ARTIFICIAL INTELLIGENCE SUPPORTING CROSS-BORDER COOPERATION IN CRIMINAL JUSTICE

JOINT REPORT PREPARED BY eu-LISA AND EUROJUST

JUNE, 2022

eu-LISA

EUROJUST

Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers machine learning technique for natural language processing
CBDCJ	Cross-Border Digital Criminal Justice
CCTV	Closed-circuit television
CEF	Connecting Europe Facility
CEPEJ	The European Commission for the Efficiency of Justice
DG HOME	European Commission Directorate-General for Migration and Home Affairs
DG JUST	European Commission Directorate-General for Justice and Consumers
e-CODEX	e-Justice Communication via Online Data Exchange
ECRIS-RI	European Criminal Records Information System – Reference Implementations
ECRIS-TCN	European Criminal Records Information System for Third-Country Nationals and stateless persons
EES	Entry-Exit System
ETIAS	European Travel Information and Authorisation System
Eurojust	European Union Agency for Criminal Justice Cooperation
eu-LISA	European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice
EUDPR	Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ L 295, 21.11.2018, p. 39
EUI	European Union Institution
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
ID	Identity Document
IT	Information Technology

IoT	Internet of Things
JHA	Justice and Home Affairs
JIT	Joint Investigation Teams
ML	Machine Learning
MoU	Memorandum of Understanding
NER	Named Entity Recognition
NERC	Named Entity Recognition and Classification
NIST	The National Institute of Standards and Technology
NLP	Natural Language Processing
NMT	Neural Machine Translation
PDF	Portable Document Format
RDF	Resource Description Framework
sBMS	shared Biometric Matching Service
SIS II	Schengen Information System Second Generation
US	United States of America
VAT	Value Added Tax
VIS	Visa Information System
W3C	World Wide Web Consortium

CONTENTS

ABOUT THE AGENCIES

THE EUROPEAN UNION AGENCY FOR THE OPERATIONAL MANAGEMENT OF LARGE-SCALE IT SYSTEMS IN THE AREA OF FREEDOM, SECURITY AND JUSTICE (EU-LISA)

Is an EU Agency established to provide a long-term solution for the operational management of large-scale IT systems, which are essential instruments in the implementation of the asylum, border management and migration policies of the EU.

The Agency currently **manages Eurodac**, the second-generation Schengen Information System (**SIS II**) and the Visa Information System (**VIS**). In addition to these, eu-LISA is **developing** the Entry/Exit System (**EES**), the European Travel Information Authorisation System (**ETIAS**) and the European Criminal Records Information System – Third-Country Nationals (**ECRIS-TCN**). These systems and the pre-existing ones are being **built/adapted to ensure Interoperability** – improved **access to information stored in EU information systems** and identity management at EU level. To that end, dedicated interoperability components are being developed.

The Agency was established in 2011 by Regulation (EU) No 1077/2011 and **started its activities on 1 December 2012**. In 2018, eu-LISA was given a larger mandate, which is detailed in Regulation (EU) 2018/1726.

eu-LISA's headquarters are in **Tallinn, Estonia**, and its operational centre is in **Strasbourg, France**. There is also a business continuity site for the systems under management based in **Sankt Johann im Pongau, Austria** and a Liaison Office in **Brussels, Belgium**.

EUROJUST, THE EUROPEAN UNION AGENCY FOR CRIMINAL JUSTICE COOPERATION

Is a unique hub based in **The Hague, Netherlands**, where national judicial authorities work closely together to fight serious organised cross-border crime. Eurojust's role is to help make Europe a safer place by coordinating the work of national authorities – from the EU Member States as well as third States – in **investigating and prosecuting transnational crime.**

Eurojust operates on the basis of **Article 85 of the Treaty of Lisbon** and the Eurojust Regulation (EU) 2018/1727, which became applicable on **12 December 2019**. The Regulation determines Eurojust's mandate, governance structure, data protection regime and the framework for establishing agreements with non-EU countries.

Each participating EU Member State seconds a National Member to Eurojust. The National Members form the **College of Eurojust, which is responsible for the Agency's operational work**. The College, in turn, is supported by the Eurojust Administration, which includes, among others, case analysts, legal advisors and data experts. The Agency's custom-built premises provide secure meeting facilities with possibilities for interpreting into all EU languages, and a meeting room specially developed for coordination centres, from which joint action days can be monitored and coordinated in real time.

Eurojust has developed a cohesive international network that **grants prosecutors around the European Union access to more than 50 jurisdictions worldwide.** The Agency has signed cooperation agreements with a dozen non-EU countries, several of which have seconded Liaison Prosecutors to Eurojust to work on cases with their counterparts in the College. Eurojust also works closely with other EU agencies and partners that support the various stages of the criminal justice chain, including law enforcement and anti-fraud bodies.

EXECUTIVE SUMMARY

Over the past few years, significant groundwork has been done by both Eurojust and eu-LISA, as well as the European Commission to **speed-up digital transformation and the adoption of IT solutions based on artificial intelligence (AI) in the field of Justice and Home Affairs (JHA).** For example, the study on **Cross-border Digital Criminal Justice**, commissioned by DG JUST, performed an assessment of business needs in cross-border criminal justice cooperation and proposed a set of solutions tackling those needs. These solutions included, for example, the **creation of a collaboration platform for Joint Investigation Teams (JITs)**[1] and a **redesigned Eurojust Case Management System**, where AI-based components were suggested for both. **Artificial intelligence was also defined as a priority area in the Action Plan for e-Justice 2019-2023.**

This report goes a step further in exploring the relevant technologies and possible use-cases for the application of AI within the judicial field, specifically focusing on the **support of cross-border collaboration in criminal justice.** The first section of the report sets the scene, **outlining the legal and policy context for the introduction of new technologies, including AI, in the field of justice.** The second section of the report is dedicated to the **discussion of relevant technologies** and **their possible applications in the context of cross-border criminal justice cooperation**.

