data access might be restricted (for example, data on attempted and successful border crossing attempts). Potential fundamental rights' implications are identified, which revolve around the use of Machine Learning (ML) and resulting effects, such as potential bias. With data sources that include personal opinions, the right to non-discrimination may also be affected. However, only anonymised or aggregated data will be used as input for the AI-Tool. Furthermore, with sources where political agendas are concerned, the neutrality of the AI results might be hindered. Legal implications might also be considered in terms of the **location** in which data sources are hosted. The data source assessment shows that this is not transparent for more than a third of data sources. Of the remaining data sources, the majority of providers store their data in the EU, while the other store their data outside of the EU.

3.6.2 Data assessment findings: data sources per forecasting category

Following the general overview of the characteristics of the long list of data sources in Annex C, the analysis is narrowed down to the selections of data sources required to inform the three forecasting categories A, B, C, which will be incorporated into the AI-Tool. This serves to show which types of data sources would be needed per forecasting category, which is against the background of the AI-Tool – depending on the required output – ultimately requiring information from either, or a combination of two or all forecasting categories A, B and C as input data. Each of the three forecasting categories will be addressed in turn below. First, a brief explanation of each forecasting category is offered, followed by a short summary of the identified data sources, and a table outlining the data source and/or provider, as well as the respective related direct indicators.

Forecasting category A

Forecasting category A refers to the forecasting of potentially critical situations in third countries and in the EU, which could have an impact on setting off irregular migratory movements. This assessment also includes the overall potentials of emerging crisis and conflicts in the future. As such, forecasting category A encompasses drivers of irregular migration, which is reflected in the selection of respective relevant data sources. For one, this is based on (historical) trend data on underlying conditions, such as socioeconomic factors known to be drivers of irregular migration, but also sudden events, ranging from natural disasters to situations of conflicts, policy changes, and more. In this regard, it must be noted that while black swan events may be challenging to predict in the first place, given their inherent definition of being both unexpected and unprecedented, they must nonetheless be swiftly recognised as such upon occurrences.

The table below alphabetically lists 16 data sources which have been identified as being relevant to inform this forecasting category. Nearly a third of the identified sources relates to statistical data. Two-thirds of the identified sources refer to innovative sources, such as particular conflict forecasting tools focus on changes to situations in countries of origin, particularly in terms of political and socio-economic changes, as well as other drivers of irregular migration movements. Other notable innovative sources used in this context also include **real-time data**, such as real-time news data and meteorological data, e.g. on events such as sudden floods destroying livelihoods in countries of origin. Nearly all sources are publicly available.

Table 3.3 Data sources and indicators for forecasting category A

Forecasting category A – critical situations in third countries and in the EU, i.e. drivers of irregular migration	
Data source/provider	**Related direct indicators (non-exhaustive)**
Armed Conflict Location & Event Data Project (ACLED)	• Political risk and tensions in country of origin, transit or destination country.
Country reports by embassies (MS) and EU delegations (e.g. migration sections' reports)	• Political risk and tension in country of origin; • Print, broadcast, and web formats in third-country reporting on political, economic developments and/or migration intention; • (Labour) conditions for irregular migrants along transit countries and in destination countries, level of state support and general level of welcoming.
Crisis Group Reports	• Political risk and tensions in country of origin, transit or destination country.
EU conflict Early Warning System (EWS)	• Political risk and tensions in country of origin, transit or destination country on the basis of structural factors.

Forecasting category A – critical situations in third countries and in the EU, i.e. drivers of irregular migration	
Europe Media Monitor	• Political risk and tensions in country of origin, transit or destination country.
Global Pulse (UN)	• Political risk and tensions in country of origin, transit or destination country.
JRC Global Conflict Risk Index	• Structural conditions in a country (political, security, social, economic, geographical, environmental)
Gallup World Poll Survey	• Expression of migration intention.
GDELT project	• Print, broadcast, and web formats in third-country reporting on political, economic developments and (possibly relevant) events.
Global Terrorism Database (START)	• Political risk and tensions in country of origin, transit or destination country.
Human Development Index (UNDP)	• Poverty level and socio-economic situation in the country of origin (life expectancy, level of education, per capita income, etc.).
Integrated Conflict Early Warning System (ICEWS)	• Political risk and tensions in country of origin or transit.
Meteorological weather data (several providers)	• Extreme weather conditions (e.g. drought, flooding); • Identified migratory routes/patterns; • Number of smuggler boats and routes on the Mediterranean.
OECD International Migration Database	• Presence and size (stock) of diasporas; • Inflows and outflows of foreign population by nationality.
Online search data	• Web search terms about migration intentions, including on routes and modi operandi.
Satellite Pictures (several providers)	• Identified migratory routes/patterns; • Number of settlement close to EU Schengen border.
UNHCR data	• Migration and net migration; • Data on global taxonomies; • Resettlement and population statistics.
Uppsala Conflict Data Program	• Political risk and tensions in country of origin, transit or destination country.
World Bank reports and data on World Development Indicators	• Historical/trend data on migration and net migration; • Population growth (demographic change); • Annual remittances from members of a country's diaspora to persons in their country of origin;

Forecasting category A – critical situations in third countries and in the EU, i.e. drivers of irregular migration	
	• Information on other World Development Indicators.

Forecasting category B

Forecasting category B covers the events between irregular migrants' setting off from their countries of origin, along transit countries, into the EU. This is so, because the output of this forecasting category, i.e. levels of (change in) irregular migrants' arrivals to the EU's external borders, depends on the identification of activities along various irregular migration routes. Hence, to achieve the overall output, the flows along established and new irregular migration routes, i.e. information on transit, must first be established. To this end, data sources, including driver data sources on activities on established and new routes, as well as outputs of forecasting category A are drawn from. Following this step and on the basis of the information established in the first step of this forecasting category, irregular border crossings into the EU are forecast as outputs. Accordingly, a mix of drivers and outputs is incorporated into forecasting category B.

The table below alphabetically lists 32 data sources, which have been identified as relevant to inform forecasting category B. Nearly half of the data sources pertains to **administrative data,** and a further third are **innovative data** sources, which are gathered for operational purposes. As such, the indicators range widely from some informing on changes in countries of origins, to some on activities in transit countries, such as the changes of routes, as well as to sources conveying information on trends and situations in the EU external border region, such as numbers on illegal crossings, apprehensions and SAR incidents, the vulnerability of border areas, numbers of people in refugee camps, data on asylum applications, and more.

While these data sources primarily refer to data on already on-going irregular migratory activities both along the way from countries of origin and in the immediate EU external border region, the data sources nonetheless comprise of a mix of qualitative sources, such as reports and statistical data on medium and long-term trends. This is against the background that the availability of historical data makes more precise forecasts possible and offers a baseline for forecasts, however, it is crucial that such data is enriched with information on dynamic developments within and outside of the EU to take into account sudden developments. Innovative data which particularly informs short-term developments, include real-time news monitors, online search data, satellite imagery, and outputs of other prediction and forecasting systems. One data source which is of high importance, but not mentioned in the table below, is output from forecasting category A, considering that information on drivers and shocks setting off irregular migratory movements are an important factor for forecasting (flows towards) irregular border crossings.

Table 3.4 Data sources and indicators for forecasting category B

Forecasting category B – irregular arrivals to the EU's external borders, incorporating movements along routes	
Data source/provider	**Related indicators (non-exhaustive)**
Conference calls on routes (Europol, national LE)	• Number of smuggler boats on the Mediterranean; • Number of caught migrants in territorial waters of third-country; • Number of illegal border crossings.
Country reports by embassies (MS) and EU delegations (e.g. migration sections' reports)	• Political risk and tension in country of origin; • Print, broadcast, and web formats in third-country reporting on political, economic developments and/or migration intention; • (Labour) conditions for irregular migrants along transit countries and in destination countries, level of state support and general level of welcoming.
Data on (internal) border checks, national visas and resident permit applications (Member States)	• Schengen visa applications and refusals (short-term) in the Schengen States and visas issued by other EU states; • Refusals of national visa or resident permit applications; • Number of refusals of Schengen visa extension applications.
Dublin data from Member States	• Schengen visa applications and refusals (short-term) in the Schengen States and visas issued by other EU states (Bulgaria, Romania, etc.); • Number of asylum applications; • Refusals of national visa or resident permit applications.
EASO Early Warning Reports	• Migration and net migration; • Annual remittances; • Schengen visa applications and refusals (short-term) in the Schengen States and visas issued by other EU states (Bulgaria, Romania, etc.); • Number of migrants in reception centres; • Number of asylum applications; • Refusals of national visa or resident permit applications; • Number of refusals of Schengen visa extension applications; • Number of smuggler boats on the Mediterranean; • Number of caught migrants in territorial waters of third-country; • Number of illegal border crossings.
eu-LISA – Visa Information System	• Data on Schengen visa applications and decisions, i.e. issued and refused visas.

Forecasting category B – irregular arrivals to the EU's external borders, incorporating movements along routes	
Europe Media Monitor	• Political risk and tensions in country of origin, transit or destination country.
European Asylum Dactyloscopy database (Eurodac)	• Number of Eurodac hits for both irregular border crossings and lodging of asylum applications.
Eurostat	• Age; sex; citizenship; country of birth; country of previous residence; level of human development of the country of citizenship, birth and previous residence of immigrants; • Asylum and Dublin statistical trends; changes in enforcement of immigration legislation; residence permits; • Asylum statistics on children; residence statistics on children; enforcement of immigration legislation on children.
Frontex Risk Analysis Reports Consolidated Annual Activity Reports RAU Forecast of Irregular Border Crossings Data on movements between non-Schengen and Schengen Frontex Situational Pictures	• Political risk and tensions in country of origin, transit or destination country; • Number of smuggler boats on the Mediterranean; • Number of irregular border crossings from Non-Schengen to Schengen country; • Number of caught migrants in territorial waters of third-country; • Number of alerts from EU Immigration Liaison Officers (ILOs) and Frontex staff deployed in third countries; • Number of illegal border crossings; • Number of SAR incidents; • Vulnerability of border area.
Gallup World Poll Survey	• Expression of migration intention.
GDELT	• Print, broadcast, and web formats in third-country reporting on political, economic developments and (possibly relevant) events.
ICPRS data	• Political risk and tensions in country of origin, transit or destination country; • Migration and net migration; • Extreme weather conditions (drought, flooding); • Vulnerability of border area; • Number of migrants in refugee camps.
Internal Displacement Monitoring Centre (IDMC)	• Identified migratory routes/patterns; • Number of migrants in refugee camps.

Forecasting category B – irregular arrivals to the EU's external borders, incorporating movements along routes	
Interviews with case workers and migrants at border crossing points and arrival centres (MS)	• Identified migratory routes/patterns; • Migration intention; • Drivers of migration.
IOM's Displacement Tracking matrix system (DTM)	• Identified migratory routes/patterns; • Number of migrants in refugee camps.
MapBox	• Population distribution.
Member States' information on migration policies	• Migration and net migration; • Annual remittances; • Inflows and outflows of foreign population by nationality; • Number of migrants in reception centres; • Number of asylum applications; • Number of overstayers (after EES implemented).
Meteorological weather data (several possible providers)	• Extreme weather conditions (e.g. drought, flooding); • Identified migratory routes/patterns; • Number of smuggler boats and routes on the Mediterranean.
Migration Data Portal	• Migration and net migration; • Activities and patterns on identified migratory routes.
Mixed Migration Foresight Project	• Migration and net migration; • Activities and patterns on identified migratory routes; • Vulnerability of border area.
OECD International Migration Database	• Presence and size (stock) of diasporas; • Inflows and outflows of foreign population by nationality.
Online search data	• Web search terms about migration intentions, including on routes and modi operandi.
Qualitative information by experts on cultural aspects	• Information on aspects of cultural distance between countries of origin, transit and potential destination countries, such as common language(s), colonial links, trade connection, size of diaspora, etc.
Qualitative information on migration policies (e.g. Visa dialogues and liberalisation developments in the Schengen area and in	• Identified migratory routes/patterns; • Social media forecasts (obtained from third parties) in third-country reporting on political, economic developments and/or migration intention.

Forecasting category B – irregular arrivals to the EU's external borders, incorporating movements along routes	
third countries) (European Commission)	
Quantitative information from geographical maps (several possible providers)	• Geographic (border) distance between countries or origin, transit and destination and estimations of duration of time required to go from one point to the next via various means of transport (on foot, by motor vehicle, train, planes, boat, etc.).
Reports by the European Migrant Smuggling Centre Reports (Europol)	• Number of smuggler activities in ports and boats on the Mediterranean; • Identified migratory routes/patterns.
Reports: International open migration data, demographic and social statistics (UNSD)	• Poverty level and socio-economic situation in the country of origin; • Migration and net migration; • Identified migratory routes/patterns.
Satellite Pictures (several providers)	• Identified migratory routes/patterns; • Number of settlement close to EU Schengen border.
Schengen Information System (SIS II)	• Number of irregular border crossings from Non-Schengen to Schengen country.
UN Global Pulse	• Extreme weather conditions (drought, flooding).
UNHCR data	• Migration and net migration; • Data on global taxonomies; • Resettlement and population statistics.
Wittgenstein Centre for Demography and Global Human Capital	• Migration and net migration.
World Bank reports and data on World Development Indicators	• Historical/trend data on migration and net migration; • Population growth (demographic change); • Annual remittances from members of a country's diaspora to persons in their country of origin; • Information on other World Development Indicators.

Forecasting category C

Last but not least, forecasting category C refers to predicting events following irregular border crossings into the EU. This covers the forecasting of levels of (change in and location of) asylum applications lodged in the EU and secondary movements towards destination countries within the EU. Forecasting category C thus necessarily builds on the outputs of forecasting categories A and B.

The table below alphabetically lists 14 data sources, which have been identified as being relevant to inform forecasting category C, i.e. forecasts for occurrences after irregular entry into the EU, such as secondary movements and asylum applications. Nearly half of the data sources are **statistical** sources, while a third is **administrative** data. The sources and indicators encompass a range of systems and databases and qualitative reports, to establish trends relating to the forecasting output on the basis of past and recent data, such as numbers of visa applications, previous levels of asylum applications, as well as other information on (net) (irregular) migration and related factors, such as varying levels of remittances, and the size of diasporas. Further, data on attempted border crossings and apprehensions are included, as well as data on policies by the EU and the relevant Member States along entry points and secondary movement routes.

Further, similarly to forecasting category B, two important sources for forecasting category C are not specifically referred to in the table below, namely forecasting categories A and B. While A informs B, B is particularly important for forecasting category C, considering that in order to forecast occurrences within the EU, such as the numbers of secondary movements and (resulting) asylum applications, it is necessary to know how many irregular migrants have crossed or are intending and attempting to cross the border into the EU in the first place.

Table 3.5 Data sources and indicators for forecasting category C

Forecasting category C – secondary movements and asylum applications	
Data source/provider	**Related indicators (non-exhaustive)**
Country reports by embassies (MS) and EU delegations (e.g. migration sections' reports)	• Political risk and tension in country of origin; • Print, broadcast, and web formats in third-country reporting on political, economic developments and/or migration intention.
Data on (internal) border checks; Data on national visas and resident permit applications (Member States)	• Schengen visa applications and refusals (short-term) in the Schengen States and visas issued by other EU states; • Refusals of national visa or resident permit applications; • Number of refusals of Schengen visa extension applications;
Dublin data from Member States	• Annual remittances; • Schengen visa applications and refusals (short-term) in the Schengen States and visas issued by

Forecasting category C – secondary movements and asylum applications	
	other EU states (Bulgaria, Romania, etc.); • Number of asylum applications; • Refusals of national visa or resident permit applications.
EASO Early Warning Reports	• Migration and net migration; • Annual remittances; • Schengen visa applications and refusals (short-term) in the Schengen States and visas issued by other EU states (Bulgaria, Romania, etc.); • Number of migrants in reception centres; • Number of asylum applications; • Refusals of national visa or resident permit applications; • Number of refusals of Schengen visa extension applications; • Number of smuggler boats on the Mediterranean; • Number of caught migrants in territorial waters of third-country; • Number of illegal border crossings.
Europe Media Monitor	• Political risk and tensions in country of origin, transit or destination country.
eu-LISA –Visa Information System	• Data on Schengen visa applications and decisions, i.e. issued and refused visas.
Europol reports	• Activities and patterns on identified migratory routes; • Number of smuggler boats on the Mediterranean; • Number of caught migrants in territorial waters of third-country; • Number of illegal border crossings; • Number of SAR incidents.
Eurostat	• Age; sex; citizenship; country of birth; country of previous residence; level of human development of the country of citizenship, birth and previous residence of immigrants; • Asylum and Dublin statistical trends; changes in enforcement of immigration legislation; residence permits; • Asylum statistics on children; residence statistics on children; enforcement of immigration legislation on children.
Frontex • Risk Analysis Reports • Vulnerability	• Political risk and tension in country of origin; • Number of smuggler boats on the Mediterranean; • Number of irregular border crossings from Non-Schengen to Schengen country; • Number of caught migrants in territorial waters of

Forecasting category C – secondary movements and asylum applications	
• Assessment of border area(s) • Consolidated Annual Activity Reports • Illegal border crossing data • RAU Forecast of Irregular Border Crossings • Data on movements between non-Schengen and Schengen areas • Frontex Situational Pictures	third-country; • Number of alerts from EU Immigration Liaison Officers (ILOs) and Frontex staff deployed in third countries; • Number of illegal border crossings; • Number of SAR incidents; • Vulnerability of border area.
Interviews with case workers and migrants at border crossing points and arrival centers (MS, Frontex)	• Identified migratory routes/patterns; • Migration intention; • Drivers of migration.
Meteorological weather data (several providers)	• Extreme weather conditions (e.g. drought, flooding); • Identified migratory routes/patterns; • Number of smuggler boats and routes on the Mediterranean.
Online geographical maps (several possible providers)	• Geographic (border) distance between countries or origin, transit and destination and estimations of duration of time required to go from one point to the next via various means of transport (on foot, by motor vehicle, train, planes, boat, etc.).
Online search data	• Web search terms about migration intentions.
Qualitative information by experts on cultural aspects	• Information on aspects of cultural distance between countries of origin, transit and potential destination countries, such as common language(s), colonial links, trade connection, size of diaspora, etc.
Reports and data on World Development Indicators (World Bank)	• Migration and net migration; • Annual remittances • Other World Development Indicators.
Satellite Pictures (several providers)	• Identified migratory routes/patterns; • Number of settlement close to EU Schengen

Forecasting category C – secondary movements and asylum applications	
	border.

Legend type of migration source

Statistical
Administrative
Innovative

This step of the data source assessment shows that several data sources are available for all three types as per the tables above. All data sources, which are mentioned in the three tables above, are operational from a technical perspective.

3.6.3 Excluded data sources

While relevant for the purpose of irregular migration forecasting, social media data will not be included in as data sources in the AI-Tool, primarily due to an unstable regulatory framework in terms of accessibility of social media data, as well as due to potential data protection issues. Nonetheless, it can be noted that social media data can theoretically be used by an AI-Tool, although it must also be noted that varying assessments exist as to the significance of their input and their overall forecasting potential compared to other sources. For example, the usefulness of incorporating data from single platforms might vary depending on different levels of usage among different countries and different nationalities. As to opportunities offered by social media platforms, Facebook and Twitter are known as common data sources for analysing public sentiments, movements or general public perceptions. Through natural language processing, algorithms identify previously coded sentiments in words and thus, visualises the public opinion and tendency towards socio-political themes. However, social media data is not only limited to sentiment analysis but can be expanded to real-time irregular migration assessments. The Facebook Data Science Team, for example, offers the possibility of mapping internal and international irregular migrations alongside each other, which cannot be done easily through traditional surveys. The project focused on so-called 'coordinated migration', defined as cases where a significant share of a population migrates as a group to a different city. To study these between-city coordinated irregular migrations, the team examines aggregated, anonymised data of all users who list both cities (departure and destination) on their Facebook profile.

3.7 Forecasting irregular border crossings: assessment and ranking of potentially relevant data sources

As outlined in chapter 3.6, the next step after establishing which data sources are relevant for which forecasting category, is to focus on those forecasting categories and relevant sources therein which are required for the forecasting of irregular border crossings in particular. As such, forecasting categories A and B are addressed in this sub-chapter, considering that forecasting category A informs forecasting category B, which ultimately covers the actual forecasts of irregular border crossings.

For both forecasting categories, the identified relevant data sources were sorted into high, medium, and low relevance, and each data source was described as to the information it offers, with additional cautioning on potential limitations. A full overview of characteristics of each data source can be found in Annex C. This is an intermediary step towards the case study in chapter 4, which narrows down even further and focuses on irregular border crossings at the Greek sea and land borders in particular.

3.7.1 Methodology

Practically, for each of the two forecasting categories A and B, the respective lists of relevant data sources required for each (as presented in the tables in chapter 3.6 above) were sorted horizontally into high, medium, and low relevance for the respective forecasting category. The relevance of data sources is only assessed theoretically in an ex-ante setting and was not tested by an AI-Tool. Relevance is measured by the potential degree of usefulness and contribution to the desired output by a forecasting tool:

- *High relevance:* The data sources show a great potential of usefulness and contribution to the forecasting tool covering a wide range of indicators that can be translated into the AI algorithm. The datasets are easily accessible and can be integrated into the internal assessment, or are so relevant that despite possible issues, they are deemed important enough to be assigned 'high relevance';
- *Medium relevance:* The data sources present a medium potential of usefulness and contribution to the forecasting tool. The indicators should be used complementarily, as they inform underlying trends and can thus contribute by an early notice of shifts of trends;
- *Low relevance:* The data sources show a low potential of usefulness and contribution to the forecasting tool. The data sets are not strictly necessary for the analysis and predictability of the respective crisis and irregular migration movements. However, the presented data sources in this category can be used complementarily for the assessment, and further serve to provide broad background information, also informing trends to some extent.

It must be noted, however, that data sources which might be of high importance in relation to other forecasts, such as secondary movements and asylum applications, might to some extent also influence (routes towards)

irregular border crossings into the EU. Such sources may therefore still be listed among the low or medium relevant data sources for the forecasting of border crossings, which in such cases should not be seen as an overall low value of respective data sources, but only in relation to the particular purpose of the output.

Vertically, a traffic light system was added, to indicate the level of risk and/or difficulty assigned to sources, e.g. in terms of legal or financial access restrictions. Low risk is indicated by green colour, medium risk by orange and high risk is coloured red. This was also influenced by the qualitative descriptions of possible limitations and risks related to the use of respective data sources. For example, these include the risk that variables that do not adequately capture the phenomenon that should be forecasted in a particular scenario, or the data sources are not suitable for the assessment since they lack legal feasibility and/or timeliness, or are missing indicators on generalisability, completeness or coherence.

3.7.2 Forecasting category A

Forecasting category A is covering potential critical situations both within the EU and in third countries, which could have an impact on irregular migration as potential drivers. In order to assess the potential crisis situations both within the EU and third countries, different types of data sources are taken into account. The data sources entail qualitative as well as quantitative datasets from different entities, including public and international, as well as non-profit organisations. The different entities produce a large number of different data sets based on past conflicts and thus generate a risk potential for future crisis situations. The forecasts should not only be based on statistical data from the past (historical data), but also on predictions that are derived from qualitative analyses of reports and interviews. The ultimate results of forecasting category A are outputs on drivers of irregular migration.

Table 3.6 Data sources for forecasting category A

Data sources and relevance for forecasting category A		
TL	**Data source**	**Reasoning**
Potential high relevance		
	World Bank reports and data on World Development Indicators	• The World Bank Reports indicate and assess irregular migration trends; • The World Bank further publishes useful data on hundreds of indicators (World Development Indicators). Such information, e.g. on changes to underlying economic situations in countries of origin, population growth, annual remittances from diasporas to countries of origin, etc. might assist in profiling countries on structural indicators, or as

Data sources and relevance for forecasting category A		
		part of past data to be used for the training of models; • Issue of timeliness of both reports and World Development Indicators. Further, access to metadata requires login.
	OECD International Migration Database	• The OECD's international migration database shows inflows of foreign population by nationality, the outflow of foreign population by nationality, the inflow of asylum seekers by nationality, the presence and size of diasporas in other countries, and more. It is publicly accessible to stakeholders in several data formats.
	ACLED	• Since ACLED collects real-time data on locations, dates, actors, fatalities, and types of all reported political violence and protests, the data set gives an accurate indication of a potential crisis situation; • Furthermore, the integrated Conflict Pulse presents prediction on trends in conflict actor behaviour a week into the future or explores historical predictions being exactly suitable for the assessment of Type A data sources. The dataset is publicly available and for free.
	Crisis Group Reports	• The Crisis Group identifies past, current and possible conflicts and crises around the world, assessing different factors and actors involved; • The global conflict tracker Crisis Watch assesses historical data and gives predictions on future scenario enabling a better understanding and forecasting of conflicts in different regions which are highly needed for Type A data assessments. The dataset is publicly available and for free.
	GDELT	• Provides access to real time monitoring of open source news from most, if not all, relevant countries. Identifies locations, themes, events, etc. This includes reports on legal and policy changes, but also on detected migration movements and other local occurrences. The output is available in multiple data formats and should consequently be easily usable for the AI-Tool; • Timeliness-reliability trade-off: Possible issues relating to the reliability of data, i.e. how well events are actually described. Further, the accuracy of output may be questioned in terms of neutrality and distortion through over- or underreporting, reporters' opinions, fake news,

Data sources and relevance for forecasting category A		
		and possibly controlled media outlets. These issues can be addressed through some data processing.
	Uppsala Conflict Data Program	• The world's main provider of data on organized violence and the oldest ongoing data collection project for civil war. UCDP researchers regularly publish research on organized violence, its causes, escalation, spread, prevention and resolution, in top scientific journals and books. UCDP also operates and continuously updates its online database (UCDP Conflict Encyclopaedia) on armed conflicts and organised violence, in which information on several aspects of armed conflict such as conflict dynamics and conflict resolution is available; • This interactive database offers a web-based system for visualising, handling and downloading data, including ready-made datasets on organized violence and peace-making, all free of charge
	JRC Global Conflict Risk Index	• JRC Risk Index is the quantitative starting point of the EU's conflict Early Warning System. It is an index of the statistical risk of violent conflict in the next 1-4 years for each country in the world. It assumes that structural conditions in a country are linked to the occurrence of violent conflict; • The GCRI is exclusively based on quantitative indicators from open sources. Using linear regression model, including historical data to train the model in order to make accountable predictions. The data is only available with restricted access which raises transparency issues.
	UNHCR	• The UNHCR data provides information on migration and net migration, it depicts the current motion of irregular migration shocks coming from Africa or the Middle East, and it provides data on global taxonomies, as well as resettlement and population statistics; • The data does not predict the irregular migration routes per se; however, the actual depiction of current irregular migration flows show an accurate picture and enables a forecast for the immediate future. Since the datasets take into account different data sources, the risks of falsely interpreting the prediction are very low.
Potential medium relevance		
	Country reports by	• Country reports by embassies can provide qualitative information on local occurrences,

Data sources and relevance for forecasting category A		
	embassies (MS) and EU delegations (e.g. migration sections' reports)	including irregular border crossings in third countries along routes and third countries along the EU external border. Information could include populations' sentiments, relevant countries' policies, and other elements relating to irregular migrants' presence and possibly also intentions in a given country; • Possible issues of access (unless collected by EU delegations), timeliness, and format.
	Europe Media Monitor	• Free of charge analysis of news in 70 languages across many countries, thereby able to cover local and smaller events which might not be covered by large new firms or which might otherwise take longer to be reported via official channels. It can provide information on third neighbouring countries' official and unofficial policies and changes therein; • Question of the accuracy of output – neutrality and distortion due to over- or underreporting based on journalists' opinions, fake news, and possibly controlled media outlets.
	Integrated Conflict Early Warning System (ICEWS)	• ICEWS combines a database of political events and a system using these to provide conflict early warnings. Event data consists of coded interactions between socio-political actors (i.e., cooperative or hostile actions between individuals, groups, sectors and nation-states); • The dataset makes predictions on conflict and their process indicating a possible scenario. There are weekly updates, and the data is publicly available and for free.
Potential low relevance		
	Satellite Pictures	• Satellite pictures can be used to identify mobility patterns, incl. activities on routes and concentrations on certain areas in border regions; • Possibly limited access and accountability mechanisms might be required.
	Global Terrorism Database	• The Global Terrorism Database presents data on terrorist attacks until 2014, indicating location, time and motifs behind terror attacks in different countries; • However, the database does not indicate recent conflict dynamics or crisis that might lead to the emergence of terror attacks. The lack of this information as well as the limited timeline of the dataset (only until 2014) exacerbate clear

Data sources and relevance for forecasting category A		
		predictions and would not be suitable for a forecast.
	Gallup World Poll Survey	• This source offers information on migration intention in countries of origin. As such, it is of low relevance for prediction actual border crossings but serves as a good indicator in an earlier step, whereby it helps to place focus on countries in which intention to migrate may be growing; • Possible limitations in relation to terms of services (costs).

