OPINION ON A PRIOR CONSULTATION REQUESTED BY EUROPOL ON THE DEVELOPMENT AND USE OF MACHINE LEARNING MODELS FOR OPERATIONAL ANALYSIS
(Case 2021-0130)

1. PROCEEDINGS

On **21 October 2020,** the European Data Protection Supervisor ('EDPS') received by Europol a request for informal consultation [redacted] regarding:
(i) the appropriate legal basis for the development and use of Machine Learning ('ML') models in the context of a specific Joint Investigation Team ('JIT', i.e. a specific cross-border criminal investigation) and Europol's support to JIT countries and;
(ii) the need for a prior consultation under Article 39 of Regulation (EU) 2016/794 ('the Europol Regulation', or 'ER') [2] (Case 2020-0982).

On **23 October 2021**, the EDPS requested additional information/clarifications from Europol as to their assessment regarding the appropriate legal basis for such processing operations and as to the data sets to be used and the algorithms and programs that would be incorporated in the specific project.

On **16 November 2021**, Europol provided further clarifications as to the envisaged processing operations. [redacted]

On **27 November 2020**, the EDPS provided staff level/informal advice as to the second issue of the informal consultation, which centred on whether, the processing operations described in Europol's initial request for consultation and further clarified in the email of 16 November 2020 represent a '*substantial change to the manner of processing'* large amounts of personal data by using new technologies and in particular by developing and relying on ML models for identifying and prioritising decrypted communications and is therefore subject to prior consultation under Article 39 ER. The EDPS deferred the answer on the appropriate legal basis for a later stage as this would require further analysis.

[redacted]

[2] 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, OJ L 135, 24.5.2016, p. 53-114.

[redacted]

On 22 January 2021, Europol shared with the EDPS for information a draft Data Protection Impact Assessment ('DPIA')[redacted], which was at that point being carried out by the data controller with regard to the development and use of a 'Machine Learning toolbox' for operational analysis in the specific operation. The draft DPIA concluded that a prior consultation with the EDPS under Article 39 ER was not necessary.

On **3 February 2021**, the EDPS by a letter addressed to Europol's Data Protection Officer ('DPO') confirmed the informal advice offered regarding the need for a prior consultation and reclassified Europol's communication of 22 January 2021 to a notification opening the Article 39 ER procedure under case number **2021-0130**.

On **10 February 2021**, Europol's DPO submitted to the EDPS a formal notification regarding a new type of processing operation 'Machine Learning Toolbox' under Article 39 ER[redacted] The notification included the final version of the DPIA shared with the EDPS on 22 January 2020 as well as the identification of five additional risks for the rights and freedoms of the data subjects and their respective mitigation measures.

On **11 February 2021**, a meeting took place between the EDPS and Europol's Data Protection Function ('DPF') and operational staff. After a first analysis of the DPIA submitted on 10 February 2020, the aim of the meeting was twofold: (i) to pose additional technical questions with regard to the processing operations under consideration and (ii) to provide further guidance and in particular to explain what kind of information and assessment is expected to be part of a DPIA regarding the development of ML models, ahead of the EDPS' formal opinion on the matter.

On **12 February 2021**, the EDPS addressed in writing the additional questions to Europol.

On **18 February 2021,** Europol provided their answers.

In parallel to the cases mentioned above[redacted], the EDPS opened in **May 2019** an own initiative inquiry regarding Europol's use of production data (including operational data) for data science purposes. [7] With this inquiry the EDPS aimed to gain a better understanding of whether Europol intended to/or already processed operational data for data science purposes (i.e. for the training, testing and validation of ML models) and whether such processing operations were compliant with the Europol Regulation. In the context of this case the EDPS addressed a first set of questions to Europol on 14 May 2019 and received the Agency's answers on 6 August 2019. Europol clarified what they considered as processing of operational data for data science purposes, their will to invest in research into the advanced techniques available (state-of-the-art), their commitment to draft and put in place relevant policies and the categories of staff that had access to operational data for such purposes. Further to this written communication, the topic was discussed during the bi-monthly meeting of 14 October 2019.

[redacted]

7 Case 2019-0264.

The EDPS addressed a follow-up letter to Europol on **3 November 2020** requesting the Agency to provide, by 30 November 2020, more information as to the policies they had put in place regarding the use of production data (including operational data) for data science purposes, the appropriate legal basis for such processing operations, the ongoing or planned projects, the algorithms used/ to be used in the projects and the safeguards put in practice.

On 26 November 2020, Europol requested an extension of the deadline announcing: (i) the finalisation of the ongoing process description on the use of production data for data science; (ii) the development of a new policy on the use of production data for data science, aligned with the process description and referring to a new policy outlining the safeguards that must be in place and (iii) the development of a new document on the technical and organisational measures that need to be in place to ensure the appropriate safeguards, in particular to ensure that data are not processed for any other purposes.