Specifically, two categories of technologies are **explored**:

NATURAL LANGUAGE PROCESSING (NLP)

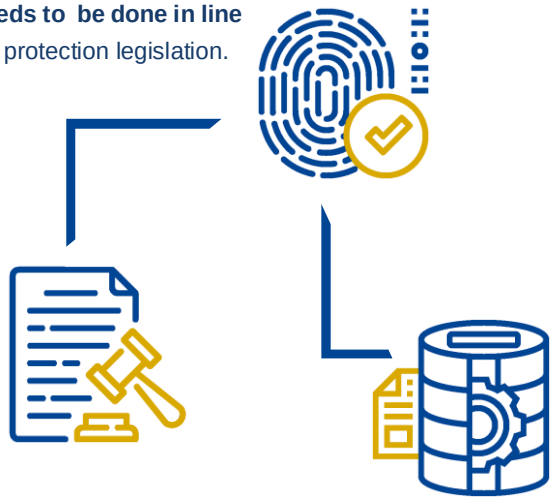
COMPUTER VISION

Both technologies are particularly relevant in applications where **the processing of large-scale unstructured data is necessary.** Unstructured data in text format includes documents of different types (e.g. contracts and invoices, including paper-based), emails and other written communications, etc. Unstructured data that can be processed using computer vision includes static as well as dynamic image data (including live video streaming). Specifically, to facilitate the work of judicial practitioners, **NLP technologies can be used for** the **creation of common ontologies** and **semantic networks in the justice domain**, for voice assistants and chatbots, automated translation in cross-border investigations, for text summarisation and for named entity recognition for investigative purposes or for anonymisation of documents. NLP technologies in combination with other approaches, such as knowledge graphs, for example, can also be **effectively used in legal research in order to facilitate the identification of relevant statutes and case-law. Computer vision**, and **in particular biometric recognition** technologies, can be deployed in the context of criminal investigations, in particular where the **analysis of large volumes of image or video data** is necessary **in the context of identification of potential victims or perpetrators of crime**, for example.

In addition to the processing of unstructured data, these technologies in combination with other approaches (e.g. generative adversarial networks) **can be used to effectively anonymise both text-based and visual data**. Specifically, with the use of named entity recognition tools, text-based documents can be anonymised by removing all possible identifiers (e.g. names, dates, places, etc.), ensuring that the data is rendered anonymous in such a manner that the data subjects are not or no longer identifiable. Similarly, by applying a combination of facial recognition and Generative Adversarial Networks (GANs) for the generation of synthetic biometrics, video streams can be anonymised by replacing real faces with synthetic images, thus protecting the identities of people captured on video.

[1] The legislative proposal establishing a collaboration platform to support the functioning of JITs was launched by the European Commission on 1 December 2021 and is currently under negotiations in the Council and in the European Parliament.

In most cases, such technologies are already available on the market, but may need adaptation to the specific context. However, prior to the implementation of these technologies in practice, authorities considering such applications should perform a robust risk assessment, as well as assess the legal, ethical implications, as well as implications for fundamental rights and freedoms of individuals of the use of such technologies. It is also important to conduct the testing and evaluation of such technologies in real-life conditions (using test data sets only), in order to ensure that their performance meets the relevant standards, in particular as regards accuracy and bias. Therefore, we suggest that a risk-based approach shall be developed and applied to determine the safeguards and the deployment model for AI systems in practice. **The implementation of these technologies in operational environments needs to be done in line with the relevant EU legal framework** and especially the applicable data protection legislation.



INTRODUCTION

Tackling cross-border crime has been a persistent challenge for law enforcement and judicial authorities. For financial crime, trafficking of human beings, terrorism or cybercrime, criminal activities have become borderless. These and many other cross-border criminal activities are facilitated by the use of modern information and communication technologies, including encryption and artificial intelligence. Tackling cross-border crime requires effective collaboration between authorities from different EU Member States and third countries. It also requires the effective use of modern technologies on par with those used by the organised criminal groups.

To address some of these challenges, in recent years the EU has made substantial progress in the justice and home affairs (JHA) area in terms of legislative developments and cross-border operational cooperation. Furthermore, in order to support the relevant Member States' authorities in the fight against cross-border crime, EU JHA agencies, including eu-LISA and Eurojust, have identified the need to modernise digital infrastructure, and establish secure communication channels for the exchange of evidence between national authorities, as well as with EU judicial authorities.

In 2017, eu-LISA and Eurojust signed a Memorandum of Understanding[2] with the aim of jointly updating the existing digital systems and processes and setting up new ones, in order to modernise the way criminal justice systems operate and to enhance cooperation in criminal justice matters in the EU. On the basis of this MoU and the relevant regulations, eu-LISA has become a partner for the justice community, thus expanding its involvement beyond border management and internal security through a range of activities:

- **eu-LISA** is responsible for the **technical maintenance of ECRIS RI**, a decentralised platform for the Member States for exchanging information on criminal convictions.
- **eu-LISA** has over the past few years been actively involved in the **development of ECRIS-TCN,** the **European Criminal Records Information System** for Third Country Nationals.
- **eu-LISA** was entrusted with the take over and the **operational management of the e-CODEX**[1] system, which **facilitates cross-border cooperation** among European judicial authorities **and facilitates access to justice** for both citizens and businesses.
- Finally, according to a proposal by the European Commission, **eu-LISA** will be responsible for the **development and operational management of the future Joint Investigation Teams collaboration platform**, which will support cross-border judicial collaboration.

Continuing fruitful collaboration, eu-LISA and Eurojust have been developing a set of guidelines for digitising criminal justice systems in the EU. These guidelines, will define the key principles that need to be taken into account when developing new digital systems. One of the principles is to adopt relevant IT solutions based on artificial intelligence where useful, secure and in compliance with data protection and privacy rules.

Because the use of AI solutions might bring interference with fundamental rights and freedoms of individuals, any solution to be considered for implementation will have to comply with the applicable EU legislation, and in particular the data protection legislation, ensure transparency, proportionality, data minimisation, purpose limitation, data protection by design, accountability, as well as allow for human intervention when necessary.

In addition to the above, significant groundwork has already been done by both the European Commission and agencies to speed up digital transformation and the adoption of IT solutions based on artificial intelligence in the field of JHA. For example, the study on Cross-border Digital Criminal Justice[3], commissioned by DG JUST, performed an assessment of

[2] https://www.eurojust.europa.eu/memorandum-understanding-between-eurojust-and-eu-lisa
[3] https://op.europa.eu/s/sWZX

business needs in cross-border criminal justice cooperation, and proposed a set of solutions to tackle those needs, including a few solutions regarding the use of AI-tools.

These solutions included, for example, the creation of a collaboration platform for Joint Investigation Teams (JITs), which allows judicial and law enforcement authorities of two or more States to carry out joint criminal investigations in the involved States, for a limited duration and for a specific purpose, and a redesigned Eurojust Case Management System, both of which should include AI-based components, such as automated data extraction or machine translation. Artificial intelligence was also defined as a priority area in the Action Plan for e-Justice 2019-2023[4]. The study on the use of innovative technologies in the justice field[5], coordinated by the European Commission, addressed one of the actions defined in the Action Plan, providing a comprehensive overview of the use of innovative technological solutions including AI. These studies have been followed by a set of legislative proposals aimed to further advance the digitalisation of judicial systems[6].