3.7.3 Forecasting category B

Forecasts on irregular border crossings require data on several key phases from the time-spectrum of irregular migration movements and beyond. These relate to the following:

- Underlying trends;
- Early recognition of drivers and shocks in countries of origin;
- Information on the level of flows along irregular migration routes;
- Situation in the EU external border region.

Each of these phases and points of information ultimately build on each other to inform reliable forecasts of irregular border crossings into the EU. As such, data on underlying trends are continuously informed by more up to date data sources on the three main phases. Here, data on drivers allow for the forecasting of the onset of irregular migration movements, while information on routes enriches insights on the direction and volume of movements. Data on the situations in the EU external border region influences short-term early warnings on attempted irregular border crossings. Each of the aforementioned data clusters will be addressed in turn below. Following this, the respective data sources informing each of these categories are clustered into high, medium, and low relevance for the immediate purpose of forecasting irregular border crossings into the EU in the table below.

Underlying trends

Underlying trends of irregular migration movements are incorporated by forecasting category B to provide context to any recognition of already established drivers or routes. For example, on the basis of historical information on socio-economic aspects in a given country of origin, it can be derived on which countries particular emphasis needs to be placed. This relates to structural factors in a given country and the development of each of the latter, particularly in terms of economic development, demographic change, and others. Here, the on-set of irregular migration movements can be traced on a long-term trend scale, whereby higher average income levels indicate a likely onset of irregular migration movements, while low and high

average income levels typically indicate low irregular migration movements.[35] Further, information on traditional routes can be derived from trend data, whereby it can be established that irregular migrants from a certain country tend to take certain routes towards certain destination countries. Such information can be compiled on the basis of statistics on previous irregular migration movements, figures on asylum applications, the size and geographic distribution of diasporas within the EU, and more.

On this basis, if a shock or drivers are forecast or recognised for a particular region or country of origin, this immediately allows conclusions on the likelihood of the potential volume of irregular migrants and the route they might be taking, including potential entry points at the EU external border. These trends can then be enriched by the AI-Tool with more recent data on the aforementioned drivers in country of origin, and other indicators from along irregular migration routes, i.e. transit, to confirm or shift predictions on the actual irregular migration movements which may have been kicked off. Likewise, data from the EU border region is incorporated into this forecast and correction of (previous) underlying trends, considering that policy changes and shifts in responses by countries result in the shifting of trends and might thus establish new ones, which needs to be factored in by the AI-Tool. Likewise, in a related step, information on visas and exit data can assist in establishing trends on overstayers, who may have crossed the border into the EU regularly, but have stayed beyond the permitted period of time and thus ultimately also stay the EU irregularly.

Early recognition of drivers and shocks in countries of origin

This sub-category relates to data sources providing information on the general background and events in countries of origin. The incorporation of data on, among others, underlying socio-economic and political factors, and the occurrence of sudden critical situations and migration shocks is important, as these might set off irregular migration movements in the first place. Recognising and/or forecasting this at the earliest possible point in time is of value, as this assists in facilitating an early view on potential changes to pressure to existing and/or new irregular migration routes and thus ultimately also on forecasting irregular border crossings. As such, this category relates to incorporating data input on drivers of irregular migration from a long-term perspective. Such data can be derived from the outputs of forecasting category A, as outlined further above.

Information on the level of flows along irregular migration routes

Next, data on movements recorded on known routes of irregular migration serves as an important indicator for upcoming pressure on EU external border regions. Such data can be derived from country reports provided by embassies in origin and transit countries, and from Europol and Frontex reports which inform on situations along known irregular migration routes, based on information submitted by third countries on numbers of irregular migrants apprehended along the way.

While the underlying trends have established the best-known routes and to which border crossing points into the EU they lead, the combination of

[35] JRC, *International Migration Drivers: A quantitative assessment of the structural factors shaping migration* (Luxembourg: Publications Office of the European Union, 2018).

identified onsets of irregular migration from certain countries with information on actual levels of flows along established and emerging routes, offers more up to date and reliable insights on the routes that are effectively likely to be taken and where they lead to. With this combination, the first crucial step of forecasting category B is covered. It must be known where irregular migrants come from, which routes they take, and to what degree these routes are utilised, to forecast where varying levels of attempts of irregular border crossings can be expected, and at which point in time.

Situation in the EU external border region

As a final step, the aforementioned information must be combined with another crucial factor, namely the situation in the EU external border region *per se.* At this stage, it is already established if and on which route how many individuals may be travelling towards the EU's external border. Consequently, the next step covers data allowing for the forecasting of actual attempts of irregular border crossings, which can be influenced by various factors. For one, this includes qualitative data on neighbouring countries' and the EU's policies, such as strict or less strict border enforcement, or the presence of SAR missions. It also includes data on activities of facilitators of irregular migration, such as the production and sale of forged or fraudulently obtained documents, the purchase of boats, the hiring of rental cars, and more. It also includes information on numbers of irregular migrants already present in third neighbouring countries of the EU, as these give indicators on the pressure for irregular migrants present in such countries to cross the EU external border and where. This further includes the incorporation of data sources allowing for the forecasting of shifts in the short term, such as news data on relevant local events, which can flag short term changes in the modi operandi of irregular border crossing attempts, but also meteorological data, which could influence numbers of attempts to cross the Mediterranean Sea for example. This, in addition to historical data, assists in forecasting where and by what means irregular border crossings will be attempted.

Data sources for each of the categories

As such, a vast range of data sources is required by forecasting category B to inform all relevant aforementioned aspects of forecasting irregular border crossings. For each of the above-introduced categories, several data sources can be identified. In the sections below, they are clustered in accordance with their relevance for forecasting border crossings specifically, which means that data informing on movement on routes and on occurrences in the border region are of higher immediate relevance than data informing underlying trends. The data sources encompass a combination of quantitative and qualitative data, and draw from information on drivers, flows, and other short-term indicators.

Table 3.7 Data sources for forecasting category B

Data sources and relevance for forecasting category B		
TL	**Data source**	**Reasoning**
Potential high relevance		

Data sources and relevance for forecasting category B		
	Outputs from A	• Influx of irregular migrants can be greatly influenced by (potential) critical situations within the EU and in third countries; e.g. war, open borders, etc.
	EU information on Union migration policies	• Information on legal and policy changes by the EU can be provided directly from its core, i.e. EU administrators or assistants.
	EASO Early Warning Reports	• EASO's Early warning and Preparedness System (EPS) is a data collection system gathering information under indicators focussing on all key stages of the Common European Asylum System (CEAS); • Countries provide monthly data to EASO within 15 days, with all 30 EU+ countries (Member States plus Norway and Switzerland) contributing; • It forecasts the number of asylum applications to plan resources and actions on asylum management (preparedness); • Questions of access and reliability, as data provided by Member States may be delayed and requiring clearance.
	Meteorological data	• Weather data assist in short term prediction on how likely it is, that individuals attempt certain border crossings. In cases of too much rain or storm, crossing seas or rivers may be considered too dangerous, and the pressure on border crossing points might lessen; • Depending on the provider chosen, possible issues of access or associated costs could arise.
	Frontex reports	• Frontex provides a variety of information on vulnerability assessment of border area(s), RAU forecast of irregular border crossings, data on movements between non-Schengen and Schengen areas, situational pictures, general risk analysis; • Questions of timeliness (some reports are only published on a quarterly or annual basis).
	Frontex illegal border crossing data	• Data on already detected illegal border crossings is too late to be incorporated into predictions of the crossings; however, such data can indicate trends, e.g. shifting pressure on particular entry points; • Possible issues of timeliness, access and format.

Data sources and relevance for forecasting category B		
	Country reports by embassies (MS) and EU delegations (e.g. migration sections' reports)	• Country reports by embassies can provide qualitative information on local occurrences, including irregular border crossings in third countries along routes and third countries along the EU external border. Information could include populations' sentiments, relevant countries' policies, and other elements relating to irregular migrants' presence and possibly also intentions in a given country; • Possible issues of access (unless collected by EU delegations), timeliness, and format.
	Europol reports	• Europol's EMSC provides weekly EPMT reports, contributes to ISAA reports and provides intelligence notifications. It draws from SIENA data in their reports, which includes information from EU law enforcement, but also from external parties. It also holds regular conference calls with law enforcement officers from countries along irregular migration routes. It could thus provide information on apprehensions at external borders, developments along routes, and modi operandi; • Possible issues of timeliness and format.
	European Asylum Dactyloscopy database (Eurodac)	• Data on asylum applications assists in indicating trends; • Data sources entail personal data and would consequently require an additional step to ensure anonymisation before data is fed into the AI-Tool.
	Member States' information on migration policies	• Qualitative information on migration policies of different Member States give more outlook on the handling and managing the intake of irregular migrants and asylum seekers in the respective Member States; • However, possible issues regarding how such data should be quantified. However, a number of indicators on asylum policies are collected by EASO under the IDS system. Here, recognition rates can be used as a proxy.
	Satellite Pictures	• Satellite pictures can be used to identify mobility patterns, incl. activities on routes and concentrations on certain areas in border regions; • Possibly limited access and accountability mechanisms might be required.
	Online search	• Google Trends/Analytics data can be used to

Data sources and relevance for forecasting category B		
	data	predict turning points in migration trends and offer real-time insight into intentions and points of interest; • When the barriers for migration are not too high, the Google Trends interest (GTI) is accurate in forecasting irregular migration, however, it does not capture populations with low technology access; • For online search data, changes of certain searches in a given country are recorded rather than levels, which is impossible to be connected to individuals.
	Schengen Information System (SIS II)	• The Schengen Information System captures all relevant information on immigration and internal information on border management among Member States; • Changes in the number of irregular border crossings from Non-Schengen to Schengen countries can be used as an indicator for asylum applications in EU external border countries and where some of these can be expected (not all irregular migrants apply for asylum in the EU arrival country); • Changes in numbers of irregular border crossings further indicate whether and from where secondary movements can be expected; • Significant issues of access (interoperability) and a separate anonymisation step is required to ensure personal data stored in SIS II is not used.
	Dublin Data from MS	• The Dublin Data from Member States captures all relevant information on regular migration and asylum seeker applications in the respective Member States; • Although the data does not entail information on irregular migration, it can assist in forecasting irregular migration by taking into account the frequency of applications, the applicants and indicated migration routes by the asylum seekers, which is thus also useful to predict secondary movements.
	Interviews with case workers and irregular migrants	• Interviews with caseworkers and irregular migrants at border crossing points and arrival centres can give an outlook into the situation in countries of origin of the different asylum seekers, including on drivers for irregular migration and chosen irregular migration routes. This can be taken into account as qualitative data

Data sources and relevance for forecasting category B		
		for predicting reoccurring irregular migration shocks; • Further, Frontex conducts debriefing interviews, and EASO launched a project on surveys of asylum-related migrants. Such data could similarly provide valuable information; • However, the caseworkers' perception of the applications and the routes may vary and are dependent on subjective indicators. This bears a risk of diminishing validity and coherence among the Member States' caseworkers. Further, issues of access, timeliness and format can arise.
	Visa Information System (VIS)	• The VIS registers data on Schengen visa applications and decisions, i.e. issued and refused visas. It further collects personal information which can be used for determining future risks, but also profiling; • A separate anonymisation step would be required to ensure personal data is not accessed. Further, issues of access exist (interoperability).
	Mixed Migration Foresight Project	• The Mixed Migration Foresight Project utilises fact-based datasets from multiple sources, combined with first-hand observations and expert judgments that generates a prediction model for regular and irregular migration shows; • This tool delivers a similar approach that is anticipated by the contracting authority; • However, the data is not publicly available, raising transparency issues for a wider public.
	GDELT	• Provides access to real time monitoring of open source news from most, if not all, relevant countries. Identifies locations, themes, events, etc. This includes reports on legal and policy changes, but also on detected irregular migration movements and other local occurrences. The output is available in multiple data formats and should consequently be easily usable for the AI-Tool; • Timeliness-reliability trade-off: Possible issues relating to the reliability of data, i.e. how well events are actually described. Further, the accuracy of output may be questioned in terms of neutrality and distortion through over- or underreporting, reporters' opinions, fake news, and possibly controlled media outlets. These issues can be addressed through some data processing.

Data sources and relevance for forecasting category B		
	IOM Displacement Tracking Matrix	• Displacement Tracking Matrix (DTM) gathers and analyses data to disseminate critical multi-layered information on the mobility, vulnerabilities, and needs of displaced and mobile populations that enables decision-makers and responders to provide these populations with better context-specific assistance. Furthermore, this tool can also be used to predict irregular migration shows since DTM indicates mobility and needs of displaced and mobile people.
	Migration Data Portal	• The Migration Data Portal is a publicly accessible point to timely, comprehensive migration statistics and reliable information about migration data globally; • The interactive world map visualizes international, publicly-available and internationally comparable migration data. The different data sets and indicators are used to make predictions on irregular migration shocks that might result in the future.
Potential medium relevance		
	Reports and data on World Development Indicators (World Bank)	• The World Bank Reports indicate and assess irregular migration trends; • The World Bank further publishes useful data on hundreds of indicators (World Development Indicators). Such information might assist in profiling countries on structural indicators, or as part of past data to be used for the training of models; • Issue of timeliness of both reports and World Development Indicators. Further, access to metadata requires login.
	Eurostat	• Eurostat's datasets on migration focus, among others, on the intake of asylum seekers and irregular migrants from non-Member States in the respective Member States but does not enclose information on irregular migration shocks or border crossings; • However, the information on asylum-seeking applications can be used to make predictions about trends of migration flows in the upcoming years.
	UNHCR	• The UNHCR data provides information on migration and net migration, it depicts the current motion of irregular migration shocks coming from Africa or the Middle East, and it

Data sources and relevance for forecasting category B		
		provides data on global taxonomies, as well as resettlement and population statistics; • The data does not predict the irregular migration routes per se; however, the actual depiction of current irregular migration flows show an accurate picture and enables a forecast for the immediate future. Since the datasets take into account different data sources, the risks of falsely interpreting the prediction are very low.
	Europe Media Monitor	• Free of charge analysis of news in 70 languages across many countries, thereby able to cover local and smaller events which might not be covered by large new firms or which might otherwise take longer to be reported via official channels. It can provide information on third neighbouring countries' official and unofficial policies and changes therein; • Question of the accuracy of output – neutrality and distortion due to over- or underreporting based on journalists' opinions, fake news, and possibly controlled media outlets.
	Internal Displacement Monitoring System	• In the form of reports as well as raw data, IDMC provides verified, consolidated and multi-sourced estimates of the number of people internally displaced or at risk of becoming displaced by conflict, violence, disasters and development projects across the world. The estimates are recorded and presented on monthly as well as annual bases. Furthermore, complements this global data with interdisciplinary research into the drivers, patterns and impacts of internal displacement across different country situations, contexts and scenario. In addition, 'Flash' updates on new displacement events across the globe, published on a daily basis on an interactive map on IDMC's website's main homepage.
	OECD International Migration Database	• The OECD's international migration database shows inflows of foreign population by nationality, the outflow of foreign population by nationality, the inflow of asylum seekers by nationality, and more. It is publicly accessible to stakeholders in several data formats.
• **Potential low relevance**		
	MapBox	• Mapbox gives information on irregular migration routes until 2017 depicting the motifs, the route

Data sources and relevance for forecasting category B		
		and the criminal activity as well as the exploitation of asylum seekers on the way on the basis of a storyline. The interactive platform delivers information on a more micro and individual level; • However, since the datasets are restricted until 2017, it does not deliver an adequate source to make predictions for future irregular migration flows that are regarded from a meta-level.
	ICPSR data	• This source offers information on population characteristics.
	Wittgenstein Centre for Demography and Global Human Capital	• This source provides projections of net migration rates per country until 2100, which can be factored into steps for early expectations of irregular migration movements.

While this section has presented an overview of which data sources are needed to forecast irregular border crossings at EU external borders in general, the following chapter covers data sources required to forecast irregular border crossings into Greece in particular.

4 MODELLING CONSIDERATIONS AND CASE STUDY

4.1 Selection of the case study

To provide a specific and concrete illustration of a case study in terms of potential output of the AI-Tool, a case study was selected by the European Commission on the basis of seven potentially relevant use cases. The detailed descriptions of the case studies were included in the draft interim report. The selected case study covers the following forecasting scenario: '**Arrivals at a point of the external borders with the following prediction variable: The number of irregular migrants who will arrive at Greece's external borders from visa obliged countries next month without a valid visa and cross.**

4.2 Objective and approach of the case study

The objective of a targeted case study is two-fold. First, to provide a clear and concrete example of how a forecast would be produced by an AI-Tool and second to illustrate the kind of modelling and design processes that are implied by different purpose-prediction decisions. The case study functions as the consecutive and final step of this study. The findings of the case study and the elaborations of the final report will support and guide the European Commission in their decision on the next steps of the development of the AI-Tool.

Concretely, it will elaborate further on issues and variables by:

- Conducting more research on how illegal border crossings are defined and collected by Frontex and issues the methodology might raise for the purpose of the case study (and subsequent assessments);
- Providing more description and strengths and weaknesses of potential data sources that theory, experience or common sense suggests may be helpful to predict this variable;
- Developing a chart on 'Cost and complexity of data sources' and 'Accuracy of the predictions' and add relevant data sources;
- Drawing implications for forecasting a similar variable in other locations. The case study will help to show what could simply be copied or scaled from a single border crossing and what would need to change. This will allow the European Commission to have a better overview of how this case study could apply to other routes.

4.3 Decision-making purposes

The selected case study is relevant for the following **decision-making purposes**:

- Adjust resources available for irregular migrant processing in Greece;
- Adjust resources available for housing and other services to irregular migrants;
- Content and intensity of diplomatic engagement with third countries neighbouring Greece;
- Level of attention that the EU should direct towards Greece's external borders;
- Level of political attention and crisis management preparation among Greece's politicians.

4.4 Data source selection and exemplary modelling consideration

This section begins with a narrative description of data sources that are likely to help in forecasting illegal border crossings at the Greek borders. It then summarises the narrative with a chart indicating **trade-offs between cost/complexity versus theoretical benefits** of including a data source, followed by a recommended **mix of data sources** to structure and test in the **proof-of-concept phase** (see chapter 0). Finally, guidance is provided on what needs to be considered if the presented approach should be replicated for a similar prediction variable in a different location.

To be able to begin with the modelling considerations, it is important to understand that Frontex defines illegal border crossing as detections of illegal border crossing at the EU external land, air and sea borders. Member States collect this data and report it to Frontex on a daily and weekly basis. Frontex then aggregates the data from all Member States and reports it. Detections of illegal border crossing do not represent the actual number of persons illegally crossing borders but only entails illegal border crossings that are detected by Member States' border management operations. There is almost certainly a gap between detections and actual crossings, especially at land and sea borders.[36] Further, there is currently no EU system in place which is capable of tracing each person's movements following an illegal border-crossing and establishing the precise number of persons who have illegally crossed the external border.[37]

In addition, a person may attempt to cross the border irregularly several times, either at the same or in different locations at the external border of the EU. This is of particular significance for Greece, where large numbers of irregular migrants are entering the EU via Greek islands and then move through the Western Balkans and re-enter the EU through Hungary or Croatia on the way to their desired final destinations in the EU. This means that a large number of irregular migrants who arrived in Greece and were detected

[36] Frontex, Situation External Borders (Website, 2020), accessible at: https://frontex.europa.eu/faq/situation-at-external-border/; and key informant interview 14.

[37] European Commission, Detections of illegal border-crossings; monthly statistics (Website, 2019), accessible at: https://ec.europa.eu/knowledge4policy/dataset/ds00032_en.

by the Greek authorities are probably detected and reported again once they arrive at the borders of Hungary or Croatia. Furthermore, Member States often are not able to report the nationality of the person illegally crossing, for example, when they do not travel with documentation.[38]

The above discussion of the prediction variable definition highlights a first choice in the selection of data sources: should the migration model include variables that may directly influence **detection**, as opposed to the reality of crossings? For example, those operating the migration model could attempt to quantify detection capacity. An approach to do this could involve a search for **correlations** between reported detections and variables that may theoretically be associated with reported detections, such as the operating budget of Greek border authorities. Other examples could include the **number and size of physical assets** associated with border management in Greece, or changes in enforcement-related legislation in Greece.

The **advantage** of including such variables would be that the model is more likely to capture these theoretically important influences on the prediction variable. The **disadvantage** is further cost and complexity in developing and maintaining the data sources. Reliance on Greek government data also implies that operators of the AI-Tool would need to understand how that data is generated and to be aware of changes to its structure - for example, if the Greek government changes how it reports a budget or restructures its agency responsibilities, operators of the AI-Tool would need to revisit the data source.

This example essentially summarises all of the important principles of **data source selection** in the context of the General Assessment, which are summarised below:

- There are few tools equivalent to the AI-Tool that the European Commission wants to build, so at this stage, the best guidance comes from migration theories and expert knowledge about influences on the prediction variable;
- One can estimate the cost and complexity of adding a data source;
- However, one cannot quantify in advance the improved accuracy that can be expected from adding a new data source but only provide intuitive estimates. Data sources that are already being collected for other purposes appear cheaper now but are less likely to be optimised for the forecasting purpose and are more vulnerable to uncertainty and change in future;
- From a legal perspective, the use of below-mentioned data sources does not raise concerns due to the fact that it contains, mainly statistical data. The data regarding Europe's policies on irregular migration and asylum might raise some limitations depending on the sets of data deployed. Limitations might potentially arise, as already mentioned, regarding the terms of service for all the above-mentioned sources. Living in the era of widespread disinformation and fake news might, however, influence the

[38] Key informant interview 14.

accuracy of the AI-Tool outputs. Personal bias might as well influence the forecasts. In accordance with GDELT's terms of use, there are no restrictions to use the available data for 'academic, commercial or governmental' except, of course, proper referencing.

Historical detections

It is highly likely that the pattern of previous detections will be helpful for forecastings of the same variable in the future as trends in nationalities and the number of detections might arise over time. However, additional variables are needed to incorporate other drivers of change in the trends for a future tool. It is commonly accepted that 2015 marked a change in arrivals to Greece. Frontex reports show that in 2015, detections of illegal border crossings in the Eastern Mediterranean route went from 50,834 to 885,386. [39] Trends remained high in 2016 before dropping in 2017 and increasing again in 2018. It may therefore be most relevant to use data from 2015.[40]

The use of historical detection data highlights a key principle in data source design that may be obvious in theory but can be difficult in practice. In this context, it is important to recognise that the AI-Tool should only be trained on data with the form and timing, which will also be available in the future. For example, let us imagine that operators of the AI-Tool discover that detections at week 49 appear important in forecasting detections in week 50. That is useful in theory, but its benefit in practice depends on how long it takes to get the latest detection data. If it takes three weeks to get the latest data, then this source will not be available at the time you need to generate the next forecast. Instead, you would have to train the data source on detections in week 46.

Europe's policies on migration and asylum

Within **short timeframes**, the migration literature suggests that changes in policies and communications about those policies can influence irregular migrants' decisions.[41] It may be fruitful to structure data on policy changes and announcements in order to capture this element of the European Union's own influence on the prediction variable (using MIPEX, for instance). If a major database like **GDELT** (see Annex C for a more detailed assessment) is already going to be accessed for the AI-Tool (see below), then a relatively cost-effective approach, to begin with, would be to incorporate irregular **migration-related topics** and **sentiment analysis** in relation to the EU's policy announcements. If this approach suggests any promise, then the European Commission could consider creating a **dedicated data source** based on the European Commission's capacity for monitoring European policies and communications.

---

[39] Frontex, *Migratory Routes* (Website, 2018), accessible at: https://frontex.europa.eu/along-eu-borders/migratory-routes/eastern-mediterranean-route/.

[40] Frontex, Migratory Routes (Website, 2018).

[41] European Commission, *How West African migrants engage with migration information en-route to Europe* (Luxembourg: Publications Office of the European Union, 2018).

Furthermore, over time, the combination of the EU's asylum and migration policies may have a decisive impact on irregular arrivals. For example, if the EU deported failed asylum-seekers quickly, it is likely that fewer people who fail asylum would attempt to enter irregularly. Over a longer time, lag, this policy and enforcement balance is likely to be highly influential on the prediction variable. One method of creating this data source could be to structure a panel of observers who quantify different dimensions of Europe's policies and practical enforcement of those policies. A more direct method could be to measure potential irregular migrants' own perceptions of Europe's policies and the implications for the irregular migrant (see below on migrant surveys).

Demographics and detections

Migration theory repeatedly highlights that cultural and network effects influence patterns of irregular migration. It is reasonable to assume that different nationalities or ethnicities behave in different ways in terms of risk-taking, opportunity-seeking and distributing information. Furthermore, Frontex data already provides a breakdown of countries of origin of detections, so it would be cost-efficient to include this breakdown as an input into the migration model. The same calculation would be relevant to any other demographic data that is already available in the same Frontex reporting.

Situation in third countries

The decision to migrate is more closely related to events in countries of origin, so those events are key predictors, although providing a longer lag. This can, however, form part of the model.

Second, monitoring events in third countries also means being able to monitor events in key transit countries like Turkey in our case study. Therefore, data in third countries (countries or origin, transit countries, neighbouring countries, etc. will likely be relevant predictors. Data from origin countries would provide a longer lag; whereas data from transit countries may contribute to the forecasting with a shorter lag.

For example, information coming from the home country may give people more or less optimism regarding their options. People back home are also reading the news about transit and destination countries and interpreting this in conversations with people in transit countries. The situation in the country of origin may also affect **new** flows of people from home to Turkey, after a longer delay. Similarly, information on routes or destinations received by the diaspora, people back home or provided online can influence the rest of the journey for people on the move.