Europol's final reply was submitted on **18 January 2021.** The reply included more general information and lacked details, inter alia, as to the list of ongoing or planned projects (as Europol clarified that data scientists usually work on particular ongoing operational cases, based on the priorities set by the Operations Directorate) and as to the safeguards put in practice in order to limit the risks to the data subjects by the processing of operational data for the training, testing and validation of ML models. Contrary to the information provided on 26 November 2020, Europol clarified that they do not intend to adopt specific policies regarding the processing of operational data for the abovementioned purposes as, in their view, the processing of operational data for the training, testing and validation of ML models does not represent an isolated and distinct processing operation.

According to Article 39 of the Europol Regulation, the EDPS shall issue his opinion within a period of up to two months of receipt of the request for consultation, i.e. by 22 March 2021. At Europol's request, the EDPS has decided to treat this prior consultation with urgency and not suspend this period after requesting Europol to provide additional information. As a result, the EDPS issues his opinion on the basis of the information included in the DPIA submitted on 22 January 2021, further completed on 10 February 2021 and on the answers provided by Europol on 18 February 2021.

2. DESCRIPTION OF THE PROCESSING

2.1. Background

In the context of a specific investigation, a JIT has been established [redacted] to facilitate a direct information exchange between the participants. Europol is formally associated to this JIT.

The investigation pertains to a criminal organization, [redacted]

Based on the data collected so far and analysed by non-automated means, it appears that the data is relevant for the prevention and investigation of international organized crime.

[redacted] The EDPS thus understands that the data were transferred to Europol in accordance with Article 18(3) ER and the specific restrictions contained in the Opening Order [redacted]. As for additional conditions stemming from the JIT Agreement, Europol specifies that JIT members have allowed them to receive, store, process, use and disseminate the data for operational analysis purposes in strict compliance with pre-defined Handling Codes. They have requested Europol's support for the following tasks[11]:

1) Analytical support to the case;
2) Detection and identification of High Value Targets and of high risk organized crime incidents;
3) Development of new investigations against high risk organized crime.

Europol is supporting the evaluation of data with criminal content [redacted], either in the EU or countries considered as criminal hubs outside the EU.

In light of the above, Europol has transferred the datasets provided by the JIT members to the Europol Analysis System (EAS). MS can thus request to the appropriate Analysis Project to access the data for their further use for the purposes of their ongoing investigations. [redacted]

[redacted]

2.2. Machine learning elements

In order to effectively and efficiently process the large amount of data received in this context, machine learning elements are (or will be) implemented (natural language processing, data clustering and other functionalities will assist in focusing on the right subset of the data).

2.2.1. Use of machine learning models in the operational analysis process

According to the information provided by Europol, the models and their use will be part of the analysis process of this data. They will be designed to facilitate the detection, within the datasets submitted to , in the framework of the JIT, of specific types of entities.

The ML models are first used to retrieve information from the dataset that might require immediate reaction or further analysis.

The models will also facilitate the pre-selection, for review and assessment, of communications or users of these illegal communications likely to develop activities of specific interest due to their importance and falling within Europol's mandate . They can for instance be used at a further stage of the investigation . As explained by Europol: *'in a dynamic operational analysis under the umbrella of a specific individual operational activity , it is possible to refer an entity that was retrieved manually or submitted by a concerned MS in SIENA to the machine learning models in order to search for positive matches within the dataset that can reveal the network of individuals involved, the flow of information among suspects and their associates, the command and control mechanisms, etc.*'

As a result, Europol specifies that the models are developed to help the selection of messages that could be of higher relevance and assist in prioritizing the contents to be assessed and identifying the users and the circumstances of high-risk criminal events. The models are tools to locate potentially relevant information within the dataset. [redacted]

2.2.2. Acquisition of machine learning tools and training[20]

According to the information provided by Europol, while some of the ML models are used as-is, i.e. with no additional training on Europol's operational data, other models are trained on Europol's operational data or fine-tuned using Europol's operational data[21]. Europol did not provide information about which datasets were used, their size, the way they were selected, etc.

From the information provided, the EDPS understands that the acquisition and training of ML models follows the subsequent steps:

1) Business requirements from the operational staff were provided to the Data and AI team [redacted] The Data and AI team selected the models based on the business requirements, technical criteria such as performance, size, inference speed, hardware requirements, being peer-reviewed and being available within common machine learning libraries in order to perform fine-tuning.
3) Some of the ML models are used as-is (i.e. with no additional training on Europol operational data). Other models, are trained on Europol's operational data or fine-tuned using Europol's operational data [redacted]
4) In all cases, the models are shared with a dedicated team of business users who provide qualitative feedback as well as manual annotations that are subsequently used to further fine-tune/train the models. This is an iterative process [redacted]
5) The use of ML models will involve systematic human intervention, evaluation and validation on the relevance of the output by Europol expert staff of the Operations Directorate. Human validation will be employed as an inherent step of the process. It will verify that the assessment of the source information corresponds to the search result, so as to ensure that the output of the system is faultless. Therefore no automated decision-making will take place based on the results of using the tools and every result delivered by the machine learning toolbox will be verified by the operational experts [redacted].