Similarly, eu-LISA has made significant steps towards the application of AI in its core business, and towards supporting its key stakeholders, such as Member State authorities, the European Commission and JHA agencies with developing and testing AI tools. In 2020, eu-LISA published a report on Artificial Intelligence in the Operational Management of Large-scale IT Systems[7], outlining the range of opportunities for AI within the scope of eu-LISA core business. Specifically, there are several projects where eu-LISA has already implemented or is considering deploying AI.

- For example, AI has been implemented within the scope of the shared biometric matching service (sBMS), in particular for improving the accuracy of biometric matching algorithms.
- Also, **eu-LISA**, with the support of a contractor, is currently working together with the European Commission and Member State authorities on **developing a VisaChat application**, which will support applicants in the application process for **short-stay visas** in the **Schengen area.**
- **eu-LISA** also aims to implement **AI to improve its internal processes**, in particular focusing on **infrastructure management**, service desk optimisation, cybersecurity monitoring, detection and response.

To support knowledge and information exchange regarding the development of AI solutions in the field of JHA and support capability building within this field, eu-LISA has also established a Working Group on AI involving representatives of Member State authorities, the European Commission and relevant EU Agencies. In addition, eu-LISA is supporting DG HOME in defining the requirements for the setup within eu-LISA of an AI Centre of Excellence for the JHA area.

This report goes a step further in exploring the relevant technologies and possible use-cases for the application of AI within the judicial field, specifically focusing on supporting cross-border collaboration in criminal justice. Following the introduction, the second section outlines the policy and legal context for the use of artificial intelligence technologies in cross-border collaboration in criminal justice, setting the overall scene for the report. The third section provides an overview of technologies and use cases for application of AI in cross-border criminal justice cooperation, including different applications of natural language processing and biometric recognition technologies and their potential for application in the specific use cases. The fourth and final section provides some preliminary conclusions and indicates the next steps towards the effective adoption of AI.

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019XG0313(02)&rid=6
[5] https://op.europa.eu/s/sW44
[6] https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6387
[7] https://op.europa.eu/s/u3RU

POLICY AND LEGAL CONTEXT

Since AI tools have been widely used in the private sector for years and seeing that AI can also bring about new risks or negative consequences for individuals or the society in general, there has been a growing call for legislators to develop specific regulations that would address the legal aspects of the use of these tools and technologies in general, this way ensuring a balanced approach.

The same has been true at European level, as this is not a phenomenon that can be tackled at national level only. Already in 2017, the European Council called for a way to address AI as an emerging trend, while ensuring a 'high level of data protection digital rights and ethical standards'[8]. This was further emphasised by the European Council in the Council Conclusions of October 2020 on 'Access to Justice – seizing the Opportunities of Digitalisation''[9] by stating that the application of AI in the justice sector may also contain the risk of perpetuating and possibly strengthening existing discrimination, including stereotypes, prejudices or structural inequalities, and of allowing distorted or opaque decision-making, and thereby result in the impairment of fundamental rights such as human dignity, the right to liberty, non-discrimination, privacy and data protection, and the right to a fair trial. The conclusions first note that an adequate level of digitalisation is a prerequisite for any use of AI tools, and coordination at EU level is needed to ensure synergies. The conclusions recognise that the use of AI has the potential to improve the functioning of the judicial systems, but underline that the decision-making powers of judges and judicial independence must not be influenced by AI.

The overall tone of the discussion around AI in the EU was set by the White Paper on Artificial Intelligence[10], which set key principles for the European approach to AI. More specifically for the JHA area, the challenges and possibilities of AI technologies were identified by the 2020 Digital Criminal Justice study, commissioned by DG JUST and supported by Eurojust[11]. The solutions proposed by the study were further analysed in the European Commission Communication in December 2020 on the further enhancement of digitalisation of justice[12].

The communication outlined that the use of AI applications in the justice field could be very beneficial. The communication also saw the need for better coordination at EU level, which could avoid duplication of national efforts and create significant synergies. The communication outlined some of the areas where this could be useful in the field of justice – anonymisation of court decisions, speech-to-text conversion and transcription, machine translation, chat-bots supporting access to justice and robot process automation. Additionally, the communication emphasised that while AI-based applications in the justice system are clear, there are also considerable risks associated with their use, especially from the perspective of fundamental rights.

In April 2021, the European Commission published a proposal for a Regulation laying down harmonised rules on artificial intelligence that sets harmonised rules for the development, placement on the market and use of AI systems in the Union following a proportionate risk-based approach[13].

The proposal identifies the field of justice as one of the fields where the use of certain AI tools or applications should be considered as high risk, especially with regard to AI systems intended to assist judicial authorities in researching and interpreting facts and the law and in applying the law to a concrete set of facts. Annex III of the proposal also identifies as high risk systems any systems to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences. However, high-risk categorisation should not apply to AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual

[8] European Council, European Council meeting (19 October 2017), *https://www.consilium.europa.eu/media/21620/19-euco-final-conclusions-en.pdf*
[9] The Council of the EU, 'Access to justice – seizing the opportunities of digitalisation', 2020, https://data.consilium.europa.eu/doc/document/ST-11599-2020-INIT/en/pdf
[10] https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
[11] https://www.eurojust.europa.eu/judicial-cooperation/judicial-cooperation-instruments/digital-criminal-justice
[12] European Commission 2020, https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2246
[13] European Commission, 2021/0106(COD), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

cases, such as anonymisation or pseudonymisation of judicial decisions, documents or data, communication between personnel, administrative tasks or the allocation of resources.

In terms of biometric recognition, the proposal foresees that the use of real-time remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement is prohibited unless certain limited exceptions apply. Therefore, such tools can strictly be used for the targeted search of potential victims, the prevention of specific, substantial and imminent threat to life or physical safety of persons or for the search for suspects of serious crimes, under certain conditions, as specified in Article 5 of the proposal. Moreover, the legislative proposal has special provisions applicable regarding the date of start of application, in what concerns the AI systems which are components of the large scale IT systems, including ECRIS-TCN, where the AI Act foresees an exception to the rule of 24 months following the entry into force of the AI Act (Article 83 and Annex IX).

In this context, it is important to note that on 18 June 2021, the European Data Protection Board and the European Data Protection Supervisor adopted Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [14]. Both regulators agreed that the Proposal has prominently important data protection implications. They welcomed the risk-based approach underpinning the Proposal, however called for further alignment with the GDPR and Regulation 2018/172515. The EDPB and the EDPS agree with the Proposal when it states that the classification of an AI system as high-risk does not necessarily mean that it is lawful per se and can be deployed by the user as such. Further requirements resulting from the EU data protection law may need to be complied with by the controller.