There are relatively cost-efficient data sources such as **GDELT** that would support structuring of sentiment, political, economic and social trends in countries of origin and transit countries. The limitation of **GDELT** arises from

its relatively weak filters on the quality of news.[42] The more that the operators of the AI-Tool invest in tinkering with the structure of the GDELT data series, the more likely they are to narrow down better inputs, but the more costly and complex the setup and monitoring will be.

Another option would be to incorporate the results of specific **surveys** among the most relevant sub-national populations in the top countries of origin ad in transit countries. For example, the survey results would gauge demand and intentions related to irregular migration or gauge intentions, destinations and routes related to the rest of their journey in transit countries. This could be done by way of GALLUP surveys, for whom the insertion of a specific question for specific sub-populations can be financed. Hence, this would require some efforts and also entail costs; however, it can contribute to boosting accuracy.

The interaction of origin and transit information indicates three other **key principles** in data source design for the modelling:

- **Different time lags** may be appropriate to different data sources. For example, you may find that the most influential data from Afghanistan is 90 days ago, the most relevant from Bangladesh 120 days ago, and the most relevant from Turkey 20 days ago. Discovering this will be an iterative process, i.e. you cannot guess the exact lag in the data source in advance;
- Including **multiple time lags** may capture important changes over time that are practically impossible to foresee. For example, there may be an influential relationship between Afghanistan 90 days, Afghanistan 60 days, Turkey 30 days and Turkey 20 days;
- Since it is hard to foresee interactions between multiple time lags but relatively cheap to test them, it is advisable to do so.

Meteorological conditions

Expert and anecdotal knowledge, evidence and empirical trends in irregular arrivals to the EU borders through the Eastern Mediterranean route suggest that the weather influences flows and/or detections. For example, irregular movements in the summer have been more numerous than in the winter.[43] Getting basic meteorological data is also relatively cost-efficient, so it would certainly be worth including during a testing phase. This could include: sea conditions in Turkey's national waters; Greece's national waters; and weather forecasts in Turkey and Greece. Note that the issue of timing matters here again: for example, the model could incorporate the weather in week 46 as independent variables on forecasts in week 50, and/or it could incorporate a weather **forecast** in week 46 as independent variables.

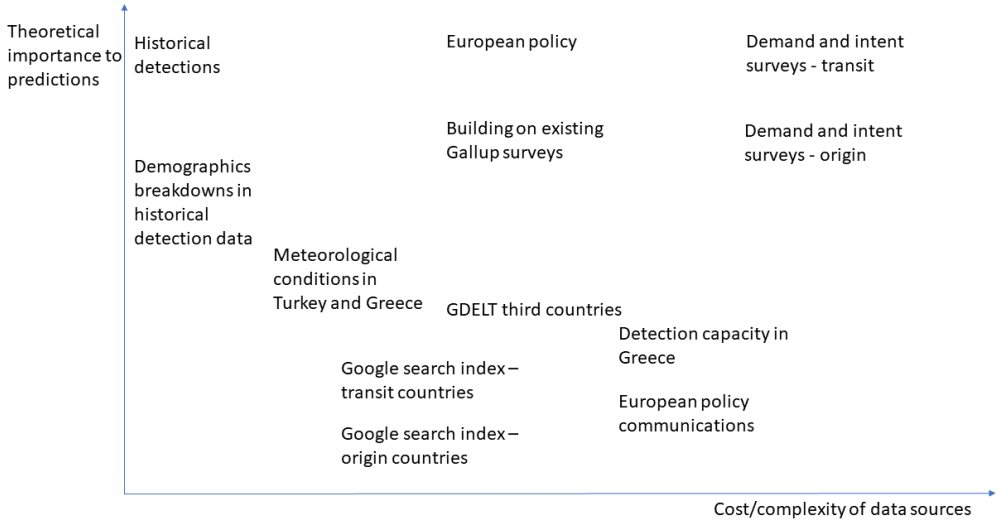
[42] Source: Insights from desk review as well as key informant interviews.

[43] Kuschminder et al, 2015; Katsiaficas, *Asylum Seeker and Migrant Flows in the Mediterranean Adapt Rapidly to Changing Conditions* (MPI, 2016).

4.4.1 Visualising trade-offs

The diagram below gives a sense of the **trade-offs** between cost and complexity on the one hand, versus the expected boost to the accuracy of different data sources. It is worth remembering that these are tentative estimates, especially in relation to forecasting accuracy. The importance of each data source can only be identified through testing, i.e. incorporating it in a **proof-of-concept model.**

Figure 4.1 Figure Cost/complexity of data sources



4.4.2 Recommended approach

Building on the narrative above and considering the further detail provided in the other Assessments, the following approach is recommended for developing a proof of concept migration model:

- Incorporate **historical detections**, including as variables all available breakdowns under the top-line data;
- Incorporate all **meteorological data** mentioned above;
- Using **GDELT**, prioritise countries of origin and transit countries, starting with the most directly relevant themes and sentiments, such as political conflict and measures of negative emotions;
- Commission highly targeted **transit country surveys** with direct and indirect measures of demand and intent. Incorporate perceptions of European policies and enforcement. If initial testing suggests these appear to be helpful, then consider a similar approach to origin country surveys;
- Experiment with structuring data on short-term European policies and announcements;
- Experiment with structuring data on **Greek detection capacity with Frontex**;
- With all of the above, test model performance with many time lags.

4.4.3 Implications for forecasting a similar variable in other locations

The prediction variable defined above is localised to Greece. If the European Commission wishes to extend or replicate the modelling approach elsewhere, the following guidelines are recommended in chronological order:

- Some data sources could be **re-used**. For example, some variables from migrant surveys in origin countries or transit countries could be plugged into a new model for illegal border crossings at Croatian borders, since the countries relevant to Greece tend to be relevant to Croatia;
- Some data sources would be **similar** but would require relocation. For example, it is likely that meteorological sources relevant to Greece would also provide data relevant to Croatia, but the new locations would need to be specified;
- The more that the data sources can be customised to the new location, the more likely that forecasting accuracy can improve. For example, let us imagine that structured data on European policies and announcements were broken down into the most relevant country. The same set of data could be plugged into a new model, and it would likely place different weights on Croatia-relevant announcements;
- Some data sources that could be highly influential would be unique to Croatia. For example, migrant demand and intent in Bosnia and Herzegovina is probably not relevant to illegal border crossings in Greece, but may be highly relevant to illegal border crossings in Croatia.

As detailed in the **Operational Assessment**, it is useful to distinguish between the work required to set up data sources and the work required to create/test models. For all practical purposes, a platform that can take structured data, train models, and test them would be able to incorporate any set of structured data sources. The comparison of Greece to Croatia highlights that most of the work in shifting from one location to another is in the data source configuration.

5 OPERATIONAL ASSESSMENT

5.1 Objective and summary of the operational assessment

Objective

The objective of the operational assessment is to provide the high-level architecture of three different scenarios for the implementation of the AI-Tool for irregular migration forecasting; in addition, it assesses the capabilities and capacities (personnel and financial), including availability, for implementing, running, managing and maintaining the proposed AI-Tool.

Summary

The present assessment focuses on the operationalisation of an AI-Tool to forecast irregular migration flows. This refers to the **technical aspects** that need to be considered to implement such an AI-Tool.

This assessment develops a **modular architecture** that can be adapted to the required needs of the forecasting tool. The Operational assessment introduces the fundamentals of AI and provides a clear architecture for its operationalisation with the aim of describing three potential scenarios to operate the AI-Tool.

This AI architecture will thus set the basis for the implementation of an AI-Tool that can utilise any kind of data sources without affecting the operation of the scenarios for implementation, which vary only in terms of their performance and reliability.

Finally, it is also important to understand that the proposed architecture and the derived AI-Tool can carry out as many models as desired to be run work in parallel to forecast different outputs.

Interlinkages with other tasks

From a technical perspective, the AI-Tool for irregular migration forecasting will rely on AI techniques and will work with any input data, provided that the underlying modelling of the forecasted phenomena is properly carried out (i.e. socio-economic modelling of irregular migration phenomena).

As such, the operational assessment within the present study focuses on the definition of three scenarios for the implementation of the AI-Tool and relies completely on the general, legislative and organisational assessments with regard to the specific context of data sources, their availability, their usability, and their capacity to explain the irregular migration phenomena to be forecasted. This operational assessment, therefore, sets the basis for the AI architecture, which needs to be considered by the actors discussed under the organisational assessment. The very same single AI-platform can therefore produce different outputs provided a team of data scientists, and migration

experts are able to model the social phenomenon of irregular migration in which stakeholders are interested in forecasting.

Finally, the operational assessment is complementary to the risk assessment, which provides an analysis of the underlying risks associated with the implementation of the proposed high-level architecture, which is declined in three different scenarios.

5.2 Assumptions

The development of the Operational Assessment is based on the consortium's internal experiences and assets such as eBDA (everis Big Data Architecture),[44] which is used for public and private entities in Europe and also trending technologies in AI, and which allows identifying the core concepts. These components are described in the following chapters of this document.

The development of the architecture **scenarios** (cf. section 5.4) presented in this analysis has considered the following assumptions:

- The data sources are already assessed and categorised in the data assessment according to their nature and typology;
- The specificities of the proposed functionalities and the integration with specific large scale IT environments will be defined in a following study, namely a design and implementation study that needs to follow up on the feasibility of the AI-Tool;
- The estimation of FTEs necessary for the operation of the AI-Tool (by scenario) are defined on the basis of architectural components developed proposed for each scenario; these will need to be recalculated when the study on a design is carried out;
- The indicative cost of operation per scenario is calculated on the basis of assumptions on the FTE costs[45] (salary grids of officials and other servants of the European Union);
- The hardware and software costs are excluded from this study; these need to be elicited during a benchmark analysis to be carried out within the frame of a design study and prior to the implementation roadmap definition.

5.3 Feasibility study high-level needs

The development of a feasible and operational AI-Tool for the purpose of irregular migration forecasting consists of the preparation of an architecture

[44] everis Big Data Architecture (eBDA) is a registered Trade Mark. It is a component architecture defined by everis during several innovation initiatives that have taken place in recent years which has resulted in BigData experience projects.

[45] Based on the grade of officials and other servants of the European Union. See: Directorate-General for Human Resources and Security (European Commission), *2019 Annual update of the remuneration and pensions of the officials and other servants of the European Union and the correction coefficients applied thereto. 2019/C 420/05 2019* (Luxembourg: Publications Office of the European Union, 2019).

model to be implemented and integrated within the landscape of Justice and Home Affairs IT environments.

In consequence, the **strategic objectives** of this Operational Assessment are:

- To present the audience with a generic AI application lifecycle;
- To present the architecture model proposed for the future implementation of the AI-Tool;
- To propose three scenarios for the future operationalisation of the AI-Tool based on different sets of functionalities.

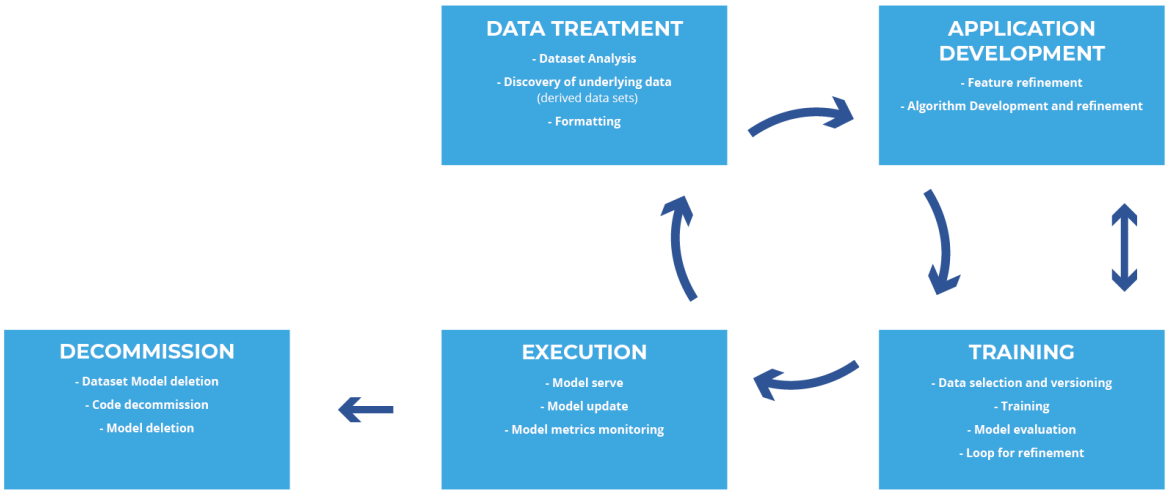
5.4 Scenarios for the operationalisation of an AI-Tool to forecast irregular migration flows

This section presents the proposed architecture for the implementation of the AI-Tool. This framework describes the generic application lifecycle of an AI-Tool and provides information on each of its phases.

The AI application lifecycle is composed of five phases, as represented in the following diagram and further described in the next sub-sections:

- During the data treatment phase, the main characteristics of the data sets to be used for the purpose of explaining and forecasting irregular migrations flows are assessed. Once a complete picture of these data is obtained, the functionalities available within the tool will allow formatting these datasets;
- During the second phase, the development of the quantitative models and their associated algorithms takes place. Once a comprehensible modelling for each migration phenomena to be reported is reached, the process will move on to the training of the algorithms;
- During the third phase, the training of the algorithms takes place. The existing data is used to teach the algorithm the optimal configuration of parameters, and then the model is run/evaluated/refined in an iterative process;
- During the fourth phase the trained model is updated and monitored;
- Finally, during the decommission phase the deprecated datasets, models, and code are disposed of.

Table 5.1 AI Application Lifecycle



5.4.1 Proposed AI architecture

The present feasibility study explores three different scenarios for the implementation of a future AI-Tool for irregular migration forecasting purposes. These scenarios are structured by a high-level architectural design which comprises the functionalities to be included at short and long term within the AI-Tool. These functionalities are described across eight main blocks as follows:

- **Block 1 - Data Management:** Within this architectural block, the tools (software solutions) that can be embarked will allow to:
  - Manage the metadata of available datasets;
  - Perform the extract/transform/load (ETL) processes to unify data according to their characteristics; and
  - Ensure the location where these processes can take place (ETL Runtime environment).

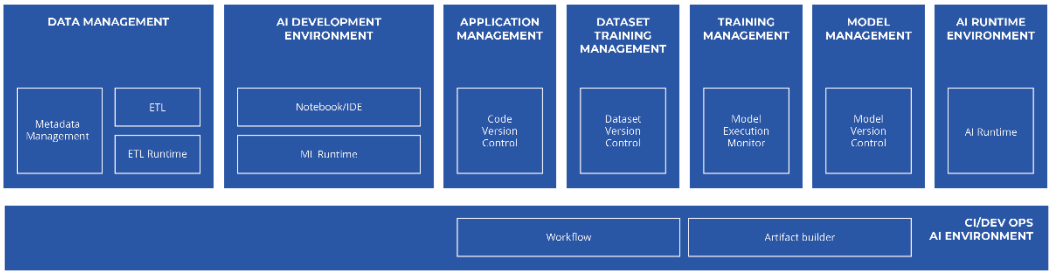
  Within this block, the Metadata Management tools will allow future developers to browse data within an existing data catalogue. In essence, the future metadata management tools will allow the staff to select the datasets they may deem pertinent for the explanation and forecasting of irregular migration flows. In addition, the *ETL and ETL Runtime* tools will allow for processing the necessary dataset in order to feed them into the model.
- **Block 2 - AI Development Environment:** Within this architectural block, the tools to be included are a Notebook/IDE (integrated development environment) and a Machine Learning (ML) Runtime environment. These two tools will be used to develop the AI models through a graphical interface, by enabling code writing and development of the model. Within this block, the staff will be able to lay down into algorithms the foreseen models for irregular migration forecasting prepared by migration experts;
- **Block 3 - Application Management:** Within this architectural block, the Code Version Control component will allow managing the different versions of the developed code and its respective updates. This block will therefore

allow to keep up with different versions of the updated forecasting models as they change through time;

- **Block 4 - Dataset training Management:** In complement of block 3, this architectural block comprises the Dataset Version Controller which will allow having versioned training datasets;
- **Block 5 - Training Management:** Within this block, the Model Execution Monitor provides information about the model training and its execution. The model's predictive performance is monitored to evaluate the need to perform a new training process. If the quality of the irregular migration forecasts previously performed deteriorates, then the monitoring functionality will indicate the need to retrain with more current irregular migration data or even to re-evaluate the entire irregular migration forecasting process;
- **Block 6 - Model Management:** Within this block, the functionality of Model Version Control will help to keep traceability of the different versioned models as previously evaluated under block 5;
- **Block 7 - AI Runtime Environment:** Within this architectural block, the execution of the models and therefore the forecasting takes place;
- **Block 8 - CI/Dev Ops AI Environment:** Within this architectural block, the Workflow and the Artefact Builder functionalities will ensure the automation of the AI application lifecycle.

The proposed architecture for the AI-Tool is visually summarised in Figure 5.2 below.

Table 5.2 AI Architecture



5.4.2 Introduction to scenarios preparation

The three scenarios presented below are incremental in terms of functionalities and requirements. In a development process, they could be seen as an incremental process that starts with the low level as a Proof of Concept (PoC), in which a simple implementation option of the AI-Tool is presented, then the medium and high-level scenarios are proposed as alternative options to proceed with either a partial solution integrated within existing large scale JHA IT systems or a complete and rather independent end-to-end solution of the AI-Tool for irregular migration forecasting.

The three scenarios mentioned above are proposed below; they are based on different levels of ambition and/or resources availability. This feasibility study

will focus on assumptions based on the following key elements in order to propose the Low, Medium and High scenarios:

- **Functional availability:** Will refer to the availability of different functionalities within the AI-Tool architecture. These functionalities allow to distinguish the scenarios across performance, reliability, and cost implications;
- **Resource availability (human and financial):** Will refer to the necessary FTEs to operate the proposed AI-Tool according to the functionalities included in each specific scenario;
- **Data availability:** Will refer to the availability of existing data infrastructure and processes that can be re-used from other IT environments, therefore differentiating each of the scenarios with respect to the data management functionalities;
- **Forecasting time horizon:** the time horizon at which the forecasting can be done is not dependant on the implemented scenario. It is a decision to be made by the end-user as part of the functional and non-functional requirements elicitation within the frame of future tool design activities;
- **Geographic scope:** This scope is independent of the chosen scenario. The availability of predictions for each area will depend on the existence of data that could help forecast irregular migration flows in each of the borders, regardless of if these data are already being collected and managed.

Based on the functionalities explained above (see section 5.4.1), the table below illustrates the differences in the architecture between the three proposed scenarios.

Table 5.3 Comparison of the three scenarios

Architecture components	Low scenario	Medium scenario	High scenario
Data Management	Not Included	Not Included	Included
AI Development Environment	Included	Included	Included
Application Management	Included	Included	Included
Dataset Training Management	Not Included	Included	Included
Training Management	Not Included	Included	Included
Model Management	Not Included	Included	Included
CI/DEV OPS AI Environment	Not Included	Included	Included
AI Runtime Environment	Included	Included	Included

5.4.3 Low-level ambition scenario

Definition

The low-level ambition scenario includes the necessary functionalities and tools to develop irregular migration forecasting models as required by the main stakeholders or end-user. The low-level scenario is the simplest of the

three scenarios discussed in this assessment, as it comprises only the most basic functionalities for the development and execution of the models. This scenario can be understood as an entry point to the development of the AI-Tool, hence providing a Proof of Concept (PoC) tool, which the European Commission may assess before deciding to further implement one of the two alternative scenarios (mid-level and high-level scenarios).

These models will use a pre-existing data management system to leverage pre-existing data in order to reach forecasts on the chosen irregular migration flows. In essence, this scenario allows for a training to be performed and predictions to be run, but it requires a pre-created environment that includes data management and infrastructure maintenance procedures.

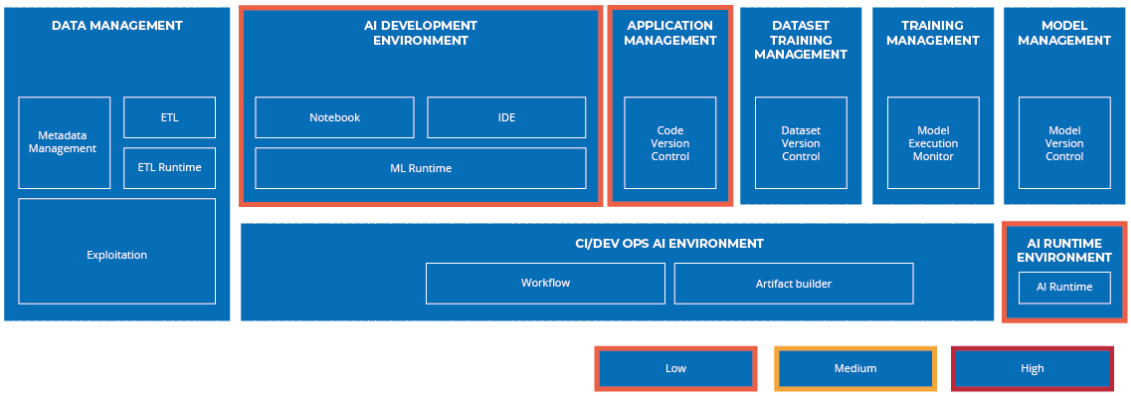
It would be a reasonable option if the tool was integrated into a currently existing system that already covers the remaining requirements. Under this scenario, the performance and reliability of the tool will be the lowest amongst the three scenarios given the limitations of the restricted functionalities available.

The low-level ambition scenario therefore comprises:
- Minimal components for AI environment;
- Minimal automatization provided by the AI-Tool itself;
- Minimal methodology to support manual tasks;
- Coverage of code and model development and runtime for AI applications;
- Low availability of functionalities;
- Low availability of data processing.

Architecture components of the low-level scenario

The low-level scenario covers the Application Development, Training, and Execution phases of the AI Application lifecycle. The functionalities included in this scenario are covered by the following components:
- AI Development Environment;
- Application Management;
- AI Runtime Environment.

Table 5.4 Architecture components for low-level scenario



Estimated needs for human resources

The estimation of roles needed for each phase of the forecasting AI-Tool life (Application Development, Training and Execution) and their responsibilities under the low-level ambition scenario are provided in the following table. These roles and responsibilities are considered to be the baseline needed for the functionalities to be covered by this scenario, which acts as an entry point or Proof of Concept tool for evaluation in the short term.

Table 5.5 Profiles needed for the low-level scenario

Required profiles	Team	Involved staff	Application Development	Training	Execution
ML Architect	Architecture Team	2 persons	• Develop, construct, test the lifecycle of AI application; • Develop the continuous integration and continuous delivery pipelines.	• Maintain the lifecycle of AI application; • Maintain the continuous integration and continuous delivery pipelines; • Preparation for model deployment; • Develop A/B testing framework and test model quality.	• Model deployment: the validated model is deployed to a target environment to serve predictions; • Provide model execution monitor to keep track of all production executions and their results; • Functional monitoring.
Data Scientist / Statistician	Data Science Team	2 persons	• Develop custom data models and algorithms.	• Analyse model performance and select the final model; • Model Validation; • Collaborate with Architecture Team to prepare model for deployment.	
Expert Data Scientist		2 persons			
Business Knowledge Expert		2 persons	• Coordinate with Data Science Team to discuss the most appropriate models.	• Coordinate with Data Science Team to discuss the model performance and its explainability and fairness	• Interpretation? • Analysis and reporting? Communication?

Total involved staff	***8 persons***	

Partial operational costs derived from estimated human resources

The approach followed to provide a cost indication per scenario is based on the estimated number of individuals involved in the operation of the AI-Tool, and on their typology. Assumptions on these expected needs are based on team expertise, practitioners' knowledge and previous experience relating the implementation of the proposed architecture. As such; these need to be refined at the stage of design of the AI-Tool. The annual operational (and partial) cost of the AI-Tool under each scenario is derived from the costs of the estimated FTEs needed to operationalise each scenario.

These staff costs are estimated on the basis of an average annual salary for officials of 152k€.[46] As such, the indicative cost of operation per scenario is calculated on the basis of assumptions on the FTE costs hence providing an indicative cost of running for each scenario; hardware and software costs are excluded from this analysis as these need to be addressed under a design study. This approach is followed for the estimation of partial operational costs for each scenario; the table below provides the expected partial operational costs for the low-level ambition scenario.

Table 5.6 Costs for the low-level scenario

Required profiles	Team	Seniority	FTE	Partial Estimated Annual Operational Costs
ML Architect	Architecture Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Data Scientist	Data Scientists Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Expert Data Scientist		Senior	0,5	76 000 EUR
Business Knowledge	Business Knowledge	Senior	0,5	76 000 EUR

[46] EC Directorate-General for Human Resources and Security (European Commission), *2019 Annual update of the remuneration and pensions of the officials and other servants of the European Union and the correction coefficients applied thereto. 2019/C 420/05 2019* (Luxembourg: Publications Office of the European Union, 2019).

Expert	Experts Team	Junior	1	152 000 EUR
Total FTEs			***5,00***	***760 000 EUR***

5.4.4 Medium level ambition scenario

Definition

The medium level ambition scenario includes all the functionalities available within the low-level ambition scenario previously described. In addition to these functionalities, the present proposed scenario enhances the architecture by including continuous integration functionalities and constant development, training and maintenance of the models.

This scenario may be seen as an option to implement an AI-Tool that will need to be integrated within existing Justice and Home Affairs large scale IT systems in order to cover all the functionalities necessary for operation.

Under this scenario, the performance and reliability of the tool will increase with respect to the low-level scenario because the added functionalities related to development, training and maintenance will considerably limit the deterioration of the system and deprecation of the models.

Like in the previous scenario, this would be a reasonable option if the tool were integrated into a currently existing system that could fulfil all data management needs. Although, on the contrary to the previous scenario, it would be more independent from any other pre-existing system.

The medium level ambition scenario therefore comprises:

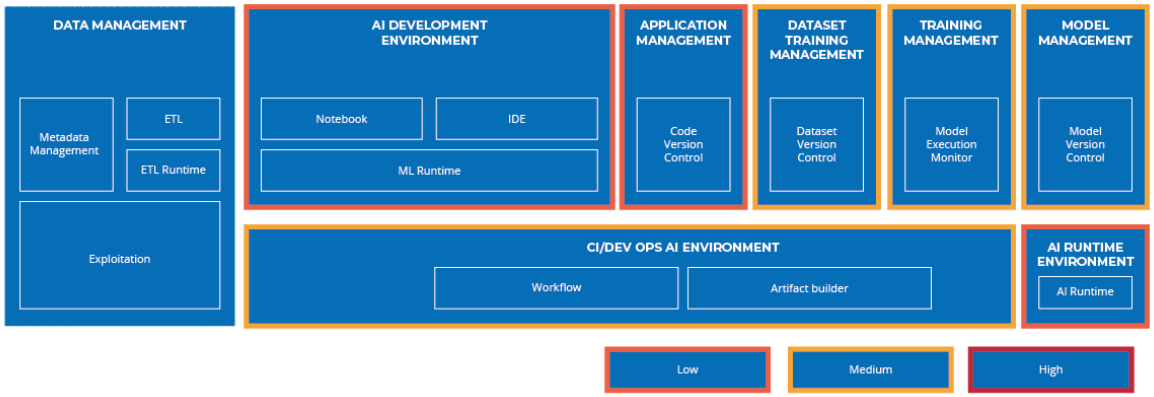
- Components to cover the complete AI lifecycle;
- Automatization for model deployment;
- Methodology covering all steps of the AI Application lifecycle;
- Coverage of all components of AI application definition;
- Medium availability of functionalities;
- Medium availability of data processing.