[redacted]

3. LEGAL AND TECHNICAL ASSESSMENT

3.1. Need for prior consultation pursuant to Article 39 of the Europol Regulation

Article 39 of the Europol Regulation subjects the following processing operations to prior consultation by the EDPS:

(a) processing of special categories of personal data as referred to in Article 30(2)[25]; or

(b) types of processing, in particular using new technologies, mechanisms or procedures, presenting specific risks for the fundamental rights and freedoms, and in particular the protection of personal data, of data subjects.

Furthermore, according to recital 50 of the Europol Regulation: '*the prior consultation mechanism is an important safeguard for new types of processing operations. This should not apply to specific individual operational activities, such as operational analysis projects, but to the use of new IT systems for the processing of personal data and any substantial changes thereto*.'

In the case under consideration, the processing operations described by Europol (in the initial request for informal consultation, Europol's email of 16 November 2020, the notification to the EDPS regarding a new type of processing operation 'Machine Learning Toolbox' and the replies to technical questions shared on 18 February 2021) represent a '*substantial change to the manner of processing*' large amounts of personal data by using new technologies and in particular by relying on ML models for identifying and prioritising decrypted communications. Moreover, the use of ML models may present specific risks for the data subjects (i.e. the users of the platform under investigation). Such risks (e.g. misidentification of data subjects, misattribution of behaviour to data subjects) have to be clearly identified and mitigated through the Article 39 ER procedure.

In view of the above, the EDPS considers that the training, testing and validation of machine learning models with operational personal data and their further use in the context of a specific operational activity **is subject to prior consultation** in accordance with **Article 39(1)(a) and (b)** of the Europol Regulation.

3.2. Formal compliance with the elements to be provided by Europol in the notification under Art. 39 ER

In accordance with Art. 39(2) ER, the prior consultation should contain at least a general description of the envisaged processing operations, an assessment of the risks to the rights and freedoms of data subjects, the measures envisaged to address those risks, safeguards and

[25] Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, concerning a person's sex life or health, plus genetic data.

security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with the Europol Regulation, taking into account the rights and legitimate interests of the data subjects and other persons concerned.

While the notification provided contains a section on the assessment of the risks and an indication of the measures taken to address those risks, even if, as explained in the following sections of this Opinion this assessment is often incomplete, the notification completely lacks any description of the processing operations. This is problematic as this is the foundation for the rest of the prior consultation process. As outlined in the EDPS accountability on the ground toolkit[26], a systematic description of the process should include the following four elements:

- data flow diagrams of the processes,
- the purpose(s) of the (different parts of the) processes,
- a description of their interactions with other processes and
- a description of the supporting infrastructure.

In the context of the case under consideration, the EDPS received from Europol a general description of the envisaged processing operations which have been described in section 2 of this Opinion. However, the information shared does not allow the EDPS to sufficiently understand the full effects of the new processing operations in detail, from the selection of models, to the use of the operational data, including how all the processes are monitored.

Indicatively and focussing on the training part of the machine learning models:

- The provided documentation does not allow an understanding of the exact steps in the process to achieve a suitable level of accuracy, resistance to bias all the while ensuring the security of all data (operational or not) used.
- The provided documentation does not address if, when and how the machine learning models will be retrained - whether this will constitute new processes or be included in the description of the abovementioned process.
- The provided documentation does not address what happens if incorrect data is used in the learning process; is the model robust enough to handle such situations? Is the model retrained after a certain number of errors in the input are detected?

In view of the above, the EDPS asks Europol **for future prior consultations to provide a data flow diagram for each purpose of the processing.**

As regards **the interaction with other processes**, the EDPS asks Europol for future prior consultations **to specifically describe the scenarios that would trigger the process under consultation and indicate any interactive processes**.

[26] https://edps.europa.eu/data-protection/our-work/publications/guidelines/accountability-ground-provisional-guidance_en

3.3. Scope of the Opinion

The Opinion of the EDPS on this prior consultation concerns the development (i.e. training, testing and validation) of ML models and their further use for the operational analysis of data collected in the context of a specific JIT.

3.4. Lawfulness of the processing operations

3.4.1. Europol's legal basis

Europol intends to use ML models for the operational analysis of the datasets provided by Member States in the context of a specific JIT. While the use of such models in the operational analysis phase (e.g. for the prioritization of the contents of the datasets or for entity extraction) clearly falls within the scope of Article 18(2)(c) ER, the processing of the operational data included in the datasets for the training, testing and validation of the pre-trained ML models raise some concerns as to whether such processing operations can rely on the same legal basis..

Article 2(c) ER defines operational analysis in a broad manner as encompassing all methods and techniques by which information is collected, stored, processed and assessed with the aim of supporting criminal investigations.

According to Article 7 of the Integrated Data Management Concept ('IDMC') Guidelines[redacted], the purpose of operational analysis is to support criminal investigations and criminal intelligence operations through all methods and techniques by which information is collected, stored, processed and assessed. For the purpose of operational analysis personal data is used specifically to determine operational action against (a group of) individuals in relation to one or more criminal offences, which may include the seizure of goods, the arrest of suspects and the deployment of investigative techniques to collect evidence.