"The use of AI has the potential to improve the functioning of the judicial systems, but the decision-making powers of judges and judicial independence must not be influenced by AI."

[14] EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) | European Data Protection Board (europa.eu)

[15] Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ L 295, 21.11.2018, p. 39

On 6 October 2021, the European Parliament adopted a resolution on AI in criminal law and its use by the police and judicial authorities in criminal matters[16]. The resolution acknowledges the benefits and risks that artificial intelligence will bring for the judicial authorities and law enforcement. The European Parliament reaffirmed that all AI solutions for law enforcement and the judiciary also need to fully respect the principles of human dignity, non-discrimination, freedom of movement, the presumption of innocence and right of defence, including the right to silence, freedom of expression and information, freedom of assembly and of association, equality before the law, the principle of equality of arms and the right to an effective remedy and a fair trial, in accordance with the Charter and the European Convention on Human Rights; stresses that use of AI applications must be prohibited when incompatible with fundamental rights. The resolution pointed to the risks related in particular to data leaks, data security breaches and unauthorised access to personal data and other information related to, for example. criminal investigations or court cases that is processed by AI systems; underlined that security and safety aspects of AI systems used in law enforcement and by the judiciary need to be considered carefully and be sufficiently robust and resilient to prevent the potentially catastrophic consequences of malicious attacks on AI systems; stressed the importance of security by design, as well as specific human oversight before operating certain critical applications and therefore calls for law enforcement and judicial authorities only to use AI applications that adhere to the privacy and data protection by design principle in order to avoid function creep.

The resolution identifies several possible use cases for AI tools for judicial authorities, mentioning as an example case-law management and additional searching possibilities. The resolution also lists several requirements for AI tools in law enforcement and the judiciary to adhere to. Specifically, page four of the resolution sets out that any AI tools in the judiciary should '*be safe, robust, secure and fit for purpose, respect the principles of fairness, data minimisation, accountability, transparency, non-discrimination and explainability, and that their development, deployment and use should be subject to risk assessment and strict necessity and proportionality testing, where safeguards need to be proportionate to the identified risks."*

On 1 December 2021, the European Commission published a package of legislative proposals dealing with the issues of digitalisation in cross-border judicial cooperation. These included first a proposal for a Regulation laying down rules on digital communication in judicial cooperation procedures in civil, commercial and criminal matters and a proposal for a Directive aligning the existing rules on communication[17]. The package also includes a proposal for establishing a collaboration platform to support the functioning of Joint Investigation Teams (JITs)[18] and a new initiative on the digital information exchange in terrorism cases, which is aimed at rendering the exchange of information between the national competent authorities and Eurojust and the European Judicial Terrorism Register more efficient[19].

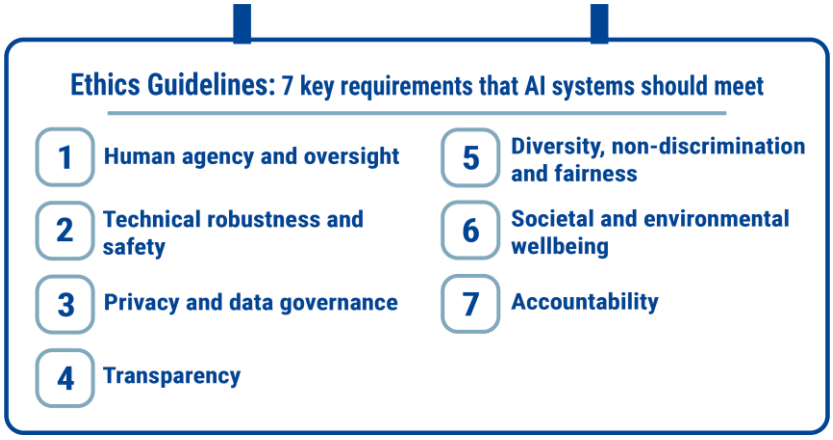
[16] European Parliament 2021, https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html
[17] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/digitalisation-justice/digitalisation-cross-border-judicial-cooperation_en
[18] https://ec.europa.eu/info/publications/joint-investigation-teams-jits-collaboration-platform_en
[19] https://ec.europa.eu/info/publications/digital-information-exchange-terrorism-cases_en

2.1. Ethical and fundamental rights considerations

Fundamental rights and ethical considerations are especially relevant when discussing the possible AI applications in the field of justice. The cornerstones of the fundamental rights framework of the EU are the Charter of Fundamental Rights of the EU[20] and the European Convention of Human Rights (ECHR). The most relevant fundamental rights to be analysed in terms of AI applications in the field of law enforcement and judiciary are human dignity (Article 1 of the Charter); the respect for private and family life (Article 7 of the Charter and Article 8 of the ECHR); the protection of personal data (Article 8 of the Charter); equality before the law (Article 20 of the Charter); non-discrimination (Article 21 of the Charter and Article 14 of the ECHR) and the right to an effective remedy and to a fair trial (Article 47 of the Charter).

In addition, the relevant secondary legislation is the General Data Protection Regulation (GDPR), Regulation (EU) 2018/1725[21] (EUDPR), the ePrivacy Directive[22] and, specifically for law enforcement and the judiciary, the Law Enforcement Directive (LED)[23], as for most of the time, AI tools are using at least some kind of personal data. Both GDPR and the LED set main data protection principles, data subjects' rights and mention automated data processing and ban any such decisions that are the result of automated processing which produce legal effects for persons and affect them, unless explicitly authorised by law[24]. As mentioned before, the EDPS and the EDPB strongly recommend to make clear in the Proposal regarding the Regulation laying down harmonised rules on AI that the use of AI systems should comply with data protection law, namely with the GDPR, Regulation (EU) 2018/1725, and the LED.

On the basis of this and other applicable legislation and practice, the High-Level Expert Group on AI analysed the application of AI and, on 8 April 2019, published the Ethics Guidelines for Trustworthy Artificial Intelligence[25]. The guidelines identify seven key requirements that AI systems should meet:



While the above ethics guidelines were not specific to the judicial field, but generally for any application of AI, a working group at the Council of Europe – the European Commission for the Efficiency of Justice (CEPEJ) has analysed the possible use cases of AI in the judicial systems and identified some examples of use cases where the use of AI would be most recommended, naming case-law enhancement, improving access to justice and law and the creation of new strategic tools on court management as the most perspective potential uses for the judiciary[26].