Architecture components

The mid-level scenario covers the Application Development, Training, and Execution phases of the AI Application lifecycle. The functionalities included in this scenario are covered by the following components:

- AI Development Environment;
- Application Management;
- Dataset training Management;
- Training Management;
- Model Management;
- CI/Dev Ops AI environment;
- AI Runtime Environment.

Table 5.7 Architecture components for medium-level scenario



Estimated needs of human resources

The estimation of roles needed for each phase of the forecasting AI-Tool life (Application Development, Training and Execution) and their responsibilities under the mid-level ambition scenario are provided in the following table. This table includes the roles and responsibilities detailed in the previous scenario and complements them with the roles and responsibilities coloured in orange. These roles and responsibilities are, therefore incrementing in order to reach the alternative scenario for the development of a tool to be integrated within a pre-existing system.

Table 5.8 Profiles needed for the medium level scenario

Required profiles	Team	Involved staff	Application Development	Training	Execution
Data Engineer	Architecture Team	2 persons	• Maintaining and validating ETL pipelines; • Maintaining data warehouses/ data lakes.	• Maintaining ETL pipelines; • Maintaining data warehouses/ data lakes.	• Maintaining ETL pipelines; • Maintaining data warehouses/ data lakes.
Senior Data Engineer		1 person			
ML Architect		2 persons	• Develop, construct, test the lifecycle of the AI application; • Develop the continuous integration and continuous delivery pipelines.	• Maintain the lifecycle of the AI application; • Maintain the continuous integration and continuous delivery pipelines; • Preparation for model deployment; • Develop A/B testing framework and test model quality.	• Model deployment: the validated model is deployed to a target environment to serve predictions; • Provide model execution monitor to keep track of all production executions and their results; • Functional monitoring.
Data Scientist	Data Scientists Team	2 persons	• Develop custom data models and algorithms.	• Analyse model performance and select the final model; • Model Validation; • Collaborate with Architecture Team to prepare model for deployment.	• Continuous training; • Functional monitoring.
Expert Data Scientist		1 person			

Business Knowledge Expert	Business Knowledge Experts Team	2 persons	• Coordinate with Data Scientists Team to discuss the most appropriate models.	• Coordinate with Data Scientists Team to discuss the model performance and its explainability and fairness.	
Total involved staff		***10 persons***			

Partial operational costs derived from estimated human resources

The approach followed to provide a cost indication for the present scenario is explained under section 5.4.3, on the partial operational costs for the low-level ambition scenario. The table below provides the expected partial operational costs for the mid-level ambition scenario.

Table 5.9 Costs for the medium level scenario

Required profiles	Team	Seniority	FTE	Partial Annual Estimated Operational Costs
Data Engineer	Architecture Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
ML Architect		Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Data Scientist	Data Scientists Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Expert Data Scientist		Senior	0,25	76 000 EUR
Business Knowledge Expert	Business Knowledge Experts Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Total FTEs			***6,5***	***988 000 EUR***

5.4.5 High-level ambition scenario

Definition

The high-level ambition scenario is an end-to-end solution to the irregular migration forecasting problem. It would include all the necessary functionalities to provide an AI-Tool that is independent of any pre-existing system. Indeed, the presence of the data management architectural block gives freedom to the technical teams to scope data sources and manipulate them for further use during the modelling phase. Data storage and flows would be designed and implemented within the same system.

This scenario may be interpreted as an alternative option to implement a stand-alone and end-to-end AI-Tool which may be independent of any pre-existing Justice and Home Affairs large scale IT system.

Under this scenario, the AI-Tool would be a standalone system that would be conceived and developed from scratch. As such, it would limit dependencies with other IT systems from any stakeholder. The performance and reliability of the tool will thus be the highest, determined by its own functionalities and maturity.

The high-level ambition scenario therefore comprises:

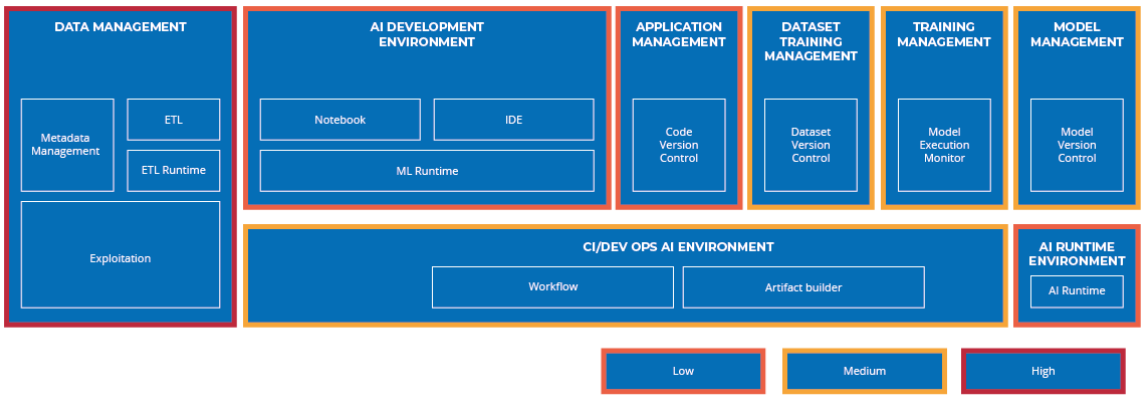
- Components to cover the complete AI lifecycle;
- Automatization for model deployment;
- Methodology covering all steps of the AI Application lifecycle;
- Coverage of all components of AI application definition;
- High availability of functionalities;
- High availability of data processing.

Architecture components

The low-level scenario covers the Data Treatment, Application Development, Training, and Execution phases of the AI Application lifecycle. The functionalities included in this scenario are covered by the following components:

- Data Management;
- AI Development Environment;
- Application Management;
- Dataset training Management;
- Training Management;
- Model Management;
- CI/Dev Ops AI environment;
- AI Runtime Environment.

Table 5.10 Architecture components of high-level scenario



Estimated needs for human resources

The estimation of roles needed for each phase of the forecasting AI-Tool life (Application Development, Training and Execution) and their responsibilities under the low-level ambition scenario are provided in the following table. This table includes the roles and responsibilities detailed in the previous scenario and complements them with the roles and responsibilities coloured in green. These roles and responsibilities are, therefore incrementing in order to reach the alternative scenario for the development of an end-to-end tool.

Table 5.11 Profiles needed for the high level scenario

Required profiles	Team	Involved staff	Data Treatment	Application Development	Training	Execution
Data Engineer	Architecture Team	2	• Analyse data quality and availability; • Design and build reliable and scalable ETL pipelines; • Design, develop data warehouses/ data lakes.	• Maintaining and validating ETL pipelines; • Maintaining data warehouses/ data lakes.	• Maintaining ETL pipelines; • Maintaining data warehouses/ data lakes.	• Maintaining ETL pipelines; • Maintaining data warehouses/ data lakes.
Senior Data Engineer		1				
ML Architect		1	• Design, the lifecycle of AI application; • Design continuous integration and continuous delivery pipelines.	• Develop, construct, test the lifecycle of AI application; • Develop continuous integration and continuous delivery pipelines.	• Maintain the lifecycle of AI application; • Maintain the continuous integration and continuous delivery pipelines; • Preparation for model deployment; • Develop company A/B testing framework and test model	• Model deployment: the validated model is deployed to a target environment to serve predictions; • Provide model execution monitor to keep track of all production executions and their results; • Functional monitoring.

Required profiles	Team	Involved staff	Data Treatment	Application Development	Training	Execution
					quality.	
Big Data Architect		2	• Administration and maintenance of the project infrastructure.	• Administration and maintenance of the project infrastructure.	• Administration and maintenance of the project infrastructure.	• Administration and maintenance of the project infrastructure.
Data Scientist Expert Data Scientist	Data Scientists Team	2 1	• Coordinate with Experts Team to better understand the problem and the main objectives; • Discuss with Experts Team the possible datasets that can be used for irregular migration forecasting; • Collaborate with Architecture Team to discuss with the datasets with Data Engineers and the possible ways of data preparation.	• Develop custom data models and algorithms.	• Analyse model performance and select the final model; • Model Validation; • Collaborate with Architecture Team to prepare model for deployment.	• Continuous training; • Functional monitoring.

Required profiles	Team	Involved staff	Data Treatment	Application Development	Training	Execution
Business Knowledge Expert	Business Knowledge Experts Team	2	• Clarify the scope and main objectives; • Suggest possible appropriate datasets; • Provide expert knowledge on irregular migration flows.	• Coordinate with Data Scientists Team to discuss the most appropriate models.	• Coordinate with Data Scientists Team to discuss the model performance and its explainability and fairness.	
Total involved staff		***12***				

Partial operational costs derived from estimated human resources

The approach followed to provide a cost indication for the present scenario is explained under section 5.4.3 on the partial operational costs for the low-level ambition scenario. The table below provides the expected partial operational costs for the high-level ambition scenario.

Table 5.12 Costs for the high-level scenario

Required profiles	Team	Seniority	FTE	Partial Annual Estimated Operational Costs
Data Engineer	Architecture Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
ML Architect		Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Big Data Architect		Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Data Scientist	Data Scientists Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Expert Data Scientist		Senior	0,5	76 000 EUR
Business Knowledge Expert	Business Knowledge Experts Team	Senior	0,5	76 000 EUR
		Junior	1	152 000 EUR
Total FTEs			***8***	***1 216 000 EUR***

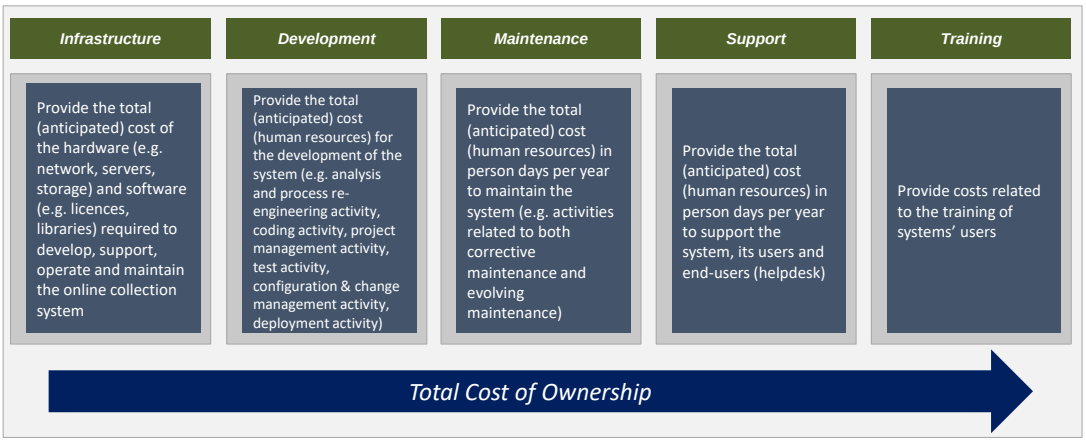
5.5 Further costs considerations

For the purpose of evaluating the costs associated with the implementation of an ICT system such as the AI-Tool for irregular migration forecasting; its functional and non-functional requirements need to be assessed against a cost analysis framework. These requirements should be identified through an elicitation process to take place within the frame of a design study of the AI-Tool.

The sum of the costs analysed under this framework provides an estimate of the Total Cost of Ownership (TCO) related to each scenario. As indicated in the ICT impact assessment guidelines[47] of the European Commission, while defining these costs, one should take into account whether these costs are an investment or operating costs. This framework is based on the VAST[48] categorisation which stipulates that for technology evaluations, five categories of costs should be analysed.

Within the frame of the present study, the operational assessment of the AI-Tool for irregular migration forecasting has provided a high-level estimation of the costs related to the maintenance category of the framework, which indicates operating costs related to the operation (operational phase) of the solution and its improvements. This analysis does not provide estimates on the other types of costs which are one-off costs to be defined once further information from a design study, benchmark analysis, and implementation roadmap is established for the future implementation of the AI-Tool for irregular migration forecasting.

Table 5.13 ICT cost model categorisation



[47] Directorate General for Informatics (European Commission), *ICT impact assessment guidelines - Practical tools and guidelines for assessing ICT implications.*

[48] Directorate General for Informatics (European Commission), (2010).

5.6 Concluding remarks

The operational assessment within the frame of this study analyses the feasibility of a high-level architecture design and proposes three different scenarios for its operationalisation. These scenarios provide a plausible solution that responds to the business needs expressed by the European Commission.

5.6.1 Implementation of a tailored architecture

In order to fulfil these objectives, this operational analysis proposes a clear set of options for implementation declined in three different scenarios.

These scenarios developed within the frame of this study can be summarised across four key characteristics:

- **Flexibility:** Enhancing the tool by including continuous integration functionalities and constant development, training and maintenance;
- **Completeness:** The scenarios are incremental, and the decision to move from one to the next level should be based on the implications of additional use cases to be solved and/or considering the PoC to evolve into a complete project;
- **Scalability:** Scenarios are not mutually exclusive, the idea is to complete the gaps;
- **Quality:** The quality of the forecasting outcome will depend on the capacity of the tool to intake data of standardised quality. Given the absence of data treatment functionalities in the proposed low and medium level scenarios, the quality of the data intake will lay beyond the control of the operating staff. This is the baseline level of quality. The high-level scenario will allow an independent data quality level, given its additional functionalities;
- **Costs:** Since at this stage, a TCO estimation is not possible, we provide below the summary of the costs of operation based on the necessary FTEs.

Table 5.14 below provides a comparative summary of the three scenarios based on the above-mentioned characteristics.

Table 5.14 Comparative summary of the three scenarios

	Flexibility	Completeness AI lifecycle	Scalability	Quality of outcome	Costs (FTEs)
Low-level scenario	--	--	Yes	Baseline	722 000 EUR
Medium level scenario	Low	Low	Yes	Baseline	950 000 EUR
High-level scenario	High	High	Yes	Independent data	1 216 000

				quality	EUR

It is important to notice that the low-level scenario represents an entry point for evaluation (a Proof of Concept tool) while the medium and high-level scenarios represent alternative options for the development of an integrated and/or an end-to-end solution in time. Further details about the development of the solution, its main characteristics functionalities, and the assessment of implementation options will need to be elicited during a design study that should cover the technical specificities related to each scenario.

5.6.2 Next steps

Our recommendation is to implement these scenarios as an incremental process based on the business needs (proof of concept, additional use cases). The steps to follow to ensure such implementation of the solution include:

- The elicitation and management of requirements;
- The declination of the proposed architecture in Application Building Blocks (ABBs);
- The declination of the ABBs into Solution Building Blocks (SBBs);
- The conduction of a benchmark analysis to select the cost-effective SBBs;
- The validation of the SBBs by the stakeholders;
- The development of the Low-level Design;
- The preparation of the Implementation Roadmap;
- The preparation of procurement process to initiate the acquisition of licenses and hardware necessary.

6 LEGISLATIVE ASSESSMENT

6.1 Objective and summary of the legislative assessment

Summary

The goal of the desk research for the **legislative assessment** was to assess the EU JHA agencies' suitability and feasibility to use and/or manage the AI-Tool, have access to its information, or supply data. To do so, the LIF explored the competences of EU JHA agencies previously identified as being the most relevant during the preparation of the Inception Report. In doing so, their legal basis was delineated. This evaluation was primarily performed by examining the relevant organisational regulations stipulating the prerogatives of the selected agencies. Furthermore, the agencies' mandates were matched against the envisaged functionalities of the AI-Tool. The main findings of the legislative assessment are as follows: i.) the EU primary legislation does provide a legal basis for the AI-Tool in question, in particular, Art. 77 of the TFEU; ii.) the development and operation of an AI forecasting tool by any of the EU JHA agencies should be aligned with all pieces of EU primary legislation. This includes the TEU, TFEU and the EU Charter of Fundamental Rights and the values these documents promote; iii.) EU secondary legislation needs to be duly considered (Dublin Regulation, Schengen Borders Code, EURODAC Regulation, VIS Regulation, EUROSUR Regulation now recasted in the new EBCG Regulation, EBCG Regulation, Qualification Directive, Directives regarding procedures related to international protection) in the design of the AI-Tool's functionalities and algorithms.

Interlinkages with other tasks

The legislative assessment is inextricably related to all the other assessments and tasks (general, operational, technical and organisational one) as it requires input from them. For instance, the design of the tool, data sources feeding the tool, its operation and functioning, who will have access to it, for what purpose and usage should be cross-referenced with the respective legal provisions. Additionally, certain legal matters need to be considered by the risk assessment. The results from the trustworthiness assessment likewise need to be considered from a legal point of view. The available technical functionalities of the tool as well as the relevant data sources, will be subject to legal feasibility analysis.

6.2 Theoretical framework

The European Union is founded on the basis of freedom, democracy, the rule of law and respect for fundamental rights, which form the cornerstones of the Union, its policies and practices. As such, and as delineated within its primary treaties- the TEU and the TFEU- the EU has competence in the context of humanitarian and refugee law. Furthermore, the vast fundamental rights framework, in particular the EU Charter of Fundamental Rights, ensures that

the EU JHA agencies operate in adherence to all such fundamental rights in the context of exercising these competences.

It is worth noting that EC has recently presented its proposal for a New Pact on Migration and Asylum, which will inevitably change the current approach to some extent. It puts an important emphasis on cooperation with third countries and encompasses a few guidance and relevant legislative proposals. For instance, a proposal regarding amendments in asylum and migration management, common asylum procedure, border screening of third-country nationals, manage crisis situations and enlarging Eurodac database. Its main aim is to improve the effectiveness of the procedures and thus enhance the trust in terms of cooperation among MSs and EU bodies concerning these matters while 'striking a new balance between responsibility and solidarity'.[49]

The adoption of legislative proposals listed in the pact might influence on the concerned tool. In particular, the envisaged amendments to the pre-entry screening of third-country nationals, Eurodac system, reforms related to the Schengen border code, the establishment of the new EU JHA agency for Asylum and the proposal about all information systems for border and migration management to all work together by 2023.

The current section will examine the up-to-date legal framework with an emphasis on EU primary and secondary legislation (both Directives and Regulations) to assess whether the implementation of the AI-Tool is feasible with respect to its envisaged functionalities and to evaluate whether a further amendment of the current legislation is required for this to be the case.

6.2.1 Primary Legislation

Following a comprehensive examination of the existing body of primary legislation, the creation of the AI-Tool for the delineated purposes has been determined to have a legitimate basis. To this aim, both the Treaty on the Functioning of the European Union and the Treaty on European Union were analysed, and the corresponding provisions were taken into account.

Article 67 (2) TFEU contains the provision for establishing a common area of freedom, security and justice through "a common policy on asylum, immigration and external border control".

Article 77 TFEU warrants the carrying out of individual checks and monitoring of EU external border crossings. (Art.77 (1) (b)). Furthermore, the TFEU also envisages "the gradual introduction of an integrated management system for external borders" (Art.77 (1) (c)).

Article 78 TFEU provides for the development of a common policy on asylum, subsidiary and temporary protection. Article 78 (g) requires the establishment of partnerships and cooperation between the EU and third countries for the better management of irregular migration flows.

Article 79 TFEU requires the development of a common immigration policy for the management of migration flows and illegal immigration.

[49] European Commission, *New Pact on Migration and Asylum* (Luxembourg: Publications Office of the European Union, 2020).

The European Commission is also granted the ability to collect any information and conduct checks for the performance of its designated tasks. [50]

6.2.2 Secondary Legislation

The proper functioning of the AI-Tool and feeding of the algorithms should be regulated through a legal framework. Therefore, amendments to the secondary legislation will be necessary in order to allow for the effective exchange of information between the relevant EU JHA agencies.

A thorough analysis of the EU secondary legislation has determined that the envisaged AI-Tool could operate within the context of the EU, more specifically that the objective to populate the tool with statistical data in order to achieve the purposes within the scope of the present study, is compatible with the EU Directives and Regulations examined, namely:

Regulation (EU) No 2019/1896 on the European Border and Coast Guard, establishes that the European Border and Coast Guard Agency (FRONTEX) shall "monitor migratory flows and carry out risk analysis as regards all aspects of integrated border management" (Article 10 (a)). It integrates the "European Border Surveillance System" (EUROSUR) within the European Border and Coast Guard in order to strengthen "the exchange of information and for operational cooperation" between Frontex and MS's national agencies, and to boost situational awareness (Article 18). Article 12 obliges the Agency to exchange information with "relevant Union bodies, offices and agencies" so that it could analyse any underlying risks, collect relevant statistical data, and assess the situation in third countries, for which the appropriate tools are to be developed (Art.12 (2)). Article 26 provides that the Agency shall "establish and maintain a European situational picture" for identical purposes, including obtaining information regarding unauthorised secondary movements. Article 28 stipulates that the EUROSUR fusion services shall be coordinated by Frontex so that information is provided on the situation at both the external borders and pre-frontier area by means of monitoring designated areas within maritime (Art.28 (c)) and air borders (Art.28 (d)), migratory flows (Art.28 (g)), media monitoring (Art.28 (h)), and analysis of information derived from large-scale information systems (Art. 28 (i)). Article 29 (3) allows the Agency to conduct a risk analysis encompassing all border management aspects, including for the development of "a pre-warning mechanism". This Regulation, therefore, in general, allows for FRONTEX to manage an AI-Tool for the purposes of forecasting irregular migration flows. However, the appropriate amendments need to be included to expressly delineate the relevant powers and the responsibilities of FRONTEX in order to host such an AI-Tool, clarifying the purposes of the Tool, how it will operate in practice, the effective population with data of the AI-Tool and the access to the end-results from it, including safeguards for the alignment of its algorithms with the fundamental human rights, for its security and its proper and effective functioning, control, monitoring and upgrade mechanisms, etc.

Regulation (EU) No 439/2010 establishes a European Asylum Support Office (EASO)[51] which strengthens the implementation of a Common European Asylum System and is responsible for managing the exchange of information on asylum between Member States (Article 3). Article 9 allows EASO to collect and analyse information for the purposes of managing asylum pressures and

[50] Treaty on the Functioning of the European Union, Article 337.
[51] REGULATION (EU) No 439/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 May 2010 establishing European Asylum Support Office.

sanctions it to either utilize "early warning systems and mechanisms" or "set up an early warning system for its own purposes" (Article 9 (3)). This Regulation, therefore, in general, allows for EASO to manage an AI-Tool for the purposes of forecasting irregular migration flows. However, the appropriate amendments need to be included to expressly delineate the relevant powers and the responsibilities of EASO in order to host such an AI-Tool, clarifying the purposes of the Tool, how it will operate in practice, the effective population with data of the AI-Tool and the access to the end-results from it, including safeguards for the alignment of its algorithms with the fundamental human rights, for its security and its proper and effective functioning, control, monitoring and upgrade mechanisms, etc.

The relevant proposals, contained within the new Pact on Migration and Asylum, were also considered. The Pact includes proposals by the European Commission for reforms regarding asylum, and where political consensus has been reached, however, are currently pending to be concluded. [52]

Proposal 2016/0131 (COD) for a Regulation of the European Parliament and of the Council on the European Union Agency for Asylum, aims at transforming EASO into the European Union Agency for Asylum and enhancing its mandate, thus improving its "gathering and analysing information on the situation of asylum in the Union and third countries"[53] including for early warning and preparedness of Member States for asylum-related migratory flows. This Proposal may thus create an opportunity for the Agency to manage the envisioned forecasting AI-Tool, albeit express clarifications would need to be made therein that forecasting, and not only early warning, is to be encompassed. The Regulation would also have to recognise the Agency as feasible to host such an AI-Tool, clarifying the relevant powers and the responsibilities of the Agency in order to host the AI-Tool, clarifying the purposes of the AI-Tool, how it will operate in practice, the effective population with data of the AI-Tool and the access to the end-results from it, including safeguards for the alignment of its algorithms with the fundamental human rights, for its security and its proper and effective functioning, control, monitoring and upgrade mechanisms, etc.

Proposal 2020/0279 (COD) for a Regulation of the European Parliament and of the Council on asylum and migration management, has set in its scope to replace the Dublin Regulation and relaunch the reform of the Common European Asylum System (CEAS).[54] In this context, it creates a common framework and a comprehensive approach to asylum and migration management which is also to be applied to situations of "irregular migration and unauthorised movements"[55], addressing its root causes.[56] This, too, includes a new solidarity mechanism, which aims to assists with migratory pressure and is set to be flexible enough in order to answer migratory

[52] European Commission, *New Pact on Migration and Asylum: Questions and Answers*, (Website, 23 September 2020) accessible at: https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_1707.

[53] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Union Agency for Asylum and repealing Regulation (EU) No 439/2010, COM/2016/0271 final - 2016/0131 (COD), p.7.

[54] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on asylum and migration management and amending Council Directive (EC) 2003/109 and the proposed Regulation (EU) XXX/XXX [Asylum and Migration Fund], 2020/0279 (COD), p.1.

[55] Ibid.

[56] Ibid, Article 3 (a.).

situations overall.[57] The European Agency for Asylum takes over the objectives of the early warning mechanism, established in the Dublin Regulation, and is also to facilitate "the development of practical tools" for the purposes of the proposed Regulation. The present proposed Regulation, therefore, may sanction the European Union Agency for Asylum to manage the envisioned forecasting AI-Tool, albeit it would also have to recognise the Agency as feasible to host such an AI-Tool, clarifying the relevant powers and the responsibilities of the Agency in order to host such an AI-Tool, clarifying the purposes of the Tool, how it will operate in practice, the effective population with data of the AI-Tool and the access to the end-results from it, including safeguards for the alignment of its algorithms with the fundamental human rights, for its security and its proper and effective functioning, control, monitoring and upgrade mechanisms, etc.

6.2.3 Fundamental Rights Framework

The legislative assessment delineates the relevant fundamental rights framework, which has to be born in mind within the scope of the AI-Tool's operation and the purposes for which it is used. Only the relevant fundamental rights provisions and whether and what type of safeguards need to be applied are examined.

The development and operation of such an AI-Tool by any of the EU Bodies/ Agencies should be aligned with the EU Charter of Fundamental Rights as it is a primary source of law as provided by Art.6 (1) TEU.[58] The potentially concerned fundamental rights and their relevance are analysed in details including the right to non-discrimination[59], the right to asylum[60], the right to respect for private and family life[61], the right to protection of personal data[62] and the right to good administration[63] in terms of this particular AI forecasting tool. Article 51 of the Charter stipulates that the Member States need to comply with the provisions in the Charter in all instances that involve their application of Union law. The FRA's Practical Guidance on Border Controls and Fundamental Rights at External Land Borders states that, when National Integrated Border Management (IBM) strategies are implemented, they still need to respect all aspects concerned with fundamental human rights.[64]

The AI-Tool application itself can neither directly contribute to nor negatively affect the exercise of fundamental rights *per se* as the main purposes of the Tool would be to forecast irregular migration at EU external borders and particularly the number of irregular migrants arriving at a point of the EU external borders. A potential risk to indirectly affect certain fundamental rights may occur on the occasion that the modelled algorithms themselves are biased or adequate control is not provided.