The definition of operational analysis thus not only includes the preparation of intelligence products by the analyst but also all the technical preparatory steps necessary to make the intelligence analysis possible. This thus includes digital forensics such as the use of decryption methods and tools and other technical steps necessary to prepare the datasets for the entity extraction process.

It is however not obvious from this definition whether the use of datasets to fine-tune, or, in other words, to further develop, train, test and validate the pre-trained ML models in view of adjusting them to the specificities of the datasets to which they will be applied, can be considered as a technique with the aim of supporting a criminal investigation and in particular the criminal analysis process. One notable difference between the use of pre-trained ML models and 'off-the-shelf' software tools that could impact such an assessment is that the former ones need to be further developed, trained, tested and validated in order to ensure that they deliver the expected results in the specific context in which they will be used.

[redacted]

We however understand that some pre-trained ML models need to be further developed, trained, tested and validated with data extracted from the datasets to be analysed in order to ensure that they perform adequately on datasets with these specific characteristics. In that sense, these operations are meant to ensure that Europol can use the ML model on these specific datasets. These operations are not valid for any other datasets with different characteristics. They would thus amount to a preparatory task in the process of operational analysis.

Given that the further development, training, testing and validation of the pre-trained ML models appears to be a preparatory task to ensure that a given tool (a ML model) is suitable for operational analysis for a given context and for the datasets on which it was trained, **the EDPS is of the view that the application of Article 18(2)(c) ER can be considered as appropriate.**

3.4.2. 'Pre-analysis' phase - compliance with Articles 18(3), 18(5) and Annex II B of the Europol Regulation

The processing of the datasets provided by JIT members for purposes of operational analysis, including the further development, training, testing and validation of the pre-trained ML models, raise compliance issues with Articles 18(3), 18(5) and Annex II B ER. Europol identified this risk [redacted] and provided for mitigation measures almost identical to the ones described in the Action Plan [redacted] provided by Europol to the EDPS on 17 November 2020 as follow up to the EDPS Decision of 17 September 2020 (relating to the EDPS own inquiry on Europol's big data challenge).

In more detail, Europol provides for the following mitigation measures: (i) the contributions accepted but still in the phase of assessment or verification/determination of the Data Subject Categorization ('DSC') will be flagged in Europol's data environment; (ii) during the extraction, data minimisation takes place based on the restrictions in the relevant AP Opening Decision (categories of data subjects, crime area, relevance and in agreement with the data provider (on what is expected/needed); (iii) the extracted data will undergo another review by the analysts of the AP, in order to further reduce the amount of data and to ensure compliance; (iv) access will be limited to duly authorised Europol staff only; (v) the machine learning toolbox will help the organisation to cluster the content of the huge volume of messages and information, and will thus mitigate the risk by supporting the process of data subject categorization.

It is our understanding that the flagging of the data is meant to address the risk that data without a DSC is further processed or integrated into the analysis work and focuses on the steps preceding the core operational analysis which are nevertheless necessary to make the

[redacted]

intelligence analysis possible. According to Europol only duly authorised staff will get access to the raw data [redacted].

In the specific case at hand, the EDPS notes that the raw dataset is treated as a contribution [redacted]. Therefore, the restrictions in its Opening Decision should apply to the dataset, including any restrictions regarding the categories of data subjects. In the DPIA[redacted], Europol mentions that [redacted] is in the process of amending the Opening Decision to include relevant data on victims that it knows are present in the dataset. Indeed, the EDPS has informally been made aware of this upcoming amendment, which would expand the scope of the AP to include certain pieces of victim information[redacted]. However, the EDPS has not yet been informed of any formal adoption of the amended Opening Decision, meaning that the dataset may currently exceed the scope of the AP it is stored in.

To ensure that the datasets provided by JIT members comply with the restrictions contained in the Opening Decision of [redacted], **Europol must formally amend its Opening Decision and notify the EDPS once this has been completed.**

It is also our understanding that flagged data will be used for the training, testing and validation of the pre-trained ML models. If this is the case, and in order for the processing operation to be in line with the EDPS letter of 4 December 2020 (reaction on Europol's Action Plan of 17 November 2020), the flagged data should only be used for this purpose and should not be searchable by other users via USE or the EAS until the DSC has been concluded. In other words, the raw data should never be available to perform the core analysis work before it is extracted, i.e. before there is assurance that the datasets provided comply with the restrictions contained in Annex II B ER and in the Opening Decision [redacted]

In view of the above, the EDPS requires Europol to provide assurance that the datasets transferred to the EAS comply with Annex II B ER and the Opening Order [redacted].

The EDPS further urges **Europol to comply with the guidance provided in his letter of 4 December 2020 for the processing of data that would exceed these restrictions**.

[redacted]

3.5. Assessment of specific data protections aspects

3.5.1. Necessity and proportionality

Compliance with the Europol Regulation also includes compliance with the fundamental data protection principles described in Article 28 ER. The principles of necessity and proportionality are of particular interest when developing ML models.