---

[20] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT
[21] Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ L 295, 21.11.2018, p. 39
[22] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) as amended by Directive 2006/24/EC and Directive 2009/136/EC
[23] Directive (EU) 2016/680. OJ L 119, 4.5.2016, p. 89-131.
[24] Law Enforcement Directive, Article 11.
[25] European Commission 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[26] European Commission for the Efficiency of Justice (CEPEJ), Ethical charter on the use of artificial intelligence in judicial systems and their environment, 2018, https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

The EDPB and the EDPS in their Joint Opinion on the proposal of the AI act draw attention to the fact that remote biometric identification of individuals in publicly accessible spaces poses a high-risk of intrusion into individuals' private lives, with severe effects on the populations' expectation of being anonymous in public spaces. For these reasons, the EDPB and the EDPS call for a general ban on any use of AI for an automated recognition of human features in publicly accessible spaces - such as of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioural signals - **in any context**. The Joint Opinion states that for consistency reasons, AI systems for large-scale remote identification in online spaces should also be prohibited under Article 5 of the Proposal. Moreover, the EDPB and EDPS recommend a ban, for both public authorities and private entities, on AI systems categorizing individuals from biometrics (for instance, from face recognition) into clusters according to ethnicity, gender, as well as political or sexual orientation, or other grounds for discrimination prohibited under Article 21 of the Charter, or AI systems whose scientific validity is not proven or which are in direct conflict with essential values of the EU (e.g., polygraph, Annex III, 6. (b) and 7. (a) of the Proposal). Accordingly, "biometric categorization" should be prohibited under Article 5 of the Proposal. Further, AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending of reoffending criminal offences, or for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of a natural person or on assessing personality traits and characteristics or past criminal behaviour, used according to their intended purpose will lead to pivotal subjection of police and judicial decision-making, thereby objectifying the human being affected. Such AI systems touching the essence of the right to human dignity should be prohibited under Article 5.

Therefore, policy developments in the EU clearly recognise the potential of AI tools for the judicial field, but also point out the possible risks of AI, which must be taken into consideration when developing/deciding to use such tools. In any case, the use of any automated tools must be lawful.[27]

"Policy developments in the EU clearly recognise the potential of AI tools for the judicial field, but also point out the possible risks of AI, which must be taken into consideration when developing/deciding to use such tools."

[27] For further information and analysis on fundamental rights and AI, The European Union Agency for Fundamental Rights (FRA) has published several reports raising fundamental rights concerns when it comes to the use of AI, which include, but also go beyond privacy and data protection. In the report 'Getting the future right – Artificial intelligence and fundamental rights' FRA highlights important fundamental rights that need to be considered when developing and using AI, such as data protection, non-discrimination, access to an effective remedy, and good administration. FRA suggests to assess fundamental rights before using AI that can have an impact on people's life.

ARTIFICIAL INTELLIGENCE TO SUPPORT CROSS-BORDER JUDICIAL COOPERATION: TECHNOLOGIES AND USE CASES

AI has been an essential component of the digital transformation of commercial enterprises for over a decade. More recently, AI has been increasingly explored in the context of the public sector as an important driver in achieving greater effectiveness, efficiency and quality in public service delivery. Some of the use cases showing clear benefits in terms of efficiency and quality of public service delivery, such as the use of chat-bots for customer assistance or the use of computer vision tools to assist in determining agricultural subsidies, are explored in the report "Overview of the use and impact of AI in public services in the EU"[28]. The area of justice and home affairs is no exception in this regard. In customs, for example, AI is being effectively deployed to support risk assessment and assist in fraud detection. AI has also become an integral part of biometric identification, leading to significant improvements in the performance of automated biometric identification systems, although certain important challenges still need to be addressed, together with relevant fundamental rights and data protection principles regarding the use of such systems.

The recent study on the use of innovative technologies in the field of justice has shown that many Member State authorities already have started to use first low-risk AI-based solutions, such as automated transcription and anonymization solutions[29] in supporting their day-to-day activities and business processes. These developments also provide grounds to further study the possible usability of AI tools also in support of cross-border collaboration in criminal investigations.

However, the compliance with the EU data protection legislation is a precondition before deciding on any use of such technologies, notably the principles of necessity, proportionality, data minimisation and data protection by design should always be considered upfront. Needless to say, the use of AI systems while processing personal data should not go beyond what is legally possible and should comply with the mandate and competence of the respective EUI(s). Besides, as some of those systems will fall in the category of high risk, there will be a necessity to conduct Data Protection Impact Assessments according to Article 39 of the EUDPR and to even carry out the prior consultation with the EDPS in certain cases as foreseen in Article 40 of the EUDPR. In this section we present several practical examples of the use of AI in the context of cross-border judicial cooperation, as well as specific technologies that may be used to address practical challenges.

3.1. Natural language processing technologies and their applications

The amount of data that is being produced on a daily basis has been growing exponentially since the advent of the Internet, and has accelerated with the development of the Internet of Things (IoT). This is further complicated by the fact that most of this information is unstructured and therefore cannot be processed using conventional techniques (e.g. statistical analysis).

Structured data, such as accounting data or airline ticketing data, is normally structured according to a certain data model, which organises different elements of data and standardises relationships or connections among those. Structured data, stored in a relational database (SQL) is relatively easy to manipulate, search and analyse without specialised data science skills and using conventional statistical analysis techniques, for example.

[28] Misuraca, G. and Van Noordt, C., AI Watch - Artificial Intelligence in public services, Publications Office of the European Union, Luxembourg, 2020, https://publications.jrc.ec.europa.eu/repository/handle/JRC120399

[29] European Commission, Directorate-General for Justice and Consumers, *Study on the use of innovative technologies in the justice field : final report*, Publications Office, 2020, https://data.europa.eu/doi/10.2838/585101