Other situations which might also have an impact on certain fundamental rights may arise, for instance, if the used algorithms are initially wrong or

---

57 Ibid, p.18.
58 Treaty on European Union, Article 6.
59 The Charter of Fundamental Rights of the European Union, Article 21.
60 The Charter of Fundamental Rights of the European Union, Article 18.
61 The Charter of Fundamental Rights of the European Union, Article 7.
62 The Charter of Fundamental Rights of the European Union, Article 8.
63 The Charter of Fundamental Rights of the European Union, Article 41.
64 European Union Agency for Fundamental Rights, *Border controls and the fundamental rights at external borders: Practical Guidance* (Luxembourg: Publications Office of the European Union, 2020).

inadequate (even if they are not discriminatory) and thus, impede the effective operation of the tool. In such situations, the results coming out of it should be regularly assessed along with adapting and correcting the inadequate algorithms. Otherwise, relying on these results/forecasts might hinder the EU JHA agencies' abilities to respond adequately and effectively in case of irregular migration flows. Therefore, poor administrative services might occur, such as the inability to review asylum applications in a timely manner.

Fundamental rights' infringements might also occur if the forecasts/results of the AI-Tool are accessed and/or used for different and/or incompatible with the primary envisaged purposes, which might be contradictory with fundamental rights. For example, if results are used to determine how to prevent people from applying for an asylum, which is completely beyond the scope of the AI-Tool. It can be limited by the existing legal framework (as it does not allow for such actions) and/or by effective control over the access and use of the tool as well as transparency in terms of the purposes for which each EU JHA agency has used the tool and how much it has contributed. Additionally, access on a need-to-know basis should be ensured. Therefore, adherence to the following rights must still be taken heed of:

1. Right to non-discrimination: Discrimination in the scope of the AI-Tool may only be examined at group/community level. What is more, such discrimination could originate not from the AI-Tool itself but solely if the specific purposes for which the AI-Tool's forecasts are used are discriminatory, as well as the indicators they are based on. More specifically, discrimination may arise if the AI-Tool, in the scope of its analysis, makes forecasts, which may correspondingly raise or lower the perceptions of risks depending on the specific group/community an irregular migration influx is forecasted to originate from. As a guarantee that this will not take place, the appropriate rules, delineating the *modus operandi* of the AI-Tool as well as its established purposes, must include the necessary degree of fairness and transparency. It must be ensured that the end-results of the Tool and their role in potential decision-making processes are not used for discriminatory treatment. Such decision-making processes should not be based solely on the end-results from the AI-Tool. What is more, suitable control must also be provided over the algorithms/scenarios and indicators the AI-Tool operates with and a periodic control must be exercised by the qualified personnel so that it is ensured no vulnerable groups' interests are negatively affected. As an additional measure, a subsequent refinement of the algorithms should be foreseen if an issue is detected (for example, not enough precise forecasting);
2. Right to asylum: The right to asylum is related to the right to non-discrimination and, once more, to the results themselves that the AI-Tool produces and the forecasts that it makes on this basis. On the occasion that a specific group/community is flagged as "risky", based on discriminatory or other criteria, this right may be negatively affected. Thus, as stated above, the appropriate and periodical controls must be exercised over the overall operation of the AI-Tool, including precision of the algorithms, if deemed needed;

3. Right to protection of personal data: Right to respect for private and family life: Personal data is not necessary for the AI-Tool operation. Therefore, no such data must be included within it, and any personal data anonymization must take place prior – namely, at the data source- which means that all data that populates the AI-Tool must already be anonymized on arrival. The right to respect for private and family life is examined in view of the use of personal data. The AI-Tool itself must also establish a regulatory framework that does not allow for the identification of individual natural persons neither within the information gathered by it nor by the results it produces as a subsequent analysis;
4. Right to good administration: This particular right is highlighted in relation to the opportunities that the AI-Tool employment presents for a better public service. The AI-Tool should be designed in a way to empower individuals to be heard, have access to documents[65] and seek reasoning for decisions that concern them. Additionally, as mentioned above, if the Agencies rely on the results from the tool to undertake certain decisions (e.g. regarding asylum application), the algorithms' quality should be ensured as well as their follow-up monitoring to guarantee accuracy and relevance.

It can be summed up that the AI-Tool itself is highly unlikely to raise any fundamental rights issues. However, the forecasts/results coming out of it might affect some of the abovementioned fundamental rights; therefore, human oversight should be foreseen.

6.3 Mandates and the competences of EU JHA Agencies and bodies

The mandates and the competences of the EU JHA Agencies identified as the most relevant ones have been carefully examined. The focus of the analysis is on the access and exchange of information among the respective agencies in terms of the tool. As a general rule, information sharing is purpose-driven, which requires reliable cooperation and a solid legal framework. A brief overview of the mandate of each EU JHA agency is included below.

6.3.1 EASO

The mandate of the European Asylum Support Office (EASO) is governed by its establishing Regulation 439/2010[66]. Its main role as an EU body is to facilitate, coordinate and enhance EU MS' cooperation on asylum in practical terms by improving the quality of the asylum process through uniform training and information on countries from where asylum seekers originate. EASO also assists MSs to fulfil their protection obligation, which is a crucial aspect within the asylum application procedure. EASO provides tailor-made assistance, supports capacity building, and develops quality control AI-Tools. Furthermore, it offers emergency support for the Member States subject to pressures (e.g. due to their geolocation, heavy burden, unforeseen flows) by

[65] "While respecting the legitimate interests of confidentiality and of professional and business secrecy", Art. 41. Para 2, lit. b., Charter of Fundamental Rights of the European Union.

[66] REGULATION (EU) No 439/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 May 2010 establishing European Asylum Support Office.

providing temporary support and assistance to repair or rebuild asylum and reception systems. Such support (e.g. deploy expert teams, facilitate the exchange of information, coordinate common actions) is usually provided to MSs upon their request. EASO is also responsible for coordinating the exchange of information regarding the matters related to asylum, resettlement, and relocation. It should guarantee compliance with international rules and procedures as well. As a key stakeholder, EASO contributes, actively, to the coherent implementation of the EU's common European asylum system (CEAS). In particular, it offers practical and technical support to MSs in terms of processing asylum applications comprehensively.

Additionally, EASO shares information and data, analyses and assessments at EU level early warning and preparedness system, provide regular regional outlook and analysis of asylum trends and push– and pull factors, as well as risk scenarios. Particularly, important is the fact that Art.9 of EASO Regulation provides the Agency to 'use existing early warning systems' and 'set up an early warning system for its own purposes', if deemed necessary. Furthermore, EASO is supposed to 'analyse data on any sudden arrival of large numbers of third-country nationals, which may cause particular pressure on asylum and reception system'.

These provide for potential legal competence of EASO to develop, maintain, host and operate such an AI forecasting AI-Tool. However, it should be noted that establishing an early warning system does not necessarily include setting up a forecasting AI-Tool, even though the early warning is supposed to comprise a predicting element itself. Therefore, once the foreseen functionalities of the AI-Tool are clarified additional assessment to the mandate of EASO should be conducted since it is clearly mentioned that such an early warning system will be only for EASO's purposes which will limit the scope of the concerned AI-Tool unless further amendments in the existing secondary legislation are introduced.

6.3.2 FRONTEX

New Frontex's Regulation[67] has been, recently adopted in order to strengthen the Agency's mandate and capabilities to support MSs in securing the EU external borders and ensure effective border control. Frontex aims to strengthen Europe's ability to respond to the challenges at its borders by sharing relevant information and providing regular risk analyses. Frontex supports MSs and third countries (TCs) in ensuring integrated border management. Thus, it is often referred to as 'a guardian' of the IBM by ensuring well-functioning border control. IBM is an EU concept introducing coherent and coordinated border management systems. It is designed with the aim to ensure that Governments maintain secure borders with as little inconvenience to travellers and cross-border trade as possible. It emphasizes coordination within and between border services, as well as international co-operation. With the adoption of the EBCG regulation, Frontex has come into existence with extended powers, competences, mandate and responsibilities to provide integrated control of the external borders by effective management of the irregular migration flows and ensuring a high level of internal security.

The Agency also helps in the prediction of challenges, which the EU countries face at their borders and support national authorities to respond adequately,

[67] Regulation (EU) 2019/1896 of the European Parliament and of the Council of 13 November 2019 on the European Border and Coast Guard and repealing Regulations (EU) No 1052/2013 and (EU) 2016/1624. PE/33/2019/REV/1.

including irregular migration. Furthermore, Frontex has an overview of European border control and migration management due to its 24/7 monitoring of the situation at the borders and beyond. Furthermore, Frontex delivers EUROSUR Fusion Services (EFS) to EU Member States and other relevant stakeholders. It encompasses automated vessel tracking and detection capabilities, software functionalities allowing calculations for detecting anomalies and predicting vessel positions, as well as precise weather and oceanographic forecasts. The Agency also conducts annual assessments (vulnerability assessment) of Europe's preparedness to face challenges at its borders and thus building resilience at Europe's borders (VA). In particular, Article 29 (3) provides that risk analysis encompass 'all aspects relevant to European integrated border management to develop a pre-warning mechanism' due to the fact that they provide detailed situational awareness of the external borders by identifying and monitoring the main/crucial driving and critical change factors. If the latter is broadly interpreted, it suggests that Frontex is also a viable option to host, maintain and operate the AI-Tool from a legal point of view based on the fact that EBCG Regulation envisages Frontex to provide early warning as well as a comprehensive and holistic approach to border management.

Here, it is worth noting that Frontex carries out also strategic risk analysis. For that purpose, it compiles data from various sources (border authorities, EU partner bodies, international organisations, its operational activities as well as open sources). These analyses provide situational awareness, which is informing about trends in irregular migration and cross-border activities affecting the EU external borders. They also serve as a basis for decision-makers to conclude upon certain priorities and undertake specific mitigation measures. Frontex strategic risk analysis is fed by collective risk analysis networks (there are six of them at the moment). With regard to that the Frontex Risk Analysis Network (FRAN) brings together its own experts and analysis with the ones from the regional risk analysis networks (Western Balkans Risk Analysis Network (RAN), Eastern European Borders RAN, Turkey-Frontex RAN and Africa-Frontex Intelligence Community) and the national risk analysis from the Member States. It enables the structured exchange of knowledge and information, performance of collaborative analysis and strategic reports regarding the situation at the EU external borders. The acquired information is used by EU Member States in arranging their border management activities as well as EU actors in their strategic and capacity building activities. Thus, Frontex has the relevant expertise to feed data into the tool as well as to analyse and carry out oversight over the results coming out of the concerned AI-Tool.

One of the main novelties, introduced by the recent Frontex Regulation is the establishment of the Standing corps and the ability of the Agency to reinforce European countries with additional border control experts, equipment as well as operational support at land, sea and air. Such a development reinforces Frontex mandate. Additionally, Frontex has the competence (Art.10) to 'develop and operate information systems' to facilitate secure data exchanges with other EU institutions with regard to occurring threats at the external borders. That provision broadly interpreted confers Frontex with the legal ground to develop, maintain and manage such an AI-Tool.

Frontex assists national authorities in return operations, including supporting the reintegration of returnees in non-EU countries. It also takes part in search and rescue operations. Frontex undertakes a key role in fighting cross-border crime by providing experts and training (e.g. common training for the

Migration Management Support Teams (MMSTs)). In accordance with the recent Regulation, Frontex will plan the EU's response to challenges at the external borders together with national authorities – rather than only reacting to crises. This provision can be widely interpreted to include the hosting and management of the AI forecasting AI-Tool as a mean to respond to these risks. Frontex develops and strengthens the EU's border management capacities through the development of integrated planning (capability development planning, contingency planning, operational planning) and its implementation. Frontex, along with the Member States, are key actors in border surveillance and control by regularly exchange border-related intelligence, situational awareness, and risk analysis. Moreover, it aims to develop technical standards for information exchange. Another essential element is the increased cooperation with third countries in terms of effective border control and return of irregular migrants. Frontex maintains a European situational picture and common pre-frontier intelligence picture that contain information on the situation at European borders and the pre-frontier area. This information is available to all Member States.

Furthermore, neighbouring Member States share the situational picture of their neighbouring external border sections with each other. The new Regulation foresees the expansion of EUROSUR fusion services to support border checks, air border surveillance, and the monitoring of irregular migration flows. EUROSUR enables the Member States to rapidly exchange information, ensure necessary cooperation and offer a joint response to challenges. Besides, Frontex is responsible for coordinating the so-called common application of surveillance tools: the Member States can request Frontex' assistance in selective monitoring of areas or vessels of interest for EUROSUR purposes by using tools like satellite imagery or ship reporting systems. This can be used to detect cases of irregular migration or cross-border crime, but also to locate a boat in distress. The logic behind these competencies might be extended to Frontex's ability to participate in the analysis of information/data obtained by the AI-Tool and to have access based on its level of restriction.

6.3.3eu-LISA

Eu-LISA is established under Regulation (EU) 2018/1726[68] to ensure the uninterrupted, efficient and effective operation of large-scale IT systems that guarantee security, rights and free movement within the EU. It provides technical and operational expertise to ensure that everyone from EU citizens to third-country nationals is free to live, work and travel. Eu-Lisa manages the operation of the SIS II system for sharing information on criminal matters to ensure the coordinated investigation of crimes that do not respect national borders, manages the Visa Information System (VIS) for fair and secure management of visa application and border entry processes. Eu-Lisa rolls out a system in terms of infrastructure, software and search. It has also improved technical controls that ensure the security of personal data and respect for the rules. Furthermore, it delivers and supports feasibility studies on various matters (e.g. Biometric Matching Service and Common Repository for Reporting and Statistics (CRRS)). Additionally, it operates the Eurodac[69]

[68] Regulation (EU) 2018/1726 of the European Parliament and of the Council of 14 November 2018 on the European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu LISA), and amending Regulation (EC) No 1987/2006 and Council Decision 2007/533/JHA and repealing Regulation (EU) No 1077/2011.

[69] REGULATION (EU) No 603/2013 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 June 2013.

system for the monitoring of asylum applications from those who may need protection under EU values and norms.

Taking into account the fact that eu-LISA has the clear mandate to host and manage all the above-mentioned large-scale systems and all the data comprised within them, eu-LISA can host the AI-Tool as a hub without using the available data for its own purposes. However, lack of clear legal provisions within its mandate and the eu-LISA limited access to restrictive EU information and documents would require further amendment of the applicable secondary legislation in this respect to ensure effective hosting and management of the Tool.

6.3.4 Europol

The EU Law Enforcement Cooperation Agency (Europol) is established by Regulation (EU) 2016/794 [70] in order to support and coordinate international law enforcement operations across a wide range of crime areas. It undertakes operational support activities to disrupt criminal and terrorist networks and prevent criminal activities. Europol supports the Member States and agencies and bodies in building operational and analytical capacity for investigations. The Agency hosts the European Migrant Smuggling Centre (EMSC). The latter predisposes the Agency's active participation in providing and accessing data to and from the AI-Tool. Europol also supports MSs in Joint Actions Days (JADs). Additionally, Europol has the mandate to 'prepare threat assessments, strategic and operational analyses and general situation reports', which task will also benefit from access to such an AI-Tool and its data.

6.3.5 European External Action Service

The European External Action Service (EEAS) is the diplomatic service/body of the EU. It assists and supports the High Representative for Foreign Affairs and Security Policy in carrying out his/her mandate. The EEAS also provides support to the President of the European Council, the President of the Council and to members of the European Commission in the exercise of their respective functions in the area of external relations. The EEAS actively participates in the implementation of the common foreign and security policy (CFSP) of the Union. The EEAS comprises of the Crisis Management and Planning Directorate (CMPD) and the Civilian Planning and Conduct Capability (CPCC). EEAS manages the early warning system of the EU able to identify factors and indicators which often are related to the occurrence of violent conflicts. It assists policy and decision-makers in managing risk factors and relocate resources, adequately and timely. However, it should be noted that it reflects solely violent conflicts. The overall nature of the EEAS legal framework and its mandate does not directly assume the hosting and management of such an AI-Tool.

It can be concluded that the current mandates of the above-mentioned EU JHA agencies do not provide an explicit legal provision in terms of hosting an AI-Tool. However, Regulations of EASO (Art.9) and FRONTEX (Art.29(3) and Art.10) contain articles which might be interpreted in a way that provides these two agencies with the legal right to develop, maintain, manage and operate such an AI forecasting tool. With regard to that, it is worth noting that there is a difference in terms of the Agency's competences, which will host the tool and have full access to its raw data and the agencies, which will

[70] Regulation (EU) 2016/794 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation (Europol) and replacing and repealing Council Decisions 2009/371/JHA, 2009/934/JHA, 2009/935/JHA, 2009/936/JHA and 2009/968/JHA.

have only access to the final outputs/forecasts of the tool. The latter ones most likely would need only minor amendments in terms of their respective legal frameworks.

6.4 Working arrangements

Working arrangements (WA) between the above-mentioned EU JHA agencies have been thoroughly examined (Annex C). The main purpose behind each of the Working arrangement is the establishment of a framework identifying the areas of cooperation based on common interest and setting the objectives and principles of mutual cooperation, including sharing data. In terms of information exchange, all of the reviewed WAs are general rather than containing particular articles about that. Therefore, more specific and goal-oriented working arrangements are necessary regarding information exchange between the respective EU JHA agencies, especially in terms of the AI-Tool. It is worth noting that the exchange of information, briefly, referred to within these WAs is often regulated under the agencies' establishing regulations (i.e. secondary legislation). For instance, Art.12 (2) of EBCG Regulation provides the Agency with the obligation to exchange information 'relevant to the performance of its tasks' with other relevant EU JHA agencies. Additionally, Art.15 of EBCG current Regulation prescribes the Agency's competence to 'take all necessary measures to facilitate the exchange of information relevant to its tasks'. Thus, the exchange of information between the Agencies should be primarily arranged within their regulatory legal frameworks (e.g. establishing regulations) and further specified under a particular WA to provide a solid legal ground. Thus, it might be concluded that Working arrangements between the EU agencies are based, primarily, on their regulations. So, they should be reviewed together rather than independently from one another.

On the other hand, there are other examples as the Europol's regulation, for instance, which establishes only the ability of the agency to exchange relevant information with EU JHA agencies rather than an obligation (Art.23). It also provides for the conclusion of WA for that purpose. While EASO is prescribed to have 'a mechanism for the exchange of information' to initiate and maintain a dialogue with 'relevant stakeholders'. In that particular case, WA *per se* might not be sufficient legal ground to ensure the necessary exchange of information regarding the concerned AI-Tool. This situation is corresponding to eu-LISA's case where 'the Agency may take all necessary measures to facilitate the exchange of information relevant to its tasks with the relevant Union agencies' (Art.37 (2)). Moreover, it provides EU JHA agencies to 'use the information received from the Agency only within the limits of their competences and insofar as they respect fundamental rights, including data protection requirements'. In other cases, additional cooperation plans are foreseen as an option to establish cooperation regarding particular activities and/or to limit the information exchange to some extend (e.g. Cooperation plans between Frontex and EASO, EASO and eu-LISA mentioned in Annex C). It should be duly noted that the cooperation plans are complementary to the working arrangements rather than separate documents providing the legal basis for the exchange of information on their own.

It can be summed up that information sharing is regulated at secondary legislation level for some EU JHA agencies, while for others, it is briefly mentioned within their regulations. So, a better and explicitly regulated exchange of data between the relevant EU JHA agencies is necessary to ensure adequate sharing in terms of the AI-Tool. This need for coherent and

improved coordination among the EU JHA agencies has been already acknowledged through the established multilateral cooperation framework. There is a possibility to include other relevant areas for cooperation once identified as of common interest included in most of the working arrangements. That provides an opportunity to extend further their exchange of information regarding the AI-Tool. It is important to note that the level of classification in terms of the exchange of information between the EU JHA agencies is not examined in detail under the working arrangements themselves. Only the fact that certain measures to ensure non-disclosure to third parties and secure retention of that data should be undertaken in each of them. Therefore, further examining the access level should be considered in terms of the AI-Tool available data. An overview of the existing working arrangements can be found in Annex G. With regard to the exchange of information between the relevant EU agencies and its legal basis, the modification of secondary legislation, working arrangement or adoption of additional cooperation plans are not necessarily independent. They complement each other, meaning, that there might be a need to amend, firstly, the secondary legislation before modifying working arrangement and respectively concluding cooperation plans.

7 ORGANISATIONAL ASSESSMENT

7.1 Objective and summary of the organisational assessment

Objective

The organisational assessment reviews what skills, resources, and coordination mechanisms are needed to host the AI-Tool and consequently, the most feasible options for locating such an AI-Tool. The results of this assessment ultimately inform the project's recommendations on where to best situate the AI-Tool and how to relay its outputs.

Summary

The decision on where to best situate the AI-Tool and how to relay its outputs hinges on three sets of tasks: operations, analysis, and dissemination. While some stakeholders have the operational capacity to build and host the AI-Tool, for example, their capacity in analysing data can vary significantly. In part, this reflects their capacities and mandates focusing on different areas of asylum and migration management that, by themselves, do not fully align with the scope of the AI-Tool (also see General Assessment and Legislative Assessment). Filling these gaps means either investing in building this operational or analytical capacity or developing a joint approach where the operational and analytical tasks are shared among multiple agencies, likely requiring investment in appropriate coordination mechanisms.

Decisions about which data sources the AI-Tool will need to access, on what timeframe, and whether any of these data sources are listed as restricted will also influence considerations of where to host the AI-Tool. The use of confidential data sources would lead to different levels of access for different stakeholders unless these data are anonymised before they are entered into the AI-Tool. In turn, sharing the outputs of the AI-Tool with different audiences will require technical knowledge about data-sharing regulations, as well as a broader awareness of what information (and in what format) is relevant to different stakeholders (e.g. policy teams vs. operational units).

During the course of our research, the main candidates that were put forward for consideration to host the AI-Tool were Frontex, the European Commission, EU-LISA, and EASO. We set out the current capacities of each candidate below, and identify areas where additional support or investment may be needed. However, this list is not definitive—other stakeholders (i.e. JRC, EEAS, Europol, Eurostat) could still contribute to the design of the AI-Tool, data analysis, and dissemination to other Union institutions, bodies, offices and agencies subject to specific working arrangements.

Interlinkages with other tasks

The organisational assessment takes stock of the insights produced in all other assessment and fits them within the broader scope of requirements for hosting an AI-Tool. This includes, for example, operational requirements to develop an AI-Tool infrastructure, legal requirements for data-sharing between different EU bodies and agencies, and general requirements to

identify the forecasting purposes and intended benefits of the AI-Tool. Moreover, the organisational assessment feeds into the other assessments by outlining the coordination mechanisms that need to be in place, including from a technical and legal viewpoint, to access and disseminate the AI-Tool's outputs to different audiences (e.g. operational units and policymakers).

7.2 Research activities and existing information gaps

An array of key stakeholders were interviewed to understand their preferences for accessing the outputs of an AI-Tool and the current or potential capacity of different agencies and institutions to host the AI-Tool. This included interviews with representatives from the European Commission, Frontex, Eurostat, EASO, EEAS, EU-LISA, and JRC, as well as representatives from the German government (regarding lessons from their Preview programme).

The consortium partners conducting this study worked closely with each other to assess the feasibility of the AI-Tool. As such, the scope and focus of the organisational assessment is informed by the analysis of other assessments relating to the development of the AI-Tool, the requirements for hosting and operating the AI-Tool, and any practical or legal constraints. Following the closed-door seminar, the analysis was further clarified, and some information gaps addressed in line with the participants' feedback.

7.3 Requirements for hosting the AI-Tool

To assess options for hosting the AI-Tool, we examined the skills and resources required to operate the AI-Tool, analyse its data, and disseminate its outputs to different stakeholders.

7.3.1 Operations

One set of considerations relates to the expertise and capacity required to operate the AI-Tool. The general, operational, and legal assessments all set out criteria that can inform our assessment of possible hosts.

Decisions about which data sources the AI-Tool will need to access, on what timeframe, and whether any of these data sources are listed as restricted could all help narrow down the list of potential hosts. So far, the data assessment suggests the AI-Tool is likely to use anonymised data instead of personal data, which would ease data collection and processing activities for operators of the AI-Tool. For example, Regulation (EU) 2018/1725 has increased accountability and data protection safeguards for Union institutions, bodies, offices and agencies dealing with personal data, but these guidelines do not apply to anonymous information. In short, using anonymised data is not only less sensitive but also legally and operationally less complex.

As described under the legal assessment, running the AI-Tool would require a legal base which currently exists under TFEU but might need to be expanded

further depending on the type of modelling the AI-Tool is intended to do. This has an impact on the organisational assessment insofar that candidates for hosting the AI-Tool likely require additional working agreements to facilitating data access and data sharing, especially if the AI-Tool will be operated jointly.

The operational and trustworthiness assessment also notes several more technical skills that stakeholders will need to possess or develop to operate the AI-Tool. These include an ability to analyse data quality and availability; to develop, construct, test the lifecycle of AI application; to develop custom data models and algorithms; to administer and maintain the project infrastructure; and to centralise information from different data models. In terms of resources, the host will need to use large data sets and run large-scale computations, in particular, where the AI models are based on machine learning and the use of algorithms with complex computations. As further outlined in Section 1.3., candidates for hosting the AI-Tool will also need to invest in their underlying infrastructure, both software and hardware.

7.3.2 Analysis

Alongside these operational requirements, another set of considerations relates to the capacity to analyse these data. Interpreting and contextualising these data will often require migration expertise, for example, to assess how irregular migration flows might be shaped by different drivers or policy contexts; whether certain "triggers" are likely to lead to internal, regional or international irregular migration; and the potential impact that irregular migration flows might have on the number of arrivals at the EU borders and the volume of intra-EU migration. The migration expertise needed to conduct these analyses also depends on the ambition level (low, medium, high) of the scenario, with higher-level scenarios likely requiring more analysts (or "Business Knowledge Experts") to help analyse the AI-Tool's outputs.

7.3.3 Dissemination

A final set of considerations relates to who will need to access the outputs of the AI-Tool and how frequently, which will inform the analysis both of where to locate the AI-Tool, and what coordination mechanisms will need to be in place. The ability to relay the AI-Tool's outputs effectively requires both familiarities with data-sharing regulations, especially in a context where sensitive and/or restricted data are being used, and a firm grasp of what information is needed by different stakeholders, including policymakers.

Interviews with different stakeholders emphasised that the outputs of the AI-Tool should be accessible and potentially be presented in different formats for different audiences. At an operational level, stakeholders recommended a succinct and data-driven product, such as a one-pager with concrete estimates (e.g. on the estimated number of arrivals at EU external borders) rather than broad ranges, that are regularly updated.[71] At a policy level,

[71] Considering that forecasting models are vulnerable to uncertainty and margins of error, this product should include information on how to interpret these estimates, including their underlying assumptions.

stakeholders recommended short briefing notes that interpret the outputs of the AI-Tool and ideally include recommendations on how to incorporate these findings into policy programming to ensure their usefulness. Across all stakeholders, however, there was a clear preference for shorter reports or briefings that analyse the raw data and which include infographics and visuals, and which can be readily shared through existing formal and informal communication channels. To simply this information-exchange, stakeholders voiced a preference for interactive dashboards where EU officials and policymakers can access (at different levels) the AI-Tool's outputs, similar to the KCMD Dynamic Data Hub. This dashboard could include reports, as described above, but also provide (restricted) access to data for operational units.[72]

7.4 Candidates for hosting the AI-Tool

Drawing on the interviews and our assessment of the skills and resources required to operate the AI-Tool and analyse and disseminate its outputs, we have identified several candidates for hosting the AI-Tool. The discussion below sets out the advantages and potential trade-offs each option offers.

