STARLIGHT Data Protection Notice

Information for the processing of personal data in accordance with art. 14 GDPR

The purpose of this data protection notice is to inform data subjects about the processing of their personal data. Considering the technical nature of the module and limitations imposed by the research design (i.e., scale), it is considered that informing those data subjects directly would involve a disproportionate effort. For this reason, this information is made publicly available via the project's website in accordance with art. 14 GDPR and with its potentially applicable derogations (art. 14 (5) (b) GDPR[1]), as an effort of enabling the data subjects to be informed about the data processing and to exercise their rights. This notice refers to the specific module of the STARLIGHT responsible for collection of data from online sources.

Data will be collected from:

i. Public social media posts from X (formerly known as Twitter), the content of which will be associated to imminent attacks against soft targets.

1. The Project

STARLIGHT aims to create a community that brings together Law Enforcement Agencies (LEAs), researchers, industry, and practitioners in the security ecosystem under a coordinated and strategic effort to bring Artificial Intelligence (AI) into operational practices. Its main objective is to empower LEAs with automated, operational, and cyber-resilient investigation, intelligence, surveillance, and control capabilities to tackle traditional and emergent criminal activities, terrorism, cybercrime, and cyber-attacks, within and at the borders of the EU. STARLIGHT also aims at contributing to the establishment of a strong EU AI-based security ecosystem, thus enhancing the EU's strategic autonomy in the field of AI for LEAs. The ecosystem will be built around an AI framework for reliable, accountable, responsible, and transparent LEA AI solutions that will enable the involvement and transfer of knowledge between LEAs, Researchers, Industry, SMEs, and Policy makers. Overall, STARLIGHT will ensure that European LEAs lead the way in AI innovation, autonomy and resilience, addressing the challenges of the present and the future, prioritising the safety and security of Europe for all. STARLIGHT solutions will be validated by six high-priority use case application areas: 1) Counterterrorism, 2) Child Sexual Exploitation, 3) Border & External Security, 4) Cybersecurity & Cybercrime, 5) Addressing Information Overload in Serious Organised Crime, and 6) Protection of Public Spaces. To follow the ever evolving and rapidly changing priorities of LEAs as well as the continuous evolution of technologies such as AI, a co-design and co-creation approach will be adopted.

[1] Paragraph 5 (b) of this Article provides for an exemption if such information proves impossible or would involve a disproportionate effort, for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. In this case, subject to the conditions and safeguards referred to in Article 89(1) GDPR, the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.

2. Data Controller

Data Controller: Centre for Research & Technology – Hellas (CERTH) / Information Technologies Institute (ITI), 6th km Harilaou - Thermi, 57001, Thermi- Thessaloniki, Greece.

Project Coordinator: French Alternative Energies and Atomic Energy Commission (CEA), RUE LEBLANC 25, 75015, PARIS 15, FR.

3. Data Processing

STARLIGHT in general aims to deliver tools to cutting-edge AI/ML and data-driven technologies for: (i) generating high-quality multilingual and multimodal data, (ii) understanding the physical (sensors and data-gathering devices related to robotics and IoT systems) and cyber world (online sources from Surface/Deep Web, Darknets) through the exploitation and provisioning of advanced and resilient AI/ML methods and tools capable of handling large volumes of multimodal data, fusing, correlating, analysing and ultimately generating knowledge and intelligence in an explainable, transparent, and accountable manner, (iii) analysing, predicting, anticipating, detecting and mitigating current and future cyber-threats and attacks to AI LEA systems, including adversarial AI. The purpose of data collection in this project is to extract useful information by citizens observations though social media platforms, around imminent attacks against soft targets. Within this Data Protection Notice you have been given the contact details of the relevant project representatives (see Section 2) should you have any questions regarding the processing of your personal data.

What personal data is being processed?

The processed data stemming from the social media posts, with publicly available accounts and with full respect of the terms and conditions of the relevant social media platform will include:

- Social media account information, including the username, as well as the number of friends, followers, lists, favourites, and statuses (i.e. posts).
- Social media posts including replies, textual and multimedia content uploaded by social media users, together with relevant metadata (i.e., public metrics, language, publication time).

No special categories of personal data (art. 9(1) GDPR) are foreseen to be collected (at least not intentionally), nor data relating to criminal convictions (art. 10 GDPR). All data will be collected in accordance with the licences and terms & conditions of the data provider. All data will be gathered only from public accounts, with the permission defined by the social media platform (i.e. X) and in compliance with the respective terms of use, including the ones referred explicitly to the terms of use on behalf of minors, thus in accordance with user expectation of privacy. All collected data will be pseudonymised, including mentioned users. Data minimisation will also be applied, i.e., only data that are necessary for the purposes of the project will be processed. Further, details are provided in the “What is the purpose of the processing” section.