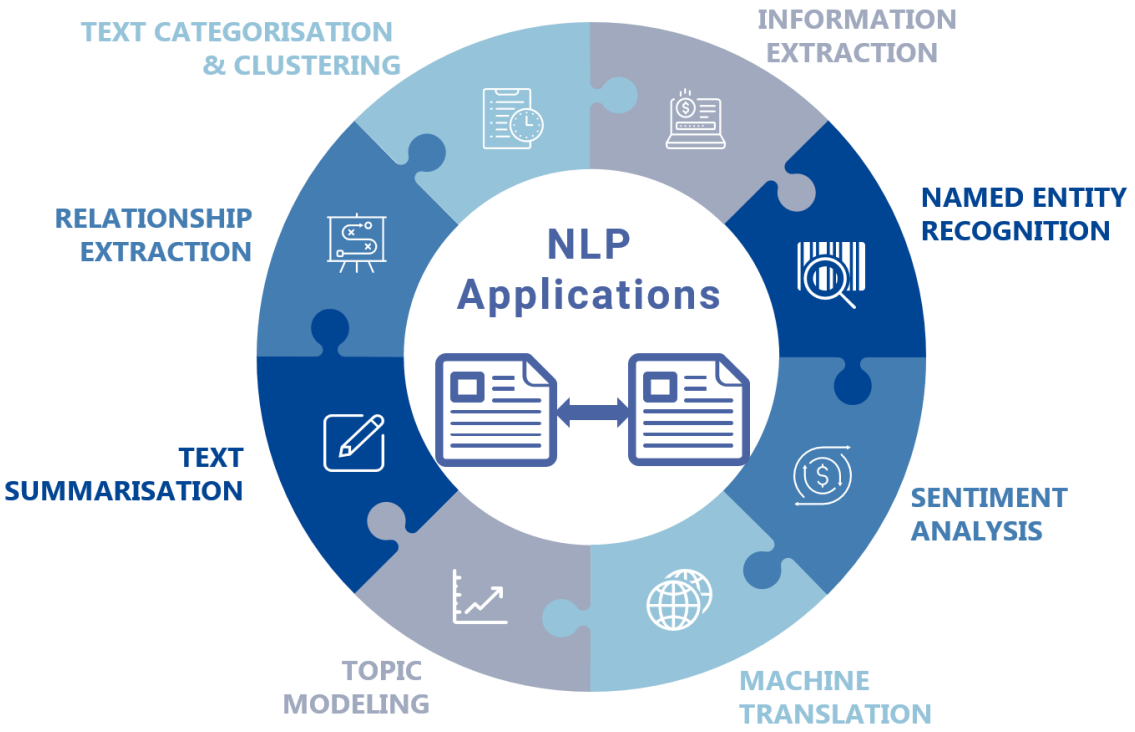
Unstructured data, on the other hand, is not organised and standardised, and is normally stored in native format in massive data storage, often referred to as data lakes. Unstructured data includes such data types as:

SURVEILLANCE DATA, AUDIOVISUAL DATA, GEO-SPATIAL DATA

TEXT-BASED MEDIA, SUCH AS DOCUMENTS, EMAIL COMMUNICATIONS AND RECORDS

DATA PRODUCED BY DIFFERENT DEVICES AND SENSORS (IOT).

Unstructured data has a number of benefits, including the speed at which it can be collected, flexibility in terms of formats and storage, and scalability; however, processing unstructured data requires specialised tools for managing unstructured data, as well as data science expertise for data processing and analysis. Judicial investigations are making use of a vast amount of unstructured data. For unstructured data, in particular in textual form (or speech, which can be processed by a machine), Natural Language Processing (NLP) technologies are essential. The figure below provides a high-level overview of the application of NLP technologies, and some of these applications relevant in the specific context of cross-border judicial cooperation is discussed in more detail directly below.



▲ *Figure 1: High-level overview of Natural Language Processing applications*

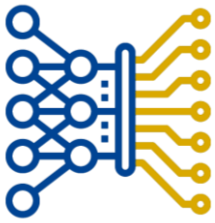
3.1.1. Automated document processing

One of the most obvious use cases for AI in the context of judicial cooperation is automated document processing. Automated document processing normally consists of at least two components: computer vision for optical character recognition and natural language processing for document analysis and classification. Automated document processing systems address the following business needs in environments where document processing is a significant part of the activity:

- conversion of scanned paper documents, PDFs and images into searchable and editable documents;
- processing of high volumes of standardised documents;
- classification of documents and creation of searchable archives.

Automated document processing can be an integral part of a case management system, supporting the prosecutors and administrative staff involved in criminal investigations with back-office functions relating to document processing. The system can effectively extract the data necessary to classify the document and register it as relevant to a specific case, thus significantly reducing the need for manual document processing. Such automation could be useful in order to automate the creation of Case Information Forms, which are used to collect, store and access information containing no personal data in a structured and systematic manner to support National Desks with relevant knowledge, experience and best practice[30]. Furthermore, with the conversion of paper-based documents, PDFs and images into searchable documents, the system would allow to further explore the possibility to deploy other AI-driven tools, including machine translation, e-Discovery and analytical tools for unstructured information, such as named entity recognition, described in more detail below. It is important to note that the use of optical character recognition tools to process scanned documents may be prone to errors. Therefore, the application of such tools to processing documents that may constitute evidence, must be done with extreme caution.

3.1.2. Machine translation



Another application of AI that is highly relevant in cross-border collaboration is the machine translation. There is a wide range of machine translation use-cases where these can be applied in the context of cross-border judicial cooperation, including in the work of Joint Investigation Teams (JITs). As indicated in the JITs evaluation reports[31], one of the most common challenges that members of the JITs face is the need to communicate in multiple languages, and analyse evidence in multiple languages. In practice, this often results in the need to translate large amounts of materials, frequently containing specialist language or using less common languages. Even if machine translated documents will not be possible to be used as evidence, it will at least provide a quick overview and indications which parts of the documents might be most important to have officially translated. The need for machine translation was also identified in the Cross-border Digital Criminal Justice study[32], with a specific focus on a translation engine with Text-to-Speech functionality for immediate translation and consultation during JIT meetings. Automatic translation is also seen as one of the possible functions that could be offered in the context of the future JIT Collaboration platform[33].

Generic automated translation systems are widely available on the market today, including as open source solutions and tools[34]. However, when used for translation of domain-specific texts, generic translation systems will not be effective, especially in smaller, less common languages. One of the challenges relating to the use of generic translation systems to translate domain-specific texts is the way terminology is handled. Generic machine translation systems do not distinguish terminology from other phrases, as out-of-domain translation hypotheses will prevail due to their higher frequency in the parallel corpus. Some of the issues relating to the translation of terminology includes the use of incorrect translation equivalents from different domains, the use of obsolete variants, and the inconsistent use of terminological synonyms. Therefore, integration of domain-specific terminology in translation is essential to produce high quality translation. In addition to terminology, one of the persistent issues in machine translation is that the meaning of speech or text depends on a specific context, where human interpretation is often necessary in order to determine context-appropriate translation.

[30] https://www.eurojust.europa.eu/sites/default/files/Publications/AnnualReport/EUROJUST-CAAR2018_EN.pdf
[31] e.g. the third JIT evaluation report: https://www.eurojust.europa.eu/third-jit-evaluation-report
[32] Cross-border Digital Criminal Justice, Final Report. https://op.europa.eu/s/tULv (p. 174).
[33] European Commission 2021, Analytical supporting document accompanying the proposal on the Regulation on the JIT collaboration platform, https://ec.europa.eu/info/sites/default/files/3_2_178653_analytic_regul_jit_en.pdf.pdf
[34] References to some of those can be found here: https://opensource.com/article/17/6/open-source-localization-tools

Terminology integration in statistical machine translation is a well-studied area, where a range of established methods exist to impose correct terminology on statistical machine translation systems. For neural machine translation tools, the improvement of terminology translation is a very active area of research and so far existing experiments have not delivered significant improvements for neural machine translation (NMT) systems.