7.4.1 Key candidates

Our analysis to date has identified several candidates for hosting the AI-Tool: Frontex, the European Commission, EU-LISA, and EASO. We have explored how they currently perform according to the criteria set out above, summarized briefly as follows:

- *The European Commission*. The advantages of hosting the AI-Tool at the European Commission (e.g. DG HOME, F2) include their ability to access confidential data, their experience in disseminating data to other EU JHA agencies, institutions and governments, and their significant analytical capacity. However, they currently lack the operational capacity to host an AI-Tool, so this would require building an IT system and determining how best to host the AI-Tool. The central contact point for operating the AI-Tool would likely be within DG JUST, where DG HOME's IT unit is located. DG DIGIT could support this set up through its experience in using AI models and applying machine learning;
- Frontex. The advantages of hosting the AI-Tool at Frontex include their ability to access confidential and timely data (reported by Member States on a daily basis) and their experience in analysing and disseminating data on irregular migration trends. Despite their existing operational resources dedicated to monitoring, producing and delivering informed situational reports and alerts, such as the Frontex Situation Centre (FSC), they require additional capacity to build an IT system capable of running complex

[72] For example, if tailored to the case study the intended purpose of these reports would be based on the decision-making purposes described in Section 4.3., and include data on policy changes and announcements, demographics and detections, the situation in third countries, and meteorological conditions.

computational models for the AI-Tool. The central coordination point for such a system would likely be within the Frontex' ICT Unit, responsible for the development, maintenance and support of the Agency's ICT infrastructure. While Frontex also has some analytical capacity, for example, through the Risk Analysis Unit, their mandate is focused on supporting Member States and Schengen Associated Countries in managing their external borders (including working with and in third countries, drawing on their Regional Risk Analysis and Liaison Officers Network) and migration management related to returns;

- EU-LISA. The advantage of hosting the AI-Tool at EU-LISA lies in its operational capacity to manage and host this type of large-scale IT system in-house. However, unlike Frontex, this would require giving EU-LISA the mandate to build and host the AI-Tool for the purpose of strengthening Europe's ability to respond at its borders. It would also require close coordination with another Agency that could then analyse and disseminate the outputs of the AI-Tool. Currently, there is limited interoperability between EU-LISA's work and that of EASO and Frontex, which limits the possibility of accessing and triangulating different data sources, such as under Eurodac. However, the planned Central Repository for Reporting and Statistics (CRRS) can potentially resolve this issue;
- EASO. The advantage of hosting the AI-Tool at EASO lies in its operational and analytical capacity based on EASO's early warning system, which is starting to look more at trigger points in third countries and pull factors in the EU, among other things. EASO could also use existing communication channels with other EU JHA agencies and institutions developed through its early warning system to disseminate the AI-Tool's outputs. However, these capacities would need to be strengthened to collect, process and interpret larger amounts of data, especially in areas where EASO traditionally does not operate (e.g. border control). Similar to Frontex, the central coordination point for operating the AI-Tool would likely be within the EASO's own ICT Unit. Similar to EU-LISA, operating the AI-Tool may also entail expanding EASO's current mandate to be able to fully support the EU's border response by forecasting irregular migration flows.

7.4.2 Support from other stakeholders

Our analysis and key informant interviews reveal there are several other stakeholders who, while not currently in a position to host the AI-Tool, could still contribute to its design, data analysis, and dissemination to other Union institutions, bodies, offices and agencies subject to specific working arrangements. These include:

- JRC could leverage its experience in conceptualising forecasting AI-Tools to support the design of the AI-Tool. JRC already has similar experience working with EEAS on their conflict risk model, for example by reviewing the conflict definition for the model through an expert workshop or discussing timeframes, looking at structural risk factors to understand better and pinpoint when the risk of conflict would materialise;
- EEAS could support data analysis efforts by providing third-country assessments at regular intervals (e.g. monthly, quarterly or annually), provided they are declassified. This can serve two different purposes: on

the one hand, it may help calibrate forecasts in line with up-to-date information on the irregular migration trends and policy environment in a given third country and, on the other hand, it can help validate the AI-Tool's forecasts in the longer term;

- Europol could support the data collection and analysis of smuggler activities in ports and boats on the Mediterranean and help identify changes in migratory routes and patterns. The European Migrant Smuggling Centre (EMSC), hosted by Europol, already has the capacity to support police and border authorities to this end and has established close cooperation with Frontex. Europol could also support disseminating the AI-Tool's outputs, for example through the annual EMSC report or through more timely reporting via bi-weekly or monthly general situational reports;
- Eurostat could have a limited role in supporting the AI-Tool by providing data for long-term analyses and providing guidance on how to collect and process large amounts of data. Although Eurostat provides harmonised data that are heavily used by specific DGs and the Member States, their time limits of two months (i.e. Member States have to provide statistics within two months) mean they cannot be used for short-term forecasts of 1-4 weeks.

7.5 Conclusion

At present, there is not a single clear candidate for hosting the AI-Tool. Instead, there are who each bring different advantages to the table, but who would also need additional support to meet all of the operational, analysis, and dissemination requirements. While some stakeholders have the operational capacity (although not currently the legal base) to build and host the AI-Tool, they do not have the capacity to analyse these data, while other candidates currently have the analytical but not the operational capacity. In line with the legislative assessment, the mandates of all these different candidates would need to be expanded, although this requires smaller amendments for Frontex and the European Commission compared to EASO and eu-LISA.

This raises the question of whether it is preferable to invest in the capacity of a single host, or whether instead, the EU should consider a joint approach where the operational and analytical tasks are shared among multiple agencies. In the first scenario, the decision about where to host the AI-Tool would hinge on the EU's ability to invest in building a host's organisational, analytical, and dissemination capacity (and whether there is more scope to allocate funding for some of these activities compared to others). In the latter scenario, the decision would centre on the extent to which the EU could invest in coordination mechanisms, such as additional working agreements, to facilitate data access and data sharing.

Options for hosting the AI-Tool may be further narrowed down by weighing the eventual format of the AI-Tool (which may lend itself to one host or

another) and the incentives of potential candidates to cooperate on the AI-Tool's operation, analysis and dissemination. In both scenarios, it is key that appropriate measures are put in place to analyse and filter the AI-Tool's outputs before they are circulated to a wider audience, both to avoid data misinterpretation and to better tailor information to different audiences depending on the type(s) of decision-making the AI-Tools aims to support (e.g. policy, operations). Ultimately, these considerations are also subject to the question of who will own the outputs of the AI-Tool, since ownership may influence the willingness of stakeholders to, for example, invest resources (e.g. budget, staff capacity) to support its design and operation.

8 TRUSTWORTHINESS ASSESSMENT

8.1 Objective and summary of the trustworthiness assessment

Objective

The objective of the trustworthiness assessment is to assess the trustworthiness of the AI-based tool (human agency and oversight, technical robustness and safety, transparency, accountability).

Summary

In 2019, the European Commission High-level Expert Group on Artificial Intelligence (AIHLEG) published the Ethics Guidelines for Trustworthy AI. The Guidelines define the three components of trustworthy AI and put forward a set of four ethical principles and seven requirements that AI systems should meet to be deemed trustworthy.

The present assessment proposes the framework, techniques and AI-Tools that should be used throughout the AI system's entire life cycle in order to ensure trustworthiness. The trustworthiness assessment is structured by four trustworthiness components: Fairness, Explainable Artificial Intelligence (XAI), Functional Monitoring, and Governance.

Interlinkages with other tasks

The trustworthiness assessment is complementary to the operational assessment; it provides the theoretical basis for the understanding of how the proposed architecture of the AI-Tool to be implemented can be understandable, reliable and bias-free. In addition, this assessment should in principle be complementary to the legal assessment and the organisational assessment as it lays down the elements to be considered from a perspective of legal and organisational constraints on the use of technology for decision making.

8.2 Trustworthiness assessment framework

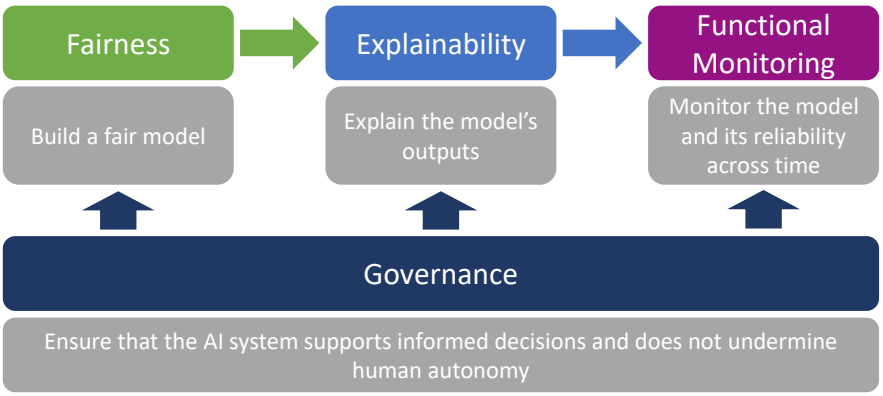
Given the objective of this assessment which is to evaluate the capability of the studied AI-Tool to comply with the following criteria, we develop the proposed trustworthiness assessment framework that will help address the requirements as outlined by the European Commission.

- **Human agency and oversight:** the AI system should empower human beings, allowing them to make informed decisions, foster their fundamental rights, and have at disposal proper oversight mechanisms;
- **Technical robustness and safety:** the AI system should be resilient and secure. It needs to be safe, accurate, reliable and reproducible to ensure that unintentional harm is minimized and prevented;

- **Transparency:** the AI system, data and business models should be transparent, traceability mechanisms should ensure that stakeholders are adequately informed of the system's capabilities and limitations;
- **Accountability:** the AI system should include mechanisms to ensure responsibility and accountability, in particular, based on audibility and adequate and accessible redress.

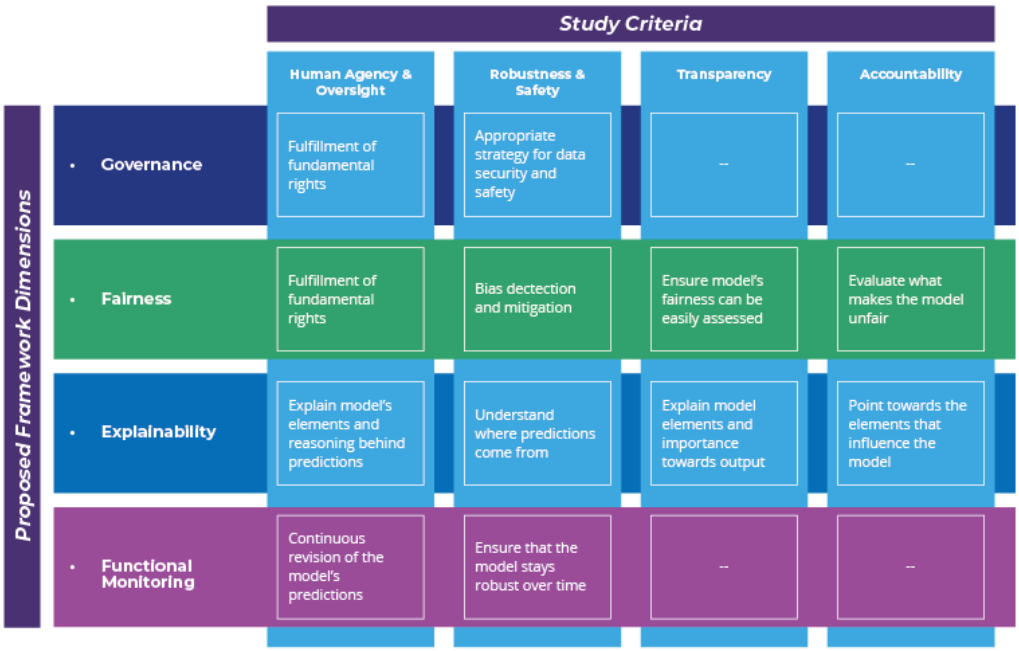
In order to address these criteria, we develop below the proposed trustworthiness assessment framework, which covers the different trustworthiness requirements based on four dimensions: Fairness, Explainable Artificial Intelligence (XAI), Functional Monitoring, and the transversal dimension of Governance.

Table 8.1 Trustworthiness framework



This framework allows to address the study criteria (see list above) according to the following mapping, which addresses the findings of the European Commission High-level Expert Group on Artificial Intelligence (AIHLEG) Ethics Guidelines for Trustworthy AI criteria.

Table 8.2 Mapping between the proposed framework and the European Commission High-level Expert Group on Artificial Intelligence (AIHLEG) Ethics Guidelines for Trustworthy AI criteria



8.2.1 Fairness

This section describes the Fairness dimension of the trustworthiness framework, which relates to the techniques used within AI in order to ensure that unintentional harm is minimized and prevented.

The AI-Tool in scope will forecast irregular migration flows to support decision making on irregular migration and asylum-seeking within European borders. The developers of the system should ensure that the tool uses adequate machine learning models and that the data is not biased against any particular attribute such as nationality, race, gender, age or religion, and sexual orientation. Otherwise, the algorithm would perpetuate discriminatory trends.

This section describes the main steps consider during the conception of an AI-Tool; it also provides an analysis of the techniques for bias detection in the datasets as well as the approaches for bias mitigation.

Overview of the steps to create a fair AI-Tool

In order to create a fair AI-Tool, four steps must be followed.

A. Design Fairness Goals

When designing a fair AI-Tool, two core elements must be covered: a) the main objectives of the tool, and b) the identification of the potential harm that the tool can cause.

In order to determine this, the following questions can be used:

- What forecasts is your tool making?
- How do you plan to use these forecasts?
- What are the short and long term goals?

In order to identify the potential harm associated with the AI-Tool, an analysis of the potential source of bias must be conducted. The most common sources of bias are the following:

- **Sample bias -** occurs when one population is overrepresented or underrepresented in a training dataset. An example of this would be an asylum seekers' AI-Tool that has been predominantly trained on irregular migrants from dominant nationalities on the training dataset;
- **Label bias** - occurs when the annotation process introduces bias during the creation of training data. For example, the people labelling the data might not represent a diverse group of locations, ethnicities, languages, ages, and genders, and can bring their implicit personal biases into their labels. This can lead to labels that are skewed in ways that yield systematic disadvantages to certain groups;
- **Outcome proxy bias** - occurs when the machine learning task is not specified appropriately. For example, if one would like to forecast the likelihood of a person crossing a border, using arrests as a proxy is biased because arrest rates are greater in countries with the more restrictive political regimes.

B. Evaluate and prepare your data

Once fairness goals are defined, the dataset should be evaluated and prepared. The evaluation of the data should be done against the fairness goals previously defined in order to measure whether the data is fairly represented.

The evaluation of data should follow a three-step process:

- **Identification of protected attributes**: which are the sensitive variables that the tool is going to examine in order to ensure that the data is fair? Indeed, to avoid discrimination, sensitive data should not be revealed by the AI system. The first step of the process is to identify these sensitive data as protected attributes. A protected attribute can be defined as a data feature that partitions a population into groups whose outcomes should reflect the reality;
- **Categorization of privileged and unprivileged groups**: which are the groups prone to bias? Once the protected attributes have been clearly identified, they need to be categorized into privileged and unprivileged groups. The aim is to transform the protected attribute values (or numerical feature) into categorical features;
- **Application of fairness metrics**: how can fairness be measured? A metric is a precise quantitative way of measuring the unfairness concept. Based on the previously defined goals select and apply fairness metrics in order to identify bias in the data.

C. Create a Fair Model

During the application of fairness metrics, we might detect that our features below or above the thresholds. In these cases, we should apply techniques in order to mitigate the results. There are three main categories of fairness algorithms that can be applied for data and algorithms mitigation:

- **Pre-processing:** to pre-define an algorithm with the objective of running a pre-defined algorithm, called a classifier, and analyse whether the prediction follows a similar distribution as the labelled data;
- **In-processing:** to make changes during the prediction process to solve the problem, with the objective of using the same dataset and obtaining an unbiased prediction of the migration phenomenon of interest;
- **Post-processing:** to edit posteriously so that the fairness constraints are satisfied. The key idea is to find a proper threshold using the original score function for each group. Firstly, the training model is developed in order to apply the predictions and secondly, the mitigation algorithm is applied.

The following table summarises the different types of algorithms that can be used for bias mitigation.

Table 8.3 Sample overview of algorithms for bias mitigation (non-exhaustive list for illustrative purposes only)

Pre – processing	**Change the statistical properties of the original data so that it produces fair predictions when used in training**			
	Reweighing Algorithm Modifies the weights of different train examples.	***Disparate Impact Remover*** Edits feature values to improve group fairness.	***Optimized Pre-processing*** Modifies Training data features and labels.	***Learning Fair Representation*** Learns fair representation by obfuscating information about protected attributes.
In- processing	**Change the behaviour of algorithms by introducing constraints and adversarial learning.**			
	Adversarial Debasing Uses techniques to maximize accuracy and reduce evidence of protected attributes in predictions.	***Prejudice Remover*** Adds a discrimination-aware regularization term to the learning objective.	***Meta Fair Classifier*** Takes the fairness metric as part of the input and returns a classifier optimized for the metrics.	
Post – processing	**Change the outputs that the algorithms provide so that the predictions are fair.**			
	Reject Option Classification Changes predictions from a classifier to make them fairer.	**Calibrated Equalized Odds** Optimizes over calibrated classifier score outputs that lead to fair output labels.	**Equalized Odds** Modifies the predicted label using an optimization scheme to make predictions fairer.	

D. Application of the model

The last step of the process to create a fair AI model is to apply the model. If bias has been detected and the process previously defined applied, we need to apply fairness metrics in order to confirm that the model has been fairly corrected. Testing the performance of the model can be done by using different practical testing techniques:

- **Targeted Tests**: Check the performance of the model on a specific group;
- **Comprehensive Tests**: Check the performance of the model on all groups, by including sufficient data for each subgroup;
- **Adversarial Testing** - Searches for rare but extreme harms.

There are previously defined methods and tools that help implement these fairness metrics, most of the available as open-source libraries for different programming languages.

8.2.2 Explainable Artificial Intelligence (XAI)

The second dimension of the trustworthiness framework is the application of Explainable Artificial Intelligence (XAI), which supports the transparency and accountability of the AI-Tool.

Results from the use of AI systems are drawn from complex processes and computations, which are rarely interpretable without appropriate technical knowledge. Therefore, it is important to embark explanatory tools to explain how the outputs are obtained in order to increase the reliability of the AI-Tool.

The use of artificial intelligence explainability (XAI) can solve these concerns by enabling users to understand and trust model predictions without tampering with its learning performance. Within the context of the present study, XAI translates to business benefits on irregular migration forecasts as follows:

- Identifying problems in the migration data and its features so that it can improve the model performance. This would translate into an improvement in the decision making processes concerning irregular migration actions;
- Giving a sense of control and safety since the forecasting of irregular migration AI-Tool's operating staff (see section 5.4) knows at every moment what the behaviour is, and therefore safety guidelines can be applied, and alerts can be triggered when necessary;
- Building trust around the irregular migration forecasting model with the stakeholders, who will be able to have a deeper understanding of the reported results;
- Solving compliance issues related to accountability and regulation. (E.g. Adherence to GDPR regulation where 'Right to Explain' is a must-have for a system).

Explainable Artificial Intelligence (XAI) will gain in importance, and focusing on the design of explainable irregular migration forecasting models will require a significant effort from the AI-Tool's operating staff.

Operational tools for model explainability

The operationalization of explainable artificial intelligence can be carried out through several tools for classification and interpretation depending on the type of model that is being assessed for explanation.

Broadly, AI models can usually be categorized into two core groups of algorithms:

- "White box" – the functioning of the algorithms can be easily explained;
- "Black box" – the functioning of the algorithms requires additional explainability techniques to understand how they work.

Table 8.4 White box vs Black box models

	"White box" Modelling – Traditional statistics	"Black box" Modelling – Machine Learning
Characteristics of	• Computationally simple; • Descriptive on the existing data; • Emphasis on introspection, form, casual effects, and process.	• High computational complexity; • Emphasis on speed and quality of forecasting; • High-performance model.
Example algorithms	• Linear/logistic Regression, Decision Tree, K-Nearest Neighbour, Rule based learners, General Additive Models, Bayesian Methods.	• Tree ensembles (Random Forest, XGBoost, CatBoost…), Support Vector Machines, Multi-layer Neural Network, Convolutional NN.

AI interpreters can either be model-specific or model-agnostic. Model-specific interpreters are specifically designed to explain a given algorithm, such as a neural network, and model-agnostic interpreters can be implemented to explain any kind of algorithm.

Finally, the choice of the tools for model explanation will depend on a number of criteria.

- **Local vs global explanations:** if the explanation needs to address specific features or if the explanation needs to happen on the global setting;
- **Model agnostic vs model-specific:** model-specific methods will exploit the knowledge of the architecture of a specific model to explain it. While this can be an advantage, it does not provide flexibility in the choice of framework. On the other hand, model agnostic methods will not exploit any prior knowledge on the implemented model, making them easier to maintain at the expense of a potential loss of precision;
- **Computational cost:** Many of these methods take a substantial amount of time to output the explanation of the model's forecasts.

8.2.3 Functional monitoring

The third dimension of the trustworthiness framework is Functional Monitoring, which proactively detects the changes over time in the distribution of the input variables present in the training, test and production datasets.

Data quality is of the highest importance, and hence the datasets are carefully screened and curated before being put into use for analysis. Within the context of the present study, it is important to highlight that single datasets containing relevant information to forecast irregular migration phenomena are uncommon. In particular, this happens because data has been collected over a long period of time by several actors and in a completely heterogeneous manner. In addition, the constantly changing migration context and the environment has an impact on the distribution of the available data.

As indicated above, the AI-Tool for irregular migration forecasting will use data of different nature as inputs. Therefore, the machine learning models within the tool will need to take into account these data characteristics by applying functional monitoring.

There are different techniques to carry out functional monitoring, most of them surrounding the monitoring of data drift. The most common way to do this is by monitoring the covariate shift, which refers to the change in the distribution of the input variables present in the train and test datasets.

8.2.4 Governance

AI systems should support human autonomy and decision-making; they should act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and at the same time foster fundamental rights and allow for **human agency and oversight**.

Table 8.5 Human agency and oversight

Human agency	Human oversight
AI systems should support individuals in making better, more informed choices in accordance with their goals. AI systems can sometimes be deployed to shape and influence human behaviour through mechanisms that may be difficult to detect, since they may harness sub-conscious processes, including various forms of unfair manipulation, deception, herding and conditioning, all of which may threaten individual autonomy. The overall principle of user autonomy must be central to the system's functionality. Key to this is the right not to be subject to a decision based solely on automated processing when this	Human oversight helps to ensure that an AI system does not undermine human autonomy or causes other adverse effects. Oversight may be achieved through governance mechanisms such as: • HITL: capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable; • HOTL: capability for human intervention during the design cycle of the system and monitoring the system's operation; • HIC: capability to oversee the overall activity of the AI system and the ability to decide when and how to use the system in any particular situation.

Human agency	Human oversight
produces legal effects on users or similarly significantly affects them.	

Given the sensitive nature of the AI-Tool discussed in this study, it is important for human oversight to be a fundamental element taken into account when designing it, and the development of the AI-Tool should be done with the idea of analysts being the ones dealing with the output of the AI-Tool before these are distributed to the end-users.

For this purpose, the definition of a governance function on the AI-Tool for irregular migration forecasting shall carry out the overall management of the AI-Tool's availability, relevance, usability, integrity, and security. Governance practices shall therefore be developed in order ensure:

- **Quality –** ensuring that the AI-Tool is correct, consistent, and its data is also of quality. Data quality is a measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability and whether it's up to date;
- **Availability –** ensuring that the AI-Tool and the data it needs to run is available and easy to consume by the business functions that require it;
- **Usability –** ensuring that the AI-Tool's functionality is clearly documented, with its data clearly structured, documented and labelled;
- **Integrity –** ensuring data retains its essential qualities as accuracy and consistency over its entire lifecycle;
- **Data security** – ensuring data is classified according to its sensitivity, level of confidentiality and defining processes for protecting data from intentional or accidental destruction, modification or disclosure.

Addressing all of these points requires the right combination of people skills and organisation, internal processes and services, and the appropriate technology.

9 RISK ASSESSMENT

9.1 Objective and summary of the risk assessment

Objective

The objective of the risk assessment is to develop a risk assessment framework related to the future implementation of the proposed AI-Tool for irregular migration forecasting. This framework helps identify potential risks related to the technical and operational dimensions of the tool, and to propose initial mitigation strategies to address them.

Summary

The risk assessment provided within the frame of this study relies on desk research and experience from the implementation of large scale IT systems; it identifies potential risks which underlay the implementation of the proposed architecture of the AI-Tool for irregular migration forecasting. The risks presented in this analysis are addressed across the following steps:

- Identification of the risks raised by the implementation and the use of the AI forecasting tool;
- Evaluation of the likelihood of the identified risk (frequency of occurrence: Unlikely, Occasional or Frequent);
- Estimation of the impact of the identified risk (Low, Medium or High).

It is important to notice that this analysis builds upon risk assessment frameworks ISO 31000, TOGAF, and BABOK to identify the risks and propose a mitigation strategy.

9.2 Risk assessment methodology

The methodology presented below is used to establish a list of risks and assess them. The risk analysis presented in this document classifies risks with respect to their impact the potential future implementation and deployment of an artificial intelligence platform for early warning and forecasting of irregular migration towards the EU. These risks are classified across three main types of risks:

- Business risks;
- Organisational risks; and
- Technology risks.

The risks considered mainly address the business and architectural domain in terms of the scope of the project, the stakeholder relationship and available resources, the technology in use, and external factors. Since all phases of the implementation of the future AI-Tool for irregular migration forecasting will be affected by risks, it is useful to carry out a preliminary risk analysis to identify, and classify the risks associated with the implementation of the tool;

and to define an initial mitigation strategy to contain, minimise, and solve adverse effects.

A first analysis of each risk is performed by mapping its likelihood (frequency of occurrence) against the magnitude of the impact in case the risk turns into an event. This mapping provides insight on how important the identified risks are in order to define and prioritize their mitigation measures.

The following guidelines are based on existing risk management best practices. The likelihood (frequency of occurrence) is scored as follows:

- **Frequent**: likely to occur often and/or continuously over the course of a transformation cycle;
- **Occasional**: occurs sporadically;
- **Unlikely**: will probably not occur more than once during the course of a transformation cycle.

The magnitude of cost impacts, associated with each risk, are assessed using the following criteria:

- **High impact**: infers critical financial loss that could result in more than one line of business leading to a loss in productivity and no return on investment;
- **Medium impact**: infers a minor financial loss in a line of business and a reduced return on investment on the IT investment;
- **Low impact**: infers a minimal financial impact on a line of the business' ability to deliver services and/or products.