In more detail, the development of ML models needs to be driven by the proven ability of the model to fulfil a specific and legitimate purpose and not by the availability of the technology. In assessing necessity, Europol should demonstrate that their purposes could not be accomplished in another reasonable way. Therefore, it has to be assessed whether there is a real need for the processing of personal data in order to achieve the purpose; whether the processing effectively addresses this need; and whether the same purpose can reasonably be achieved with other less invasive means.

In the DPIA, Europol explains that the operational analysis tools that are developed to detect the relevant entities in the datasets provided by Member States in the context of a specific JIT and identify high value targets(s) and high risk criminal event(s), will expedite the analysis and will assist in providing, in a timely manner, relevant information to the concerned MS, third parties or Union bodies, in order to use this intelligence for investigating or preventing serious and organized crime (actionable intelligence). According to Europol, the volume of the dataset makes the analysis a labour intensive and lengthy process resulting in a decrease of the operational value of the intelligence in question and raising the risk that relevant inferences are not reached and not shared with the respective operational partners.

Although this information can be considered as the starting point of the necessity assessment (since it justifies that the use of the ML models will assist the Agency to effectively carry out its tasks), the EDPS is of the view that **there are still elements missing for such an assessment** to be carried out thoroughly. In particular, it should be further explained and documented why the use of personal data is necessary for the development of the ML models [redacted] and why the specific pre-trained ML models were selected instead of others in order to justify that the least intrusive solution (from the personal data perspective) was selected. As a second step it should be further explained and documented how the datasets used for these purposes were strictly necessary (see section 3.5.2. below on the data minimisation principle).

When assessing proportionality, the interests of the Agency need to be weighed up against the rights and freedoms of individuals. In relation to ML models, Europol needs to assess in particular any detriment to data subjects that could follow from bias or inaccuracy in the algorithms and the data sets being used.

Therefore, these elements are going to be further analysed in the following sections.

[redacted]

3.5.2. Compliance with the data minimisation principle

The above section outlined the way the necessity and proportionality assessment should be carried out when fine-tuning, or, in other words, when further developing, training, testing and validating the pre-trained ML models in the context of operational data processing at Europol. Another assessment that should be carried out concerns compliance with the data minimisation principle laid down in Article 28(1)(c) of the Europol Regulation. Europol should thus demonstrate that the personal data it processes is adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

[redacted]

At the stage of writing this Opinion, the EDPS has not been informed of the actual volume or characteristics of the personal data that have been used to further fine-tune models. Europol only informs that '*different tasks require different types of data*', indicating which categories of data were necessary for the development of the pre-trained ML model.[37] The EDPS considers that **this information should be clearly tracked throughout the fine-tuning cycle, retained afterwards for (internal and external) auditing purposes and conveyed to the EDPS when so requested.**

As a guideline, when documenting which personal data from the original set are to be used during the fine-tuning stage, the EDPS points to the criteria of the Europol Regulation mentioned earlier: Europol should provide a detailed explanation of why the selected subset of data is adequate, relevant, and limited to the amount necessary for the purpose of fine-tuning the model.

[redacted]

3.5.3. Risks related to bias

As ML models learn from data, where the training data is unbalanced or reflects discrimination, they may produce outputs which have discriminatory effects on people based on their particular characteristics. Therefore, processes should be in place as well as technical and organisational measures to manage possible risks.

In the letter of 18 February 2021 addressed by Europol to the EDPS (including the Agency's answers to the EDPS' additional questions), Europol explains that the training of the ML models with operational data was subject to strict data protection safeguards aimed at addressing the issue of data bias. According to Europol, the fact that the Agency used pre-trained models that were not exclusively trained on law enforcement data is an element that addresses the risks related to bias.

Although this is an important element in addressing such risks, there is no explanation on Europol's side as to the following: (i) regarding pre-trained models that are going to be further trained by using Europol's operational data (e.g. BERT-base language model), how Europol would avoid transferring possible biases included in their own data; (ii) regarding pre-trained models that are going to be used 'as is' (e.g. Insight-face-v3 face similarity model), whether their accuracy rates (in the specific case the 99.8% accuracy on Labelled Faces in the Wild (LFW)) are valid for different categories of data subjects, e.g. for different ethnicities or different ages.

Therefore, the EDPS is of the view that there are still elements missing in order to be able to determine whether a thorough assessment regarding data bias was carried out.

In view of the above, **the EDPS asks Europol to**:

- adopt procedures that would allow Europol to identify and remove, or limit, any bias in the data used to further train the relevant models;
- verify that the training data used do not reflect discrimination and, should this be the case, replace with a different bias-free set of data;
- adopt a process that would allow the regular monitoring of the models regarding biases and their readjustment or retraining.

3.5.4. Risks related to statistical accuracy

As mentioned above, one of the mistakes that can readily occur when dealing with Machine Learning is related to the training data. As a key step in using this type of technology, it is crucial to ensure that the training data reflects the data that ultimately will be processed by the model.

This is due to the fact that if the training data does not statistically reflect the operational data, the model will suffer from sample bias (sometimes called selection bias) i.e. the model will

only provide answers that it 'knows' and thus will only give answers that 'look like' the training data. Those answers will most likely be erroneous.