Therefore, to ensure that the translation system works well, it needs to be developed as a specialised system. There are different methods for this. Some of the generic machine translation tools based on NMT and trained on generic corpus. Retraining such models on specialist in-domain resources is one way to develop a specialised system. Other methods use a combination of different techniques and add in-domain knowledge to the system, thus allowing the production of high-quality translations of domain-specific texts, even when applied to low-resource languages. Most of these systems operate via application programming interfaces (APIs), and can be implemented on private cloud or on-premises infrastructure, in particular in use cases where data protection and security are a major concern, especially where operational personal data is concerned.

One of the technology providers relevant in public administration context is the Connecting Europe Facility Digital programme (CEF Digital) with the eTranslation tool[35]. The eTranslation tool covers 24 official EU languages, as well as Norwegian and Icelandic. In addition to the official EU languages, the eTranslation includes Arabic, standard Chinese, Japanese and Russian. One of the major benefits of the eTranslation tool provided by CEF Digital is that it already includes domain-specific engines, including the engine based on the corpus provided by the European Court of Justice. Domain-specific engines can also be built upon request by EU bodies and institutions that need a specialised translation engine. It is important to stress that customisation is not a drag-and-drop operation and requires major resources, whether with the CEF eTranslation or any other provider. In most cases, in particular when using NMT models, the development of domain-specific engines will require a large in-domain parallel corpus, which is not always available and may be very expensive to generate.

"The implementation of automated translation tools in the workflow of JITs may significantly improve performance by reducing time spent on translating evidence and making evidence more directly accessible to the team."

The CEF Digital eTranslation tool is available as a web service and can be integrated as a building block in an online application. In those use-cases where high levels of security are required, that cannot be ensured by standard https internet protocols, eDelivery[36] can be used as a secure delivery channel with the eTranslation web service[37]. eDelivery is used in e-CODEX and therefore will meet the necessary security requirements.

The implementation of automated translation tools in the workflow of JITs may significantly improve performance by reducing time spent on translating evidence and making evidence more directly accessible to the team. Sworn translation will still be necessary to ensure the admissibility of translated evidence in court; however, automated translation can deliver major value at the stage of investigation and significantly reduce the time and costs for sworn translation, considering that only post-editing will be necessary in most cases instead of translation.

[35] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation
[36] eDelivery is a Connecting Europe Facility Building Block designed to ensure secure transmission of documents via the internet. More information on eDelivery is available here: https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eDelivery
[37] More information available here: https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=251168527

3.1.3. Automated summarisation systems

Another application of natural language processing technologies that may be relevant in the context of cross-border criminal justice cooperation is the use of automated text summarisation systems. Text summarisation systems are useful in applications where large amounts of information need to be processed in a limited amount of time, in particular in situations where processing of such information by humans is not feasible and where precision is not mission-critical. For example, text summarisation systems have effectively been used in academic environments, where the number of academic papers published every year that need to be reviewed by researchers is growing rapidly. These tools allow academics to extract from research publications only the most relevant information, significantly reducing the time needed to perform systematic literature reviews, for example in the biomedical domain where the amount of information is growing exponentially[38].

Text summarisation is a non-trivial task for computer systems, as they lack the human ability to make sense of information in context and the language ability that is necessary to represent processed information in a way that is legible to human consumers[39]. Although automated text summarisation originates in the 1950s[40], major advances in this field did not emerge until the early 2000s. Similar to the developments in other areas of natural language processing, and AI more generally, these advances are largely connected to the emergence of the internet and the wider availability of large text corpora, which are necessary for the development of advanced methods relying on machine learning algorithms, including neural networks.

Text summarisation solutions are largely based on two techniques: extraction and abstraction. Extractive summarisation produces summaries by selecting a subset of sentences from the original text using statistical methods. Unless additional post-editing tools are applied or human post-editing is performed, summaries produced by extractive methods are not easily readable. Abstractive text summarisation relies on machine learning methods, including deep neural networks, with the aim of producing human-readable summaries of text containing the most relevant information[41]. Most of the text summarisers available today are based on extractive approaches and include algorithms and techniques such as TextRank[42], LexRank[43] or Latent Semantic Analysis.

Looking at the field of justice more specifically, before any potential use of the summarisation tools it needs to be considered that the summaries might be incorrect. Even though summarisation solutions can help organisations deal with large volumes of textual information, such as seized documentation, automated summarisation cannot match human performance and can only help to provide a high-level and quick understanding of the content of documents, thus making the information more accessible for in depth human analysis. In any case, documentation used for evidence will need to go through human verification and analysis and the users need to be aware of not rely on the output of the system only. Although significant advances have been made over the past two decades in developing text summarisation systems, today there are no one-size-fits-all solutions, as systems trained on generic text corpus will not be suitable for processing specialist text. A cursory literature review on the subject of automated legal text summarisation suggests that extractive techniques are still mostly dominant in this field, but there are increasing attempts to combine extractive and abstractive techniques to improve the quality of the summaries produced.

---

[38] Mishra, R., Bian, J., Fiszman, M., Weir, C.R., Jonnalagadda, S., Mostafa, J. and Del Fiol, G., 2014. Text summarization in the biomedical domain: a systematic review of recent research. Journal of biomedical informatics, 52, pp. 457-467; Wang, M., Wang, M., Yu, F., Yang, Y., Walker, J. and Mostafa, J., 2021. A systematic review of automatic text summarization for biomedical literature and EHRs. Journal of the American Medical Informatics Association, 28(10), pp. 2287-2297.

[39] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B. and Kochut, K., 2017. Text summarization techniques: a brief survey. arXiv preprint arXiv:1707.02268.

[40] See e.g. Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of Research and Development, 2(2), pp. 159-165. https://doi.org/10.1147/rd.22.0159.

[41] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B. and Kochut, K., 2017. Text summarization techniques: a brief survey. arXiv preprint arXiv:1707.02268; Wojciech Kryscinski, Romain Paulus, Caiming Xiong, and Richard Socher. 2018. Improving abstraction in text summarization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp 1808-1817.

[42] A graph-based ranking model for text processing, used to identify keywords and the most relevant sentences.

[43] An unsupervised approach to text summarisation based on graph-based centrality scoring of sentences.