Combining the two factors helps assess the importance of the risk under evaluation. This analysis helps carry out the impact assessment on the basis of the following pairwise likelihood/impact scores:

- **High Risk (H):** Implying significant failure of the implementation effort resulting in certain goals not being achieved, therefore leading to a project failure with severe consequences;
- **Moderate Risk (M):** Implying noticeable failure of implementation effort threatening the success of certain goals;
- **Low Risk (L):** Implying that certain goals will not be wholly successful.

These cross products of these parameters provide the level of risk (high, medium, or low), therefore indicating the importance of identified risks, given their likelihood and magnitude of cost impacts. These parameters are operationalised in section 9.4 of the present report.:

Table 9.1 Risk likelihood vs magnitude of cost impact

	Likelihood (frequency of occurrence)		
Magnitude of the cost impact	**Frequent**	**Occasional**	**Unlikely**
High	High risk (H)	High risk (H)	Medium risk (M)
Medium	High risk (H)	Medium risk (M)	Low risk (L)
Low	Medium risk (M)	Low risk (L)	Low risk (L)

9.3 Analysis and mitigation of risks underlying the implementation of the AI-Tool for irregular migration forecasting

During the risk identification, an extensive list of potential uncertainties over the project was narrowed down to specific known risks, in order to anticipate events during the implementation and have a mitigation strategy. The risk analysis helps determine the strategy to address the identified risks throughout the project implementation by applying a contingency planning, tracking and evaluation of the risks.

9.3.1 Business risks associated with the implementation

EU Classified information (Risk 1): This risk addresses the transferring EU Classified Information to non-EU actors (i.e. third-country service providers). This risk may be mitigated by defining a specific section in the security strategy plan for the AI-Tool to ensure that non-EU stakeholders (e.g. third-country service providers do not have access to secure communications or classified information.

EU autonomy (Risk 2): This risk addresses potential issues with the strategic autonomy for the EU is affected by poor decision making due to the misuse of results reported by the AI-Tool for irregular migration forecasting. This risk may be mitigated by defining a specific section in the security strategy plan based on the trustworthiness framework in order to ensure the ability of the EU to set priorities and set decisions. This plan should operationalize the necessary backstops to avoid the misuse of results drawn from the AI-Tool.

Information leaks (Risk 3): This risk addresses potential information leaks, including to criminal or terrorist groups, and foreign intelligence. This risk may be mitigated by defining a specific section in the security strategy plan to ensure Data Leak Prevention. Possible solutions are:

- Encryption;
- Endpoint protection;
- Email control content;
- Intelligent firewalls;
- Device controls;

- Assess security permissions;
- Control print;
- Secure back-up;
- Image text analysis;
- Educate users.

Data pollution (Risk 4): This risk addresses potential data is pollution. Depending on the chosen scenario for implementation, this risk may be mitigated by ensuring that the AI-Tools available within the phase of Data Treatment allow for consistent data profiling. This entails the need for a data profiling AI-Tool for continuous monitoring of its quality, a data pipeline to avoid data duplication and a data quality control team.

Cyber risk (Risk 5): This risk addresses potential cyber-attacks or data breaches. This risk may be mitigated by defining a specific section in the security strategy plan with:

- Cybersecurity principles training;
- Antivirus and antispyware software;
- Firewall;
- Regular software updates;
- Backup copies;
- Control physical access on computers and network components;
- Secure wi-fi networks;
- Individual user accounts for each employee;
- Limit employee access to data and information and limit authority to install software;
- Regularly change passwords.

Foreign control (Risk 6): This risk addresses potential issues with service providers who are owned by foreign investors outside the EU. This risk may be mitigated by defining a specific section in the security strategy plan to evaluate if the company is at risk based on its foreign investors.

Prior to the implementation of the AI-Tool, there is a need to define the selection criteria for the assessment of providers at the moment of carrying out the benchmark for the selection of the AI platform components.

Restricted data (Risk 7): This risk addresses potential issues raised by limited access to restricted and/or proprietary data from service providers. This risk may be mitigated by ensuring that:

- Data sources are duly assessed and classified;
- Data Scientists have the correct accreditation to the corresponding data;
- The data to be used is not constraint by any legal or business requirement.

Security standards (Risk 8): This risk addresses potential issues with the security standards not being met. This risk may be mitigated by defining a specific section in the security strategy plan to ensure security standards for

the development and use of the AI-Tool. These standards must be based on a security framework and best practices such as SABSA or OWASP-SAMM:

- SABSA security framework is used to address Risk Management, Information Assurance, Governance, and Continuity Management;
- OWASP-SAMM provides a security framework to formulate and implement a strategy for software security tailored to the specific risks derived from the use of the AI-Tool.

Delays of delivery from third parties (Risk 9): This risk addresses potential issues with external service providers and other third parties who may induce delays in the implementation. This risk may be mitigated by implementing dummy services and testing them to simulating the different external services to be consumed.

Quality of delivery from third parties (Risk 10): This risk addresses potential issues with suppliers and/or subcontractors who may fail to meet the quality of delivery. This risk may be mitigated by anticipating the quality assurance for deliveries and establishing clear KPIs to monitor it.

Social risks (Risk 11): Risks related to political and/or social instability, demonstrations, strikes lead to a deviation in the planning or not provision of the irregular migration forecasting service. This risk may be mitigated by defining a continuity of service plan or project.

Validation risks (Risk 12): These risks address potential validation issues that may take place during the implementation of the AI-Tool for irregular migration forecasting. This risk may be mitigated by ensuring that:

- Decision delays do not imply impacts on the due course of the project (approval decisions, a financial decision, procurement decisions, infrastructure decisions, etc.);
- Decisions are not ambiguous;
- Decisions are complete;
- Decisions are of high quality.

Monitoring implementation risks (Risk 13): These risks address potential implementation and integration issues when setting up and connecting the AI-Tool for irregular migration forecasting to existing large scale systems. These risks may be mitigated by ensuring that:

- The implementation follows its methodology;
- The integration of components follows the procedures;
- The integration with the business processes is successful;
- The integration with systems is successful;
- The integration with the organisation is successful.

Monitoring management risk (Risk 14): Addressing the potential lack of management or control during the implementation. This risk may be mitigated by ensuring that:

- Changes in priorities are accurate, justified, documented;
- The project team has authority to complete work;

- The project team understands the requirements;
- A change control board is created;
- A change management process is implemented;
- A change management system is implemented;
- Change requests are of quality (justified, clear, evidence-based).

Incomplete scope definition (Risk 15): This risk addresses the possibility of an incomplete scope definition, which endangers project completion due to the lack of planning and consideration within the budget of activities missing from the scope. This risk may be mitigated by ensuring that the assumptions about the project specificities are valid and as exhaustive as possible to cover all potential factors and activities that need to be considered within budget.

Low quality of requirements (Risk 16): This risk addresses the insufficient detail of requirements to be elicited. This risk may be mitigated by ensuring that the requirements are unambiguous and complete, and are aligned with the business processes, the strategy, and the existing systems the AI-Tool will integrate.

Inadequate requirements (Risk 17): This risk addresses potential issues with requirements that may not linked to the business needs (i.e. the European Commission's requirements defer from real business needs). This risk may be mitigated by ensuring that requirements are aligned with the European Commission's strategy, IT and organisational policies, business processes, and systems.

Physical access to facilities (Risk 18): This risk addresses the potential lack of access to the hosting facilities, which may create delays in the project. This risk may be mitigated by identifying all the necessary permits and manage their granting to the whole team throughout the project.

Reliability of forecasts (Risk 19): This risk addresses potential poor reliability of the results reported by the AI-Tool. It is important to notice that this risk is not related with the implementation of the AI-Tool but rather to the ability and expertise of the data scientists and migration experts in charge of modelling the migration phenomena of interest. This risk may be mitigated by ensuring that continuous modelling is carried out by a team of experts properly trained in quantitative techniques, and socioeconomic aspects of migration; in addition, continuous external and independent audits can also help mitigate this potential risk.

Inadequate decision making (Risk 20): This risk addresses potential inadequate decisions (under- or over-preparation) taken due to a lack of accuracy of the results reported by the AI-Tool. As in the case of risk 19, this risk is to the ability and expertise of the data scientists and migration experts in charge of modelling the migration phenomena of interest. The proposed mitigation strategy is, therefore, the same as in the case of risk 19 in order to ensure that continuous modelling is carried out by a team of experts properly trained in quantitative techniques, and socioeconomic aspects of migration.

9.3.2 Technology risks associated with the implementation

Technology change (Risk 21): This risk addresses the potential changes that may occur in the underlying technologies chosen for the implementation of the project. This risk may be mitigated by anticipating and be informed of technological changes and new trends in solutions.

Inadequate technologies (Risk 22): This risk addresses the potential issues with the quality of technology components. This risk may be mitigated by ensuring that technology components:

- Meet requirements;
- Are not over-engineered;
- Comply with standards and best practices;
- Are stable, scalable, interoperable, reliable, and maintainable;
- Do not have security vulnerabilities;
- Introduce third party liability.

Open Sources reliability (Risk 23): This risk addresses the potential reliability issues with open source components. This risk may be mitigated by using automated AI-Tools like continuous tracking AI-Tool to review the code.

Poor translation (Risk 24): This risk addresses the potential accuracy issues related to the translation from local languages into English. This risk may be mitigated by ensuring that the data is on reading standard formats to ease the use of the files by the translation AI-Tools to be deployed within the AI-Tool as part of the data treatment.

Merging different data (Risk 25): This risk addresses the potential issues related to merging data from different sources. This risk may be mitigated by ensuring that the merging data is on the same format to ease the merge.

9.3.3 Organisational risks associated with the implementation

Unrealistic expectations (Risk 26): This risk addresses potential unrealistic expectations due to the lack of participation of stakeholders in the definition of the project and its objectives. This risk may be mitigated by ensuring the presence of all interested parties during the implementation project kick-off, and all alignment meetings.

Install a communication strategy to thoroughly disseminate to interested parties all possible changes in the objectives and/or results of the project.

Insufficient stakeholder's sponsorship (Risk 27): This risk addresses the potential absence of high-level sponsorship which may cause a misalignment between the project's objectives and the organisational strategy. This risk may be mitigated by ensuring the involvement and motivation of an internal implementation project sponsor to align the project objectives with the organisational strategy.

Insufficient stakeholder's availability (Risk 28): This risk addresses potential lack of stakeholder availability, which would lead to a deviation in the planning of the implementation. This risk may be mitigated by:

- Preparing a communication plan for internal staff to stress the planned dates and estimated dedication for the implementation project;
- Confirming directly with the person and/or their responsible, feasible availability on the planned dates;
- Explaining the dedication of internal staff assumed for the elaboration of high-level project planning;
- Considering alternative resources or reinforcement in critical activities assigned to part-time internal staff.

Activity coordination (Risk 29): This risk addresses calendar specificities (timetables, public holidays and holiday periods) that affect the coordination of implementation activities. This risk may be mitigated by identifying the differences in calendar present in the project and take them into consideration both in the management and in the execution of the project.

Insufficient or incomplete provider's availability (Risk 30): This risk addresses potential issues of insufficient staff availability and/or its inability to continue operating (i.e. no longer being active and operational); this risk would lead to deviation in the planning of the implementation and operation of the tool. This risk may be mitigated by:

- Anticipating the staffing at a high-level plan of resource utilisation for the project (quantity, profiles and dates of incorporation of the resources);
- Defining the minimum time required to start the project from the moment the award is confirmed;
- Confirming the staff availability forecasts;
- Keeping staff requests for the project up to date.

Skills mismatch (Risk 31): This risk addresses potential skills mismatches between the project needs and the staff's specific knowledge. This risk may be mitigated by:

- Confirming directly with the person and/or their manager who possesses the knowledge and/or skills required by the project;
- Defining the knowledge and/or skills of internal personnel assumed for the elaboration of the organisational model of the project;
- Limiting the effects of learning curves leading to delays and cost overrun.

Loss of project knowledge (Risk 32): This risk addresses potential issues with unwanted rotation of resources may generate the loss of project-specific knowledge, leading to failure to meet the objectives of the implementation project. This risk may be mitigated by:

- Ensuring that a knowledge transfer policy is in place and in accordance with any known gradual rotation of staff within the organisation;
- Distributing and/or replicating project-specific knowledge to different people;
- Documenting and keep project-specific knowledge up to date.

Data Scientist outsourcing (Risk 33): This risk addresses the potential outsourcing of the Data Scientist, which would lead to loss of control on the analysis. This risk may be mitigated by ensuring contractual warranties and incentives with the Data Scientist. Develop a knowledge-sharing policy to regularly ensure knowledge flow transfer.

Legal framework (Risk 34): This risk addresses potential legal & regulatory changes that may affect the implementation project. This risk may be mitigated by ensuring that the AI-Tool does not incur in legal liability due to regulatory changes.

9.4 Evaluation of risks

This section presents the initial evaluation of the risks assessed above. This evaluation has been carried out on the basis of expert knowledge on the implementation of large scale ICT project implementations; this evaluation is provided for indicative and illustrative purposes and does not present an exhaustive list of risks associated with the implementation of the AI-Tool for irregular migration forecasting.

9.4.1 Business risks evaluation

Table 9.2 Business risks evaluation

Reference	Risk	Likelihood	Cost Impact	Risk Assessment
Risk 1	EU Classified information	Unlikely	High	Moderate Risk (M)
Risk 2	EU autonomy	Unlikely	High	Moderate Risk (M)
Risk 3	Information leaks	Unlikely	High	Moderate Risk (M)
Risk 4	Data pollution	Unlikely	High	Moderate Risk (M)
Risk 5	Cyber risk	Unlikely	High	Moderate Risk (M)
Risk 6	Foreign control	Unlikely	High	Moderate Risk (M)
Risk 7	Restricted data	Occasional	High	High Risk (H)
Risk 8	Security standards	Unlikely	High	Moderate Risk (M)
Risk 9	Delays of third parties	Occasional	Medium	Moderate Risk (M)
Risk 10	Quality of third parties	Unlikely	High	Moderate Risk

Reference	Risk	Likelihood	Cost Impact	Risk Assessment
				(M)
Risk 11	Social risks	Unlikely	Medium	Low Risk (L)
Risk 12	Decision making / Validation	Unlikely	High	Moderate Risk (M)
Risk 13	Monitoring implementation	Occasional	Medium	Moderate Risk (M)
Risk 14	Monitoring management	Unlikely	High	Moderate Risk (M)
Risk 15	Incomplete scope definition	Occasional	High	High Risk (H)
Risk 16	Low quality of requirements	Frequent	Medium	High Risk (H)
Risk 17	Inadequate requirements	Unlikely	High	Moderate Risk (M)
Risk 18	Physical access to facilities	Unlikely	High	Moderate Risk (M)
Risk 19	Reliability of forecasts	Occasional	High	High Risk (H)
Risk 20	Inadequate decision making	Occasional	High	High Risk (H)

9.4.2 Technology risks evaluation

Table 9.3 Technology risks evaluation

Reference	Risk	Likelihood	Cost Impact	Risk Assessment
Risk 21	Technology change	Occasional	Medium	Moderate Risk (M)
Risk 22	Inadequate technologies	Unlikely	High	Moderate Risk (M)
Risk 23	Open Sources reliability	Unlikely	High	Moderate Risk (M)
Risk 24	Poor translation	Occasional	High	High Risk (H)
Risk 25	Merging different data	Occasional	High	High Risk (H)

9.4.3 Organisational risks evaluation

Table 9.4 Organisational risks evaluation

Reference	Risk	Likelihood	Cost Impact	Risk Assessment
Risk 26	Unrealistic expectations	Occasional	High	High Risk (H)
Risk 27	Insufficient stakeholder's sponsorship	Unlikely	High	Moderate Risk (M)
Risk 28	Insufficient stakeholder's availability	Occasional	High	High Risk (H)
Risk 29	Activity coordination	Unlikely	Low	Low Risk (L)
Risk 30	Insufficient provider's availability	Occasional	High	High Risk (H)
Risk 31	Skills mismatch	Occasional	High	High Risk (H)
Risk 32	Loss of project knowledge	Unlikely	Low	Low Risk (L)
Risk 33	Data Scientist outsourcing	Unlikely	High	Moderate Risk (M)
Risk 34	Legal framework	Occasional	High	High Risk (H)

10 MAIN OUTCOMES AND CONCLUSION OF THE STUDY

The following chapter presents the main findings, conclusions and recommendations of the present study. The first section provides a comprehensive overview of the main findings of the study, followed by the main outcomes per assessment.

10.1 Current AI forecasting landscape

Some conflict and crisis forecasting systems already exist in the EU or are being developed at prototype level. Some of these were developed by EU JHA agencies and others by EU Member State institutions. Such conflict and crisis forecasting systems can be considered to be useful for irregular migration forecasting to some extent, as theory and experience suggest that conflicts and crises can influence irregular migration patterns. The closest AI-Tool related to the objectives of this study is EASO's Early Warning and Forecasting System, which forecasts the number of asylum applications that Member States can expect by monitoring and forecasting crises in third countries. Similar to the purpose of the AI-Tool covered in this study, the already existing AI-Tools are and were intended to support the operational preparedness of recipients of the systems' outputs.

10.2 Stakeholder challenges

Various challenges and necessary factors to the development of a designated AI-Tool to forecast irregular migration movements were identified on the basis of research into already existing AI-Tools and on the basis of stakeholder interviews. Challenges relating to the use of particular data sources were identified upfront by stakeholders. Reporting cycles appear to differ among EU Member States and EU JHA agencies. This may be down to various reasons but is crucial for expectation management in terms of both the data providers and the receivers of the output of the AI-Tool into which such data is integrated. Additionally, not all Member States report their data at the same points time or within the reporting period itself. This might distort the underlying analysis of the AI-Tool accordingly. Further, especially in cases of data being provided late, data processing and integration has to be done in a more fragmented manner, rather than all at structurally at once.

Furthermore, understanding the strengths and limitations of analyses drawn from complex statistics or machine learning systems might, in some cases, be considered challenging by stakeholders, considering the relative novelty and overall complexity of such systems. Hence, to fully understand the context, input, and underlying analytical steps towards the outputs of such an AI-Tool, training on the interpretation of AI-Tool output for decision-makers will

support the analyses and use they can make of the outputs. Hence, for the AI-Tool to be useful, its immediate users, such as data scientist, and its end-users, such as decision-makers, should have multiple opportunities to provide feedback within the scope of their involvement with the AI-Tool.

Considering that there are only a few tools equivalent to the AI-Tool which the European Commission envisions, the best guidance in terms of data mapping and analysis activities comes from migration theories, research into available data sources, and expert knowledge about possible influences on prediction variables.

10.3 Data source mapping and assessment

The data source assessment of this study collected characteristics and provides analyses on 60 data sources, and offers further intensified research into data 39 of these data sources. The range of data sources covers a wide spectrum of local and global data, historical and real-time data, statistical, administrative, and innovative data, various data formats, etc. For all three forecasting categories as mentioned above, a wide range of potentially relevant, accessible and operationalizable data sources were identified. The AI-Tool could be conceptualised as offering three types of prediction categories, which reflect time-scale requirements, as well as the reality of the main phases of irregular migration movements. In this regard, prediction category A addresses the underlying situations and potential shocks in countries of origin, i.e. the drivers of irregular migration. Prediction category B covers occurrences between the countries of origin from where irregular migrants set off, i.e. the transit phase until including crossing into the EU. As such, this category first assesses shifting flows along routes in a first step, with a view to ultimately predicting irregular border crossings into the EU as output. Category C predicts occurrences within the EU related to irregular migration, namely secondary movements and asylum applications. These categories build on and inform each other, in that category B incorporates the output of category A, while category C incorporates both the output from categories A and B. Considering that there are only a few tools equivalent to the AI-Tool which the European Commission envisions, the best guidance in terms of data mapping and analysis activities comes from migration theories, research into available data sources, and expert knowledge about possible influences on prediction variables.

In terms of incorporating particular data sources into the envisioned AI-Tool, it is not possible to quantify the improved accuracy that can be expected from adding new data sources to the AI-Tool prior to building it. Instead, only intuitive estimates can be provided in this regard, as well as estimations on the costs and complexity of adding a data source. Data sources that are already being collected for other purposes may appear cheaper at the current point in time, but are less likely to be optimised for the forecasting purpose and are also more vulnerable to uncertainty and change in future.

Also, no significant legal obstacles were identified regarding the use of most assessed data sources, so long as the respective terms and conditions are complied with. Limitations might potentially arise regarding the terms of service for some data sources. Further, the on-going spread of widespread disinformation and fake news might influence the accuracy of the AI-Tool's outputs. Personal bias might also influence the forecasts.

10.4 Legislative framework of the AI-Tool and fundamental rights considerations

The main findings of the legislative assessment are as follows:

- The EU primary legislation does provide a legal basis for the AI-Tool in question, in particular, Art. 77 of the TFEU;
- The development and operation of an AI forecasting tool by any of the EU JHA agencies should be aligned with all pieces of EU primary legislation. This includes the TEU, TFEU and the EU Charter of Fundamental Rights and the values these documents promote;
- The following EU secondary legislation needs to be duly considered (Dublin Regulation, Schengen Borders Code, EURODAC Regulation, VIS Regulation, EUROSUR Regulation now recasted in the new EBCG Regulation, EBCG Regulation, Qualification Directive) in the design of the AI-Tool's functionalities and algorithms;
- Some amendments to the establishing Regulations of some of the Agencies might be necessary in order to respond to the needs of the AI-Tool. Particularly, clarifying the relevant powers and the responsibilities of the concerned Agency in order to host the AI-Tool, clarifying the purposes of the AI-Tool, how it will operate in practice, the effective population with data of the AI-Tool and the access to the end-results from it, including safeguards for the alignment of its algorithms with the fundamental human rights, for its security and its proper and effective functioning, control, monitoring and upgrade mechanisms, etc.;
- Most of the analysed data sources are deemed as feasible to be used from a legal point of view, based on their particular Terms and Conditions. In accordance with some of them, the usage of available data is regulated by a formal agreement.

Recommendations

When weighing up which agency or body will host the AI-Tool it should be considered that the most reliable way to ensure its adequate and effective functioning, control and data supply, would be through explicit legal provisions (e.g. at the level of either a regulation or directive). Thus, the legislative framework should encompass the main objectives of the AI-Tool itself, the envisaged control and monitoring mechanisms, guaranteeing its secure and compliant operation. Additionally, the proper functioning and feeding of the algorithms should be regulated through a legal framework. So, amendments to the secondary legislation will be necessary in order to allow the effective exchange of information between the relevant EU JHA agencies in terms of the discussed AI tool.

Human oversight is to be guaranteed, and no fully automated decision-making will take place. Developers of algorithms should exclude parameters that are proxies for known biases.

The AI-Tool should support appropriate accountability mechanisms that allow for actions traceability, which means, to maintain/keep track of access activities, chronology and records of the drawn conclusions and/or decisions and the information used for the purpose. Additionally, the AI-Tool should provide for follow-up control and assessment, whether further refinement of the algorithms is necessary. To achieve that an explicit legal provision should be present;

The AI-Tool should provide different access levels depending on the role and clearance of the variety of potential users. The access level and its relevant aspects (who, to what information) are matters which should be duly considered under the respective legal provisions. With regard to that, a differentiation between the hosting Agency (having full access to the raw data) and the ones, which will only have access to the final results/forecasts of the tool should be made as this will affect the accessibility of information.

As a next step Working arrangements establishing the cooperation and information exchange between the relevant EU JHA agencies should be far more specific and definite in terms of the type of information shared and its access level in order to facilitate the effective operation of the concerned AI-Tool due to the fact mentioned above (section 6.4) that most of the existing WA do not explicitly arrange the exchange of information between the respective EU JHA agencies. In addition, a Cooperation Plan in some of the cases might be concluded in order to refine the scope of cooperation and exchange of information. While in other cases, a novel more specific Working arrangement might be necessary. It should be noted that the conclusion of Working arrangements between the respective EU Agencies should be stemming from secondary legislation, namely the Regulation establishing the functioning of a particular agency and its relationship with other EU bodies in order to have a solid legal ground. Additionally, the above-mentioned Cooperation Plan is a follow-up action rather than a possibility on its own. This means that in the case when a current WA is sufficient to some extent, only a Cooperation Plan will be necessary to narrow down and specify the rules and conditions in terms of exchange of information regarding the AI tool.

Control/monitoring/oversight to ensure/guarantee the adequate and proportionate functioning of the tool. This is in accordance with 'Ethics guidelines for trustworthy AI' from 2019.[73] For example, control must be exercised over the quality of the results produced by the Tool, whether the algorithms are accurate and whether the analysis performed on the basis of input data is reasonable, enough correct and useful.

[73] Available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

The AI-Tool design should be fair as to fully comply with the fundamental legal principles of the EU, which means that specific measures should be incorporated at the design phase to ensure such compliance. For instance, principles of fairness, non-discrimination and others should be integrated in such a way that the tool would take them into account during the forecasts' process.

To sum up, all the above-mentioned recommendations from a legal point of view are complementary to each other rather than separate options and prerequisites on their own regarding the development, hosting and functioning of the AI tool.

10.5 Location and hosting of the AI-Tool

At present, there is not a single clear candidate for hosting the AI-Tool. Instead, there are who each bring different advantages to the table, but who would also need additional support to meet all of the operational, analysis, and dissemination requirements. While some stakeholders have the operational capacity (although not currently the legal base) to build and host the AI-Tool, they do not have the capacity to analyse these data, while other candidates currently have the analytical but not the operational capacity.

This raises the question of whether it is preferable to invest in the capacity of a single host, or whether instead, the EU should consider a joint approach where the operational and analytical tasks are shared among multiple agencies. In the first scenario, the decision about where to host the AI-Tool would hinge on the EU's ability to invest in building a host's organisational, analytical, and dissemination capacity (and whether there is more scope to allocate funding for some of these activities compared to others). In the latter scenario, the decision would centre on the extent to which the EU could invest in coordination mechanisms, such as additional working agreements, to facilitate data access and data sharing.

The decision about where to host the AI-Tool hinges on three sets of tasks: operations, analysis, and dissemination. Currently, none of the candidates assessed would meet all of these criteria alone without additional investment. For example, while some stakeholders have the operational capacity (although not currently the legal base) to build and host the AI-Tool, they do not have the capacity to analyse these data, while other candidates currently have the analytical but not the operational capacity.

When selecting a host, the EU will need to decide about whether they wish to invest in building the operational or analytical capacity of a single host, or if they would prefer to develop a joint approach where the operational and analytical tasks are shared among multiple agencies. The latter approach will require further investment in appropriate coordination mechanisms (such as additional working agreements) to facilitate data access and data sharing.

Options for hosting the AI-Tool may be further narrowed down by weighing the eventual format of the AI-Tool (which may lend itself to one host or another) and the incentives of potential candidates to cooperate on the AI-Tool's operation, analysis and dissemination. In both scenarios, it is key that appropriate measures are put in place to analyse and filter the AI-Tool's outputs before they are circulated to a wider audience, both to avoid data misinterpretation and to better tailor information to different audiences depending on the type(s) of decision-making the AI-Tools aims to support (e.g. policy, operations). Ultimately, these considerations are also subject to the question of who will own the outputs of the AI-Tool, since ownership may influence the willingness of stakeholders to, for example, invest resources (e.g. budget, staff capacity) to support its design and operation.