The distinction between the bias described here, and that outlined in section 3.5.3, is that with regard to the latter, a subset of the population will be 'targeted' by the bias whilst here, erroneous answers will be provided by the model because the training data does not completely represent the operational data.

For example, in a ML model which performs object detection in images, if the training data consists only of images of guns, this is not representative if the model is supposed to also detect common items such as chairs, tables etc. Thus it will only recognise guns which in the worst case scenario will lead to data quality issues or frequent no-hits.

For some ML models, Europol has opted to use the 'pre-trained' model. For those models, it is imperative that Europol determines an approach that ensures that the data used for the training reflect the realities of the environment on which the ultimate model is run.

For ML models that will use operational data for the training, the selection of the training data must be carefully considered and also take into account the data minimisation principle. Building a proper training data set will take time and careful testing and validation.

Another potential issue related to the proper distribution of the training data is linked to ensuring that the training accurately reflects the real world situation where applicable. For example, if the training data relates to the relationship between jobs and gender, it is possible that, due to cultural factors, men are more prevalent in construction work and women in nursing jobs. The reality is that women can have construction work and men can be nurses.

This type of culture bias might lead to a statistically non-representative training data, which would again lead to data quality issues. In our example, when faced with a picture of what the model detects as a man, it might attribute it more readily to a construction job, or when faced with the request to detect nurses, the model might skew the results in favour of picture of women.

Regardless of whether or not Europol decides to use pre-trained models, processes to build or check the training or validation of data sets must be built and documented. This is to ensure that a structured approach is taken to deal with this difficult problem and ultimately to avoid biases linked to a statistically non-representative training data (non-representative of the operational data or real world as required). This is valid for data sets used for the training and validation of all models.

These processes should include making statistical checks on the input and output data (statistical abnormalities in the output data is an indication of problems in the input data). The criteria on which to build these statistics and the thresholds to identify mistakes need to be defined for each model.

In view of the above and in order to address the issues identified**, the EDPS asks Europol to**:

- adopt a process on how Europol ensures that their training data and validation data are statistically sound (i.e. will not lead to the kind of problems described above);
- for the pre-trained models, analyse and determine if the input data of the pre-trained models will accurately reflect the operational data ultimately used;
- adopt processes describing how validation data and test data (for non pre-trained models) are selected, keeping in mind the data minimisation principle and the need to have representative data sets;
- put into practice the substance of the above-mentioned processes, keeping in mind the need for securing any operational data, regardless of its function.

3.5.5. Risks related to errors

Another issue that will occur in the training data is related to errors in the output data. It is fair to assume that not all entries in the training data will be accurate. For example, in the context of facial recognition, a picture of a man might be labelled as a woman.

Although a few mistakes in the tens of thousands of entries in the training data will probably not be significative (i.e. will not significantly affect the end model), an accumulation of errors might very well affect data quality in the output. For example, a batch import of data in the training set might have a systematic error in multiple entries.

Errors may be detected quickly in the training data when, for example, the validation of the model occurs. However, if a model is trained with incorrect data and used in operations, any inference from the model's output will need to be reviewed. This means that a proper tracking of how the model and its output data are used in operation is necessary.

It is worth noting that errors in the model's output, whatever their source, if left unchecked, might be detected very late in the operational processes, which will lead to a rapidly increasing number of propagation errors which ultimately will lead to poor data quality.

Similarly to dealing with statistical accuracy, processes should be defined and documented to ensure that errors are dealt with in all situations. Errors can appear in the input data because of incorrect training or validation data sets (on pre-trained models or models trained with operational data); errors can also appear in the output because the model makes a mistake; errors can also be caused by human interaction with the output data, e.g. if the output data is re-used for training or validation, the error will propagate to the model.

The risk with errors in the output data is that they may propagate in the operational data sets (the different Europol operational databases). Thus, data that is provided by ML models and data that is inferred using ML models should be marked as such so that, in case of errors, the operational staff can follow the propagation and fix the potential issues.

Furthermore, proper monitoring is crucial in these circumstances. The UAS needs to be able to use the above mentioned marking to be able to reconstruct which errors have been corrected. This will add a layer of control that propagation of mistakes have been resolved. Of course, this will only work if the DPF performs adequate audits and is informed when major errors and significant propagation of mistakes are detected.

In view of the above and in order to address the issues identified, **the EDPS asks Europol to**:

- document processes as to how Europol will deal with (detect and correct) errors in the training data and validation data;
- document processes as to how Europol will ensure that in case of output error in the models, the data that is further propagated (and is incorrect) will be dealt with;
- document the adaption of the UAS to this additional processing operation, making sure that with the UAS, it is possible to know what data have been used for training, validation and operational analysis. The UAS should also be capable of tracing the propagation of data coming from the machine learning models in the operational databases;
- implement the above-mentioned points, taking into account the need for security for all processing of operational data.