What is the purpose of the processing?

Collection and further processing of those data will support Law Enforcement Agencies towards the identification of content or activities of interest, namely activities that raise concerns. In particular, social media content will be analysed in the event that suspicious activities come to the Law Enforcement Agencies attention, or following offending behaviour where there is a need to gather further information.

Data security

CERTH in the STARLIGHT project implements appropriate technical and organisational measures to ensure an appropriate level of protection against the risks arising from processing, such as accidental or unlawful destruction, loss, alteration, unauthorised disclosure, or access. All data will be collected in accordance with the licenses and terms & conditions of the data provider. All data will be gathered only from public accounts, with the permission defined by the social media platform and in compliance with the respective terms of use, thus in accordance with user expectation of privacy. In accordance with the data minimisation principle, only the parts of the social media posts that are deemed necessary for the project's objectives will be processed subject to a privacy-by-design technique, while the majority will be deleted immediately. The server hosting this database is accessible only by authorised users through authentication (using passwords of high complexity). A firewall will also be in place to allow only specific (whitelisted) IPs to access the server and restrict the access of each whitelisted IP only to specific ports/services. Different access privileges to the data are available to ensure that the authorised users will only have access to the stored data on a need-to-know basis, i.e., that the authorised users will have access only to the stored psedonymised data needed to fulfil their tasks. Devices that will store a backup of the data will follow the same security procedures as the main server. For any remote interactions with the server (e.g., remote control or data transfer), secure protocols such as ssh/stfp are used. Any processing of the data is performed on that server. In case processing will be needed on other machines, the same security measures of the server will be applied to the respective machine. The metadata of the social media will be also stored in a local database that is secured (authentication mechanisms are enabled) and is also IP protected.

Will the collected data be shared?

The collected personal data (in their pseudonymised form) may be disclosed: (1) to all partners of the Consortium, through a password protected system; and (2) if this is required to third parties for the fulfilment of our legal obligations or is necessary for the fulfilment of the above data processing purposes and is in compliance with the applicable legal framework. The information collected will be also used to contribute towards several journal and conference publications as well as scientific contests, in line with Social Media/Web Policy. It is also highlighted that no personal data will be transferred outside the European Union (EU) or the European Economic Area (EEA). When this study is over, CERTH/ITI will be the only one responsible for the information collected.

Who will be responsible for all of the data when this study is over?

When this study is over, CERTH/ITI will be the only one responsible for the information collected.

How long will data be stored?

The storage duration of the data in their pseudonymised form will be the duration of the project plus five (5) years after the end of the project [i.e., 30 September 2030], to be available for demonstration in case of an inspection or an audit, as long as required to achieve the above purposes of processing, unless a longer retention period is required by law or for the establishment, exercise or defence of legal claims.

Will the collected personal data be used for other purposes?

All personal data collected in STARLIGHT will not be processed for any other purposes outside of those specified in this document.

Will the collected data be processed by automated tools supporting decision-making?

Your data will be used: (i) for scientific research purposes, (ii) to facilitate the functionality of other modules of the project, and (iii) for demo purposes. Data collected from you will only be used to test the capabilities of the STARLIGHT tools and you will not suffer any consequences of automated processing supporting decision-making.

What are your rights?

Your rights under GDPR are contained within articles 12-23 and 77. Some of your most important rights include:

- *Right to information:* you may request information about whether we hold personal information about you, and, if so, what that information is and why we are holding it. This information shall be provided within a reasonable period after obtaining the personal data, but at the latest within one month of receipt of request.
- *Right to access:* you may request to receive a copy of any personal information we may hold about you, if any, and to check that we are lawfully processing it.
- *Right to rectification*: you may ask us to rectify the information that we hold about you in case you consider that something is missing or is incorrect.
- *Right to erasure*: you may ask us to erase your personal data at any given moment without a specific reason.
- *Right to object*: you may request to stop processing, delete or remove your personal data at any desired moment where there is no good reason for us continuing to process it.
- *Right to data portability*: you have the right to request the transfer of your personal data in an electronic and structured form to another party or directly to you. This enables you to take your data from us in an electronically usable format and to be able to transfer your data to another party in an electronically usable format.
- Lodge a complaint with the Hellenic Data Protection Authority (https://www.dpa.gr).

Please note that the aforementioned rights may be restricted in the light of the GDPR (e.g. art. 89 par. 2) and the applicable national data protection legislation.

For the exercise of your rights and for any other data-related information you may contact us at m4d_ethics@iti.gr