Recommendations

To ensure that operational and analytical capacities to run the AI-Tool can be met, the European Commission should either invest in strengthening the operational/analytical capacity of the host or develop a joint approach where the operational and analytical tasks are shared among multiple agencies, likely requiring investment in appropriate coordination mechanisms.

The dissemination of the AI-Tool's outputs should be adapted to the needs and preferences of different audiences. For example, a policy unit would benefit from the input that is accompanied by analysis, which may take longer to produce but is more useful for policy programming purposes. Operational units would instead benefit from more action-oriented and timely input, such as concrete estimates on the arrival of irregular migrants at the EU borders;

Depending on the level of confidentiality of the data sources used in the AI-Tool, EU JHA agencies, institutions, and bodies may have different levels of access to the AI-Tool. These different access levels should already be considered in the design of the AI-Tool and in the coordination, mechanisms used to relay information.

10.6 Technical considerations

The operational assessment of this feasibility study recognises that within the current context of irregular migration and the associated response with EU early warning systems create an incentive to develop a performant AI-Tool for irregular migration forecasting. The operational analysis and proposed high-level architecture design provide a solution that responds to the business needs the European Commission is subjected to.

Furthermore, the trustworthiness assessment provides a framework comprising the techniques and tools that can help address the following needs during the design and operation of the AI-Tool for irregular migration forecasting:

- Human agency and oversight;
- Technical robustness and safety;
- Transparency; and

- Accountability.

In order to address these criteria, the trustworthiness assessment develops the framework by relying on four key dimensions: Fairness, Explainable Artificial Intelligence (XAI), Functional Monitoring, and the transversal dimension of Governance.

Finally, the operational risks associated with the future implementation of the tool have been identified and described in the risk assessment; these risks have been categorized in business, technology and organisational risks, and a mitigation strategy has been proposed.

10.7 Conclusion of the study

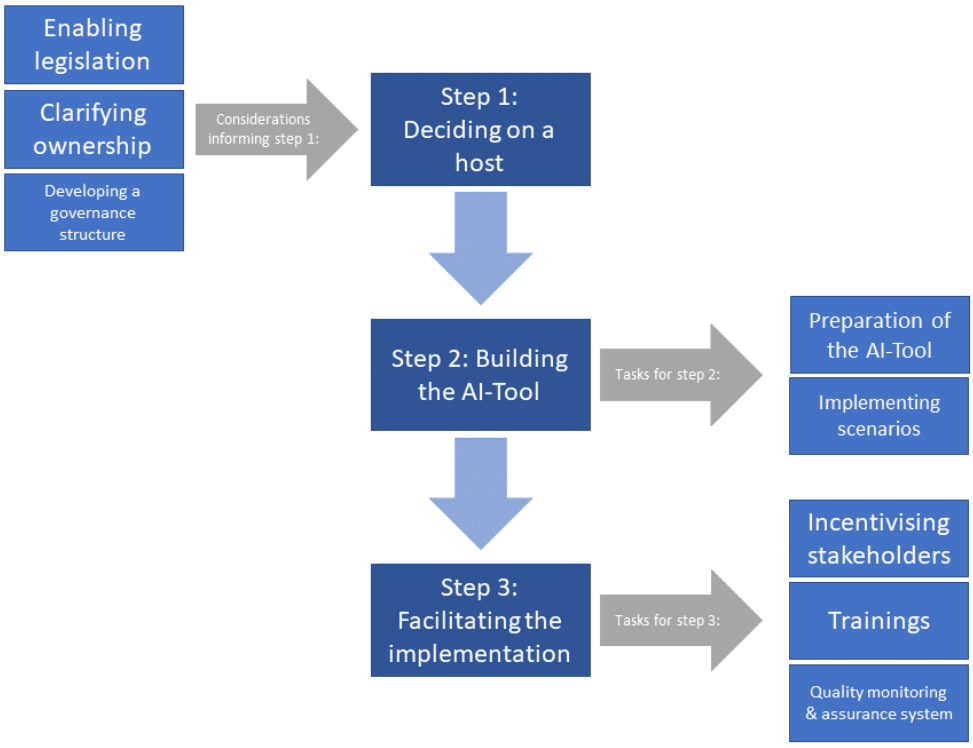
Overall, based on the various feasibility assessments which consisted of an extensive ex-ante data source assessment, desk research, stakeholder interviews as well as the findings of the Closed-doors seminar, it can be concluded that a well-performing forecasting system can be built. However, it will only be possible to precisely assess its reliability ex-post after the implementation of the AI-Tool. This study submits evidence-based recommendations to the European Commission on such steps that should be considered as per each assessed area, and ultimately on how the AI-Tool can be implemented most efficiently and respecting such standards as the EU's Ethics Guidelines on Trustworthy AI.

Table 10.1 Overview of the requirements for feasibility

11 NEXT STEPS

The following section briefly outlines the various next steps that need to be taken. Figure 11.1 below indicates whether these steps can be taken simultaneously, and which steps might be required to be taken before the commencement of others. It should be noted that while the figure indicates that the building of the AI-Tool should follow the decision on the host, these processes can effectively be done in parallel. However, the arrangement of one step following the other was based on the consideration that the eventual host might already have a specific IT architecture in place, which would have to be adapted by the AI-Tool to ensure full functionality. Hence, building an AI-Tool in parallel while a host is chosen would likely result in changes having to be made to it at a later stage to integrate it into the host's environment.

Figure 11.1 Next steps



Step 1: Identification of a host

A first necessary step towards operationalisation is the identification of a host of the AI-Tool. The decision on the host depends on whether the EU has a preference for building the capacity of one EU JHA agency or body to host the AI-Tool, or for pursuing a joint approach where the operational, analytical, and dissemination responsibilities are shared by multiple actors. To inform this decision, we recommend assessing whether it is feasible or desirable to have one stakeholder operate the AI-Tool and another analyse and

disseminate its outputs, and what cooperation mechanisms would need to be in place. Further, it needs to be considered how much scope exists for investing in additional operational or analytical capacity.

Considerations informing step 1:

A set of important considerations need to be addressed in the process of making a decision on a host. These are as follows:

- **Enabling legislation:** As outlined in the assessments, there may be a need for enabling legislation and almost certainly a formal definition of the respective roles and responsibilities of EU JHA agencies in relation to the AI-Tool. This framework should define access restrictions to the AI-Tool and information it produces, knowledge dissemination settings, hosting and management rights and coordination mechanisms. This consideration can also encompass looking into the assessments of the mandates of individual EU JHA agencies with a view to establishing the extent to which changes would have to be made depending on the preferred hosting structure.
- **Clarifying ownership:** Also, the question of ownership of the AI-Tool needs to be clarified since this underlines the point about the governance structure of the AI-Tool. Specifically, this means deciding on who will own the outputs of the AI-Tool, and then reflecting on how this ownership may influence the willingness of stakeholders to, for example, invest resources (e.g. budget, staff capacity) to help design and operate the AI-Tool.
- **Development of the governance structure:** Another important next step consists in the development of the governance structure of the AI-Tool. Synergies between different stakeholders are already in place to some extent, such as some information-exchanges and joint risk analyses, but the frameworks and legal bases for governing an AI-Tool are still missing. For example, additional working arrangements in line with individual mandates are most likely needed to facilitate cooperation on the AI-Tool. It should also be decided on whether the AI-Tool should have a central coordination information point, for example within the European Commission, where the outputs of the AI-Tool can be relayed to different EU JHA agencies, institutions and Member States depending on their needs and preferences.

Step 2: Building the AI-Tool

As indicated above, the building of the AI-Tool can theoretically be already commenced while step 1 is on-going. Nonetheless, it is recommended to wait until a host is decided on, to gain insights on technical specifications on the respective architecture used. In building the AI-Tool, the model itself needs to be prepared. The AI-Tool can be conceptualised as offering the three previously outlined types of forecasting categories (A, B, and C), which reflect time-scale requirements, as well as the reality of the main phases of irregular migration movements. In this regard, forecasting category A addresses the underlying situations and potential shocks in countries of origin, i.e. the drivers of irregular migration. Forecasting category B covers occurrences between the countries of origin from where irregular migrants set off, i.e. the transit phase until including crossing into the EU. As such, this category first

assesses shifting flows along routes in a first step, with a view to ultimately predicting irregular border crossings into the EU as output. Forecasting category C covers occurrences within the EU related to irregular migration, namely secondary movements and asylum applications.

Tasks for step 2:

For step 2, the below presented set of tasks is crucial in building the AI-Tool for its envisioned purpose.

- **Preparation of the AI-Tool's data model:** For the AI-Tool to offer the three previously presented types of forecasting categories, preparatory activities have to be conducted to this end, such as obtaining training data for the AI-Tool, and beginning to choose the appropriate architecture and building blocks.
- **Implementing scenarios:** From a technical perspective, the next steps include the implementation of the scenarios proposed under the operational assessment as an incremental process based on the business needs (proof of concept, additional use cases). The proposed architecture design of the AI-Tool outlines which functionalities are required to facilitate the operationalisation a solution that responds to the needs expressed by the European Commission. Relevant steps to follow to ensure such implementation of the solution include:
  - The elicitation and management of requirements;
  - The declination of the proposed architecture in Application Building Blocks (ABBs);
  - The declination of the ABBs into Solution Building Blocks (SBBs);
  - The conduction of a benchmark analysis to select the cost-effective SBBs;
  - The validation of the SBBs by the stakeholders;
  - The development of the Low-level Design;
  - The preparation of the Implementation Roadmap;
  - The preparation of procurement process to initiate the acquisition of licenses and necessary hardware.

Step 3: Facilitate the implementation

Taking decisions on the host and building and providing the AI-Tool to end-users without any additional steps to facilitate the implementation are unlikely to ensure that the AI-Tool will be used to its full potential. To best facilitate the implementation, relevant stakeholders need to be informed on various elements relating to the AI-Tool, including on the data on which the outputs are based, on how outputs are generated in the first place and ultimately how trustworthy they are. This has significant implications on the extent to which the outputs of the AI-Tool will be used and thus can eventually offer added value to EU processes. As such, stakeholders not only need to be informed about the AI-Tool's specificities but also trained on understanding and engaging with its output. This will assist in incentivising full use of the AI-Tool. A similarly important task is to ensure an appropriate quality monitoring and assurance system is in place, which allows for ensuring the AI-Tool continues to provide valuable output. Creating the latter can already be commenced at earlier stages, i.e. while the AI-Tool is built; however, the

existence of a functioning quality control system is crucial in the facilitation stage.

Tasks for step 3:

For step 3, the below-presented set of tasks will prove relevant in ensuring the AI-Tool is embraced by stakeholders in accordance with its purpose, and can thus be utilised to its full potential.

- **Identify incentives for stakeholders**: Furthermore, there is a need for identifying the incentives for stakeholders to use the AI-Tool and cooperate on its operation, analysis and dissemination. Relevant policy and operational units should indicate their willingness and ability to use the AI-Tool's outputs for their programming depending on where they see it to be most useful. This may require the European Commission to invest in political capital and expectation management (e.g. to inform on whether such an AI-Tool can be deemed trustworthy enough to incentivise its use in daily policymaking).
- **Provide trainings:** Next to this, expert knowledge and training of the AI-Tool are required to build the capacity of end-users if they are not analysts. This is related to interpreting the forecasts, including their caveats and limitations, understanding the quality assurance processes (including the caveats and limitations of the data fed to the tool), and how the forecasts are produced, as well as how to translate into actions, measures, policy and decision-making. It is recommended to start early in building the capacity of immediate and end-users. There will likely be three categories of user: technical operators who need to develop and maintain data sources and the actual system itself; analysts who interpret the system's outputs; and decision-makers who need to understand the strengths and limitations of analysis incorporating the AI-Tool's forecasts.
- **Develop a quality monitoring and assurance system** to process the data coming in and out of the AI-Tool in order to produce as accurate as possible forecasts.

Annex Bibliography

Acostamadiedo et al, 2020. *Assessing Immigration Scenarios for the European Union in 2030: Relevant, Realistic and Reliable?* IOM. Available at: https://publications.iom.int/system/files/pdf/assessing-immigration-scenarios-eu.pdf.

Acostamadiedo, E., 2020. *Forecasting the future of migration – many approaches, one commonality: uncertainty.* Available at: https://migrationdataportal.org/de/blog/forecasting-future-migration-many-approaches-one-commonality-uncertainty.

Aequitas, 2018. Center for Data Science and Public Policy. Available at: http://www.datasciencepublicpolicy.org/projects/aequitas.

Bellamy, R. et al, 2018. *AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias*. Available at: https://arxiv.org/abs/1810.01943.

Bijak, J. et al, 2020. 'Assessing Time Series Models For Forecasting International Migration: Lessons From The United Kingdom'. *Journal Of Forecasting* 38, no. 5: 470-487.

Bijak, Jakub, and Mathias Czaika, 2020. *Assessing Uncertain Migration Futures – A Typology Of The Unknown*. Deliverable D1.1: Quantmig Project. Southampton: Krems: University of Southampton, Danube University Krems.

Bijak, J. et al, 2017. 'Modelling in Demography: From Statistics to Simulations', in: Silverman, E. (ed), *Methodological Investigations in Agent-Based Modelling.* Springer, pp 167-187.

Bijak, J., et al, 2015. Probabilistic Population Forecasts for Informed Decision Making. *Journal of Official Statistics*, 31(4), pp.537-544.

Bijak, J., Wiśniowski, A., 2010. 'Bayesian forecasting of immigration to selected European countries by using expert knowledge', *Journal of the Royal Statistical Society.* Available at: https://www.knomad.org/sites/default/files/2017-05/m1_Apr292014_Joint_KNOMAD_UNPD_Seminar_Bijak_and__0.pdf.

Brian, H. Z., Blake, L., and Margaret, M., 2018. *Mitigating Unwanted Biases with Adversarial Learning*, arXiv:1801.07593.

Buettner and Muenz, 2016. *Comparative Analysis of International Migration in Population Projections*. KNOMAD. Available at: https://www.knomad.org/sites/default/files/2017-04/KNOMAD%20WP%2010%20Comparative%20Analysis%20of%20International%20Migration%20in%20Population%20Projections.pdf.

Carammia, M. and Dumont, J., 2018. *Can we anticipate future migration flows*? OECD/EASO Migration Policy Debate N.16. Available at: https://www.oecd.org/els/mig/migration-policy-debate-16.pdf.

Chadefaux, T., 2017. Conflict forecasting and its limits. *Data Science*, 1(2), pp. 7-17. DOI: 10.3233/DS-170002.

Charter of Fundamental Rights of the European Union.

Collett, E. and Le Coz, C., 2018. *After The Storm: Learning From The EU Response To The Migration Crisis*. migrationpolicy.org. Available at: https://www.migrationpolicy.org/research/after-storm-eu-response-migration-crisis.

Crisis Group, 2019. *Better Early than Sorry: How the EU Can Use its Early Warning Capacities to their Full Potential.* Available at: https://www.crisisgroup.org/europe-central-asia/better-early-sorry-how-eu-can-use-its-early-warning-capacities-their-full-potential.

Disney et al, 2015. *Evaluation of existing migration forecasting methods and models. Report for the Migration Advisory Committee.* CPC. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467405/Migration_Forecasting_report.pdf.

DTM, IOM, 2018. *Mixed migration flows in the Mediterranean.* Available at: https://migration.iom.int/docs/Flows_Compilation_Report_June_2018.pdf.

EASO, 2017. *Quantitative assessment of asylum-related migration. A survey of methodology.* Available at: https://easo.europa.eu/sites/default/files/publications/Quantitative_assessment_of_asylum_related_migration_1.pdf.

EASO, 2019. *The future of international protection in the EU+ by 2030: A Scenario study.* EASO/ DOP/IAU/RSS/2019/229.

EASO, Frontex. *Working Arrangement between the European Asylum Support Office (EASO) and the European Agency for the Management of Operational Cooperation at the External Borders of the Member States of the European Union (Frontex).* Available at: https://www.easo.europa.eu/sites/default/files/public/20120926-FRONTEX-EASO-working-arrangement FINAL_0.pdf.

EEAS, 2014. *Factsheet – EU Conflict Early Warning System. Available* at: http://www.eeas.europa.eu/archives/docs/cfsp/conflict_prevention/docs/201409_factsheet_conflict_earth_warning_en.pdf.

ESA. *Migration Radar 2.0, Big data for migration, Big data applications to boost preparedness and response to migration – Feasibility Study.* Available at: https://business.esa.int/projects/migration-radar-20.

Eu-LISA, 2020. *Eu-LISA and EASO Sign a Three-Year Cooperation Plan.* Available at: https://www.eulisa.europa.eu/Newsroom/PressRelease/Pages/eu-LISA-and-EASO-Sign-a-Three-Year-Cooperation-Plan.aspx.

Eu-LISA, Europol. *Memorandum of Understanding between the European Police Office (Europol) and the European Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice eu-LISA).* Available at:

https://www.eulisa.europa.eu/PartnersStakeholders/Documents/MoU_Europol_eu-LISA.pdf.

Eu-LISA, Frontex. *Working Arrangement between the European Agency for the Management of Operational Cooperation at the External Borders of the Member States of the European Union and the European Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice.* Available at: https://www.eulisa.europa.eu/PartnersStakeholders/Documents/WA_Frontex_eu-LISA.pdf.

European Commission, 2016. *Proposal for a Regulation of the European Parliament and of the Council establishing a common procedure for international protection in the Union and repealing Directive 2013/32/EU.* COM(2016) 467 final, 2016/0024 (COD). Available at: https://ec.europa.eu/transparency/regdoc/rep/1/2016/EN/1-2016-467-EN-F1-1.PDF.

European Commission, 2018. *Many more to come? Migration from and within Africa*. Available at: https://ec.europa.eu/jrc/sites/jrcsh/files/africa_policy_report_2018_final.pdf.

European Commission, 2018. *ICT impact assessment guidelines - Practical tools and guidelines for assessing ICT implications.* Available at: https://ec.europa.eu/isa2/sites/isa/files/ict_impact_assessment_guidelines.pdf.

European Commission, 2019. *2019 Annual update of the remuneration and pensions of the officials and other servants of the European Union and the correction coefficients applied thereto*. Available at: https://op.europa.eu/en/publication-detail/-/publication/96d1687f-1d95-11ea-95ab-01aa75ed71a1/language-en/format-HTML/source-120860403.

European Commission, 2019. *Ethics guidelines for trustworthy AI*. Available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

European Commission, 2019. *IOM Displacement Tracking Matrix (DTM).* Available at: https://ec.europa.eu/knowledge4policy/dataset/ds00047_en.

European Commission, 2019. *Detections of illegal border-crossings; monthly statistics.* Available at: https://ec.europa.eu/knowledge4policy/dataset/ds00032_en

European Commission, 2020. *Amended proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the establishment of 'Eurodac' for the comparison of biometric data for the effective application of Regulation (EU) XXX/XXX [Regulation on Asylum and Migration Management] and of Regulation (EU) XXX/XXX [Resettlement Regulation], for identifying an illegally staying third-country national or stateless person and on requests for the comparison with Eurodac data by Member States' law enforcement authorities and Europol for law enforcement purposes and amending Regulations (EU) 2018/1240 and (EU) 2019/818*. COM(2020) 614 final, 2016/0132 (COD).

Available at: https://ec.europa.eu/info/sites/info/files/proposal-regulation-biometric-data_en.pdf.

European Commission, 2020. *New Pact on Migration and Asylum.* Available at: https://ec.europa.eu/info/sites/info/files/new-pact-on-migration-and-asylum-package_1.pdf.

European Union Agency for Fundamental Rights, 2020. *Border controls and the fundamental rights at external borders-Practical Guidance*. Available at: https://fra.europa.eu/en/publication/2020/border-controls-and-fundamental-rights-external-land-borders.

Europol. *Agreement on Operational Cooperation between Europol and Frontex.* Europol. Available at: https://www.europol.europa.eu/agreements/frontex.

Feldman, M. et al, 2015. *Certifying and Removing Disparate Impact*. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 259-268.

Frontex, 2020. *FRAN Quarterly Quarter 3, July–September 2014*. Frontex. Available at: https://frontex.europa.eu/assets/Publications/Risk_Analysis/FRAN_Q3_2014.pdf.

Frontex. *Data Protection Notice for Frontex Application for Return (FAR).* Available at: https://frontex.europa.eu/assets/Data_Protection/Data_Protection_Notice_Returns.pdf.

Frontex. *Processing personal data for risk analysis (PeDRA) - Privacy statement.* Available at: https://frontex.europa.eu/assets/Data_Protection/Privacy_Statement.pdf.

Gartzke, E., 1999. War is in the error term. *International Organization* 53(3), pp. 567–87.

German-Foreign-Policy.com. *Crisis Prevention.* Available at: https://www.german-foreign-policy.com/en/news/detail/8053/.

GMDAC and KCMD. *Big Data for Migration Alliance (BD4M): Harnessing the potential of new data sources and innovative methodologies for migration. Joint Concept Note.* Available at: https://bluehub.jrc.ec.europa.eu/bigdata4migration/uploads/attachments/cjio8ha96019f99zvr9xq6025-big-data-for-migration-alliance-concept-note.pdf.

Good Judgment. *Super forecasting will change the way you think about the future.* Available at: https://goodjudgment.com/about/.

Ingham, L., 2018. *UN urged to develop global AI platform to predict wars.* Available at: https://www.verdict.co.uk/un-global-ai-platform-predict-war/.

IOM, 2018. *Global Migration Indicators 2018*. Available at: https://publications.iom.int/system/files/pdf/global_migration_indicators_2018.pdf.

Kamiran, F., and Calders, T., 2012. Data pre-processing techniques for classification without discrimination. *Knowledge and Information Systems*, vol. 33, pp. 1-33.

Kamishima, T. et al, 2012. *Fairness-Aware Classifier with Prejudice Remover Regularizer.* ECML PKDD: Machine Learning and Knowledge Discovery in Databases, pp. 35-50.

Kasapoglu, C., Kirdemir, B., 2019. *Artificial Intelligence and the Future of Conflict.* Carnegie Europe. Available at: https://carnegieeurope.eu/2019/11/28/artificial-intelligence-and-future-of-conflict-pub-80421.

Katsiaficas, C., 2016. *Asylum Seeker and Migrant Flows in the Mediterranean Adapt Rapidly to Changing Conditions*. MPI. Available at: https://www.migrationpolicy.org/article/asylum-seeker-and-migrant-flows-mediterranean-adapt-rapidly-changing-conditions

Kuschminder, K., de Bresser, J. & Siegel, M., 2015. *Irregular Migration Routes to Europe and Factors Influencing Migrants' Destination Choices.* Available at: https://emnbelgium.be/sites/default/files/publications/irregular_migration_routes_maastricht_graduate_school_of_governance_june_2015.pdf.

Migration Data Portal, 2020. *Future migration trends.* Available at: https://migrationdataportal.org/themes/future-migration-trends.

Migration Data Portal, 2020. *Forecasting the future of migration—many approaches, one commonality: uncertainty.* Available at: https://migrationdataportal.org/de/blog/forecasting-future-migration-many-approaches-one-commonality-uncertainty.

MPI, 2016. *Asylum Seeker and Migrant Flows in the Mediterranean Adapt Rapidly to Changing Conditions.* Available at: https://www.migrationpolicy.org/article/asylum-seeker-and-migrant-flows-mediterranean-adapt-rapidly-changing-conditions.

Peteranderl, S., 2019. *Wie Flüchtlingsbewegungen mit Social Media vorhersagbar sind*. Der Spiegel. Available at: https://www.spiegel.de/netzwelt/netzpolitik/republica-2019-wie-sich-fluechtlingsbewegungen-mit-social-media-vorhersagen-lassen-a-1266174.html.

Proposal 2020/0279 (COD) for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on asylum and migration management and amending Council Directive (EC) 2003/109 and the proposed Regulation (EU) XXX/XXX [Asylum and Migration Fund], 2020/0279 (COD).

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Union Agency for Asylum and repealing Regulation (EU) No 439/2010, COM/2016/0271 final - 2016/0131 (COD).

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, addressing situations of crisis and force majeure in the field of migration and asylum, 2020/0277 (COD).

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, introducing a screening of third country nationals at the external borders and amending Regulations (EC) No 767/2008, (EU) 2017/2226, (EU) 2018/1240 and (EU) 2019/817, 2020/0278 (COD).

Regulation (EU) 2016/794 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation (Europol) and replacing and repealing Council Decisions 2009/371/JHA, 2009/934/JHA, 2009/935/JHA, 2009/936/JHA and 2009/968/JHA. Available at: http://data.europa.eu/eli/reg/2016/794/oj.

Regulation (EU) 2018/1726 of the European Parliament and of the Council of 14 November 2018 on the European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu LISA), and amending Regulation (EC) No 1987/2006 and Council Decision 2007/533/JHA and repealing Regulation (EU) No 1077/2011, available at: http://data.europa.eu/eli/reg/2018/1726/oj.

Regulation (EU) 2019/1896 of the European Parliament and of the Council of 13 November 2019 on the European Border and Coast Guard and repealing Regulations (EU) No 1052/2013 and (EU) 2016/1624.

REGULATION (EU) No 439/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 May 2010 establishing European Asylum Support Office. Available at: https://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:132:0011:0028:EN:PDF.

REGULATION (EU) No 603/2013 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 June 2013 on the establishment of 'Eurodac' for the comparison of fingerprints for the effective application of Regulation (EU) No 604/2013 establishing the criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third-country national or a stateless person and on requests for the comparison with Eurodac data by Member States' law enforcement authorities and Europol for law enforcement purposes, and amending Regulation (EU) No 1077/2011 establishing a European Agency for the operational management of large-scale IT systems in the area of freedom, security and justice (recast). Available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:180:0001:0030:EN:PDF.

Sardoschau, S. et al, 2020. *Migration and Cultural Change.* Available at: https://ideas.repec.org/p/ces/ceswps/_8547.html.

Schmidt, T., Hooper, K. (forthcoming). *Preparing for Future Migration Trends: Using forecasting and scenario-building for forward-looking policies*. Migration Policy Institute Europe.

Sohst, R.R. et al, 2020. *The Future of Migration to Europe: A Systematic Review of the Literature on Migration Scenarios and Forecasts* . IOM's Global Migration Data Analysis Centre, Berlin and Netherlands Interdisciplinary Demographic Institute. Available at https://publications.iom.int/books/future-migration-europe-systematicreview-literature-migration-scenarios-and-forecasts.

Terms of References Request for Service n° 003.

Treaty on the Functioning of the European Union.

Twitter. *Developers: Tap into what's happening.* Available at: https://developer.twitter.com.

Twitter. *Guidelines for law enforcement.* Available at: https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support.

Vezzoli, S. et al, 2017. *Global migration futures: Exploring the future of international migration with a scenario methodology*. IMI Working Paper Series, 135, 1 – 32. Available at: https://www.migrationinstitute.org/publications/global-migration-futures-exploring-the-future-of-international-migration-with-a-scenario-methodology.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

–by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

–at the following standard number: +32 22999696, or

–by email via: https://europa.eu/european-union/contact_en

DR-02-21-127-EN-N

Publications Office
of the European Union

doi: 10.2837/932912
ISBN 978-92-76-29645-4