3.5.6. Risks related to security

Article 32 of the Europol Regulation requires Europol to take all the necessary and appropriate technical and organisational security measures to protect personal data against accidental or unlawful destructions, accidental loss or unauthorized disclosure, alteration and access or any other form of unauthorized processing.

Europol's ML toolbox environment as described in the Europol notification, will constitute a large ecosystem of automated tools, languages and models that will process sensitive categories of personal data. Several functions are employed ranging from data insertion for training purposes, data pre-processing, modelling, model evaluation and integration, visualization, packaging and deployment as well as connection add-ons with other Europol information systems. In such an ML environment, the security requirements can be more challenging both from a technological and human perspective.

In the DPIA Europol described the security environment and the security measures that have and will be implemented to ensure the confidentiality, availability and integrity of the personal data that will be used in the context of the specific investigation and in particular related with the development and use of ML models.

Regarding the training and testing of the ML models, Europol declares that it will use operational production personal data. In that context the security measures applied will provide

the adequate safeguards. To that respect a specific policy is needed to be drafted in which nominated staff shall be allocated clearly defined roles and responsibilities. Adequate access controls and logs will be established providing a fully controlled administration of the testing environment, in accordance with Article 40 of the Europol Regulation.

We note that multiple stakeholder groups are involved in the activities of the specific project, ranging from the traditional software engineers to machine learning engineers, senior operational analysts, information security staff and other experts.

Therefore, an environment such as the above requires "accountability on security", meaning that a security framework is applied and that rules on data management, data governance and risk management are clearly defined. Europol shall be aware that the use of various new tools and frameworks, the involvement of new people (stakeholder groups) and the development of new systems or combinations of systems requires a thorough analysis of the data flows and of the security risks.

Apart from the specific security risks of any of the selected ML models that has to be analysed, tackled and verified internally before its use, any cybersecurity trends will also need to be taken into account. Europol is aware that new ML tools and frameworks may lack in security in the beginning, and that the development of a ML environment involving many systems and interfaces may evoke security failures at the boundaries/interfaces. The use of new technologies in Europol technical environment are deployed with measures that ensure the prevention and early detection of any personal data breach. Europol ensures that in case of a personal data breach the EDPS as well as the competent authorities of the Member States concerned as well as the provider of the personal data are notified according to Article 34 of the Europol Regulation. A security incident response plan will be immediately activated in case of a security incident in the machine learning environment.

Europol ensures that all stakeholders have a complete understanding of security and privacy. Data protection by design and by default, data classification, data protection techniques, secure methods of authentication, privacy principles, are elements that are and have to continuously be well defined in the organization. Europol shall ensure that all stakeholders have, based on their function and role, the appropriate training and knowledge of security and privacy.

It is important that Europol clearly defines ownership and accountability for all stakeholders as data are moving to different owners [redacted] at different stages of each defined workflow and within different tools. A clear description of roles and responsibilities is an element that has not been yet provided by Europol in the DPIA.

Europol needs to perform risk assessment analysis of solutions at both tool/system level and from an end to end perspective. This will ensure that security is 'built in' into the design and that applicable security requirements are met at every point in the end to end system (where valuable data is processed, stored or transmitted). Particular attention should be paid at interfaces between the different systems or tools. Especially in case of interconnection with

other Europol systems such as SIENA, the Europol Information System, the Europol Analysis System or any other system which processes operational information, Europol needs to verify any particular security risks and document the required security measures in place. In addition, tests to verify security and quality controls at either side of those interfaces should be clearly documented and verified. It is recommended that Europol performs a combination of feature security testing and penetration assessments of each selected tool.

Europol has to ensure that good programming practices are followed during implementation of the ML models and, depending on the technologies used, appropriate vulnerabilities are mitigated.

A thorough monitoring and security hygiene of the ML toolbox environment is necessary. Specific procedures to ensure that all software components of the tools and systems are at their latest security patch, periodic access list reviews are implemented, logs are regularly reviewed are necessary. We note that at present Europol does not yet provide audit capabilities in the UAS. It is however crucial that Europol implements the possibility to verify and establish what data have been accessed and by whom at all times. The EDPS urges the implementation of this feature and that equivalent measures are at least provided for until the audit capabilities are established.

In view of the above the EDPS considers that Europol has already taken various security measures for the development and use of ML models and that these measures need to be regularly reviewed and complemented. Additionally, in order to address the issues identified above, **the EDPS asks Europol to**:

- provide for the training and testing environment of ML models as well for their use documentation on the clear description of roles and responsibilities and corresponding attribution to access controls authorizations with adequate management and control;
- have full visibility and documented description of all data flows of ML models at all times of the iterative process of development, testing and use;
- conduct security risk assessment for the ML environment and the tools;
- apply specific security testing's in the boundaries/interfaces of the ML models to ensure the integrity of the data processing;
- speed up the auditability of the ML models and ensure that the relevant system security logs are adequately monitored and reviewed.

3.5.7. Data Retention of training sets of data

Europol provides that the training sets of data, containing all the manually annotated data items will be kept for a period decided by the data controller as the data set is an important element for modifying and improving the algorithm.

In this respect, the **retention of the training data sets shall be justified and adequately recorded by Europol.**

3.5.8. Human intervention

Article 30(4) of the Europol Regulation does not allow for decisions of competent authorities that produce adverse legal effects for the data subjects to be based solely on automated processing of sensitive data. To that end Europol provides for human involvement during the process described under section 2 of this Opinion. Europol notes that the use of ML models will involve systematic human intervention, evaluation and validation by Europol expert staff of the Operations Directorate on the relevance of the output. Human validation will be employed as an inherent step of the process. It will verify that the assessment of the source information corresponds to the search result, so as to ensure that the output of the system is faultless. Therefore, no automatic decision-making will take place based on the results of using the tools and every result delivered by the ML toolbox will be verified by the operational experts.

In case the automated results are assessed as false positive or false negative, the human intervention should provide feedback for the retraining of the ML models. This feedback shall be recorded by Europol.

Europol further clarifies that human intervention is also guaranteed before the use of the ML models when training, testing and validating them. Once the ML models have been selected and adapted by the members of the Data and AI team, their results are manually reviewed by selected members of the Operations Directorate who provide qualitative feedback and manual annotations that can be used for further fine tuning the models to achieve better performance.

In view of the above, the EDPS considers that the DPIA as complemented by the document including Europol's answers to the additional EDPS questions adequately identifies and records the degree of human intervention in the process which appears to be meaningful.

However, it should be further clarified regarding the use of ML models: **(i) at which stage of the process human intervention takes place; (ii) whether and, if so, how the outcome of human intervention is used for retraining the ML models.**

4. CONCLUSIONS

In light of the above, **the EDPS considers** that, with the information provided by Europol, he **is not in place to assess** whether the notified processing operations comply with the provisions of the Europol Regulation.

Nevertheless, in accordance with Art. 39(3) ER, **the EDPS formulates a series of recommendations** that Europol should follow in order to avoid possible breaches of the Europol Regulation:

Formal requirements of the DPIA:

- In future prior consultations, provide a data flow diagram for each purpose of the processing and describe the scenarios that would trigger the process under consultation and indicate any interactive processes.

Necessity and proportionality:

- Explain and document why the use of personal data is necessary for the development of the ML models[redacted] and why the specific pre-trained ML models were selected instead of others in order to justify that the least intrusive solution (from the personal data perspective) was selected.

Data minimisation:

- Provide a detailed explanation as to why the selected subset of data is adequate, relevant, and limited to the amount necessary for the purpose of fine-tuning the model.

Risks relating to bias:

- Adopt procedures that would allow Europol to identify and remove, or limit, any bias in the data used to further train the relevant models;
- Verify that the training data used do not reflect discrimination and should this be the case, replace with a different bias-free set of data ;
- Adopt a process that would allow the regular monitoring of the models regarding biases and their readjustment or retraining.

Risks related to statistical accuracy

- Adopt a process on how Europol ensures that their training data and validation data are statistically sound (i.e. will not lead to the kind of problems described above);
- For the pre-trained models, analyse and determine if the input data of the pre-trained models will accurately reflect the operational data ultimately used;
- Adopt processes describing how validation data and test data (for non pre-trained models) are selected, keeping in mind the data minimisation principle and the need to have representative data sets;
- Put into practice the substance of the above-mentioned processes, keeping in mind the need for securing any operational data, regardless of its function.

Risks related to errors:

- Document processes as to how Europol will deal with (detect and correct) errors in the training data and validation data;
- Document processes as to how Europol will ensure that in case of output error in the models, the data that is further propagated (and is incorrect) will be dealt with;
- Document the adaption of the UAS to this additional processing operation, making sure that with the UAS, it is possible to know what data have been used for training,

[redacted]

validation and operational analysis. The UAS should also be capable of tracing the propagation of data coming from the ML models in the operational databases;
- Implement the above-mentioned points, taking into account the need for security for all processing of operational data.

Risks related to security
- Document a clear description of roles and responsibilities for the development, testing and use of ML models including the corresponding attribution to access controls authorizations with adequate management and control;
- Document all data flows of ML models at all times of the iterative process of development, testing and use;
- Conduct security risk assessment for the ML environment and the tools;
- Where applicable conduct specific security testing's in the boundaries/interfaces of the ML models with other Europol systems to ensure the integrity of the data processing;
- Speed up the auditability of the ML models and ensure that the relevant system security logs are adequately monitored and reviewed.

Data retention of training data sets:
- Justify and record the retention of the training data.

Human intervention:
- Clarify, regarding the use of ML models:
  - at which stage of the process human intervention takes place;
  - whether and, if so, how the outcome of human intervention is used for retraining the ML models.

In any case, the EDPS requests Europol to provide assurance that the processing of the datasets provided by JIT members comply with Art. 18(3), 18(5) and Annex II.B ER and Opening Decision [redacted]. Would this not be possible, Europol must implement the measures detailed in their Action Plan of 17 November 2020 and the additional **guidance provided by the EDPS in his letter of 4 December 2020.**

Done at Brussels, 5 March 2021

Wojciech Rafał WIEWIÓROWSKI