One of the key challenges in the use of abstractive summarisation for summarising specialist texts is the lack of large parallel corpora consisting of full texts and their summaries that could be used for training machine learning models. Additionally, in the case of data from criminal investigations, the principles of purpose limitation, data minimisation, data integrity and confidentiality (access rights to sensitive data) will have to be taken into account before starting collecting training data, and therefore prior anonymisation will be required. However, the results of recent experiments performed by Google engineers suggest that the application of certain techniques, including self-supervised learning, allows the development of abstractive summarisation systems with high-quality output for generic text, which can then be fine-tuned with a small number of examples[44]. This suggests that such methods could also potentially help develop well-performing abstractive summarisers in the future for domain-specific texts where large parallel corpora do not exist or are prohibitively expensive to generate.

3.1.4. NLP for evidence analysis and anonymisation

The processing of unstructured data, and therefore the application of NLP technologies, is relevant in the context of a wide range of criminal investigations, in particular relating to the prosecution of fiscal fraud (e.g. excise duty fraud or VAT fraud); financial fraud (e.g. insider trading, bankruptcy fraud, property fraud or money laundering); fraud relating to trade in specific goods (e.g. illicit drugs, intellectual property, weapons or sanctioned goods); and fraud on the internet and swindling. In all of the above, most of the data used as evidence will be unstructured, and would therefore require the use of techniques for processing unstructured data[45].

Until recently, investigations of white-collar crime would to a large extent be done manually with the support of e-Discovery[46] tools, which were based on conventional search algorithms (i.e. search by keywords). e-Discovery tools have been used for a relatively long time in the legal profession in order to automate – at least partially – the review process and therefore improve efficiency and effectiveness. With the widespread digitalisation of professional activities and the widespread use of digital communication technologies by criminals, in most cases, evidence collected in investigations of different types of the white-collar crime mentioned above include very large amounts of textual data (e.g. invoices, emails, contracts, shipping documentation, etc.). In such use cases, the use of other natural language processing techniques becomes essential.

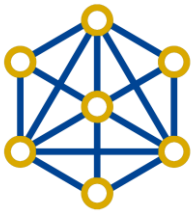
Investigators working on white-collar crime cases, such as tax fraud, for example, face a number of problems relating to dealing with big data analysis without appropriate tools in their toolset:

- conventional **search tools** result in an excessive number of search hits, which cannot be processed by human investigators;
- performing **queries on very large datasets** using conventional investigative software often results in significant **time delays** (up to 30 minutes per query) as well as possible software malfunction due to overload;
- **conventional** tools not being able to handle large datasets;
- and the resulting **large amounts of time** necessary to process the seized data.[47]

[44] https://ai.googleblog.com/2020/06/pegasus-state-of-art-model-for.html
[45] https://www.ibm.com/cloud/blog/structured-vs-unstructured-data
[46] Electronic Data Discovery used to facilitate search of information relevant to a specific law suit or investigation.
[47] van Banerveld, M., Le-Khac, N.A. and Kechadi, M.T., 2014, November. Performance evaluation of a natural language processing approach applied in white collar crime investigation. In International Conference on Future Data and Security Engineering (pp. 29-43). Springer: Cham.

A wide range of NLP techniques can be used to tackle some of the challenges identified above. For example, information extraction techniques such as named entity recognition in combination with relation extraction and network analysis can help identify relationships between certain companies, individuals or other named entities across large amounts of data.

Named entity recognition (NER) techniques can also be effectively combined with a graph-based clustering approach to identify relationships between the named entities[48]. Topic modelling and text clustering can also be helpful in identifying certain patterns in text, such as similar or topically related content, thus addressing the gap where a search by keywords does not always produce relevant results. Such approaches can be effectively applied in the context of a Case Management System in order to identify possible connections between different cases, as well as between different entities involved in these cases, thus providing a high-level overview of criminal networks, for example.



Identifying possible links between cases is also relevant for the cooperation between the different JHA agencies. For example, the legislation regulating the cooperation between Eurojust, Europol and the EPPO foresee that hit/no-hit systems should be used in order to exchange information between the agencies.[49] For the further development of these hit/no-hit systems, the NLP technology could be one of the tools to be considered in order to find more accurate links between cases.

At least in one instance, the implementation of natural language processing tools for investigative purposes suggested that they significantly outperform conventional forensic analysis software tools, and in some instances reduce the time necessary for data processing eleven-fold[50]. Similar claims with regard to efficiency gains in AI-assisted document review have been reported by some of the technology providers[51].

In addition to being used for investigative or forensic purposes, named entity recognition and classification (NERC) can be effectively used for document anonymisation or pseudonymisation. Document anonymisation is relevant in a variety of scenarios, such as anonymisation of datasets used for training of automated systems; anonymisation of court rulings prior to publication; and in those cases where anonymisation is required by law. In the context of cross-border collaboration between judicial authorities such tools are particularly relevant if they can be applied in multilingual scenarios.

One such multilingual anonymisation toolkit is currently being developed with the support of CEF Digital. The objective of the project is to support public administrations in EU Member States in complying with the GDPR. The toolkit, developed by a consortium of European companies, will support all official EU languages, including a range of under-resourced languages (Latvian, Lithuanian, Estonian, Slovenian and Croatian) and severely under-resourced languages (such

[48] Das, P., Das, A.K., Nayak, J., Pelusi, D. and Ding, W., 2019. A graph-based clustering approach for relation extraction from crime data. IEEE Access, 7, pp. 101269-101282.
[49] E.g as foreseen by art 49(1) and 50(5) of the regulation (EU) 2018/1727
[50] See Banerveld, M., et al., 2014.
[51] See e.g. https://www.luminance.com/technology.html

as Irish and Maltese). The project[52] will use natural language processing tools to create an open-source toolkit for reliable text de-identification, focusing specifically on the medical and legal fields. As argued by the authors of the project, the recent developments in deep learning architectures and the availability of large multilingual pre-trained models such as BERT[53] have allowed the performance in NERC[54] to be significantly improved. This has been further augmented by the transfer learning capabilities of such deep learning models, which allow new systems to be trained using smaller datasets of manually labelled data. Transfer learning allows the knowledge acquired in one domain or language to be used in cross-domain and cross-language settings, significantly increasing the efficiency and reducing the development costs of such models[55].

The MAPA toolkit developed by the project consortium will contain a general NERC model, which will be fine-tuned for domain-specific applications. This system can later be further tailored to domain-specific use cases. The deep-learning based NERC approach will be combined with other techniques where relevant, as for example in the case of pattern detection based on regular expressions in order to identify email addresses, ID or telephone numbers or bank accounts, etc. The MAPA toolkit will be publicly available as open-source software and will initially be specifically targeted at public administrations operating in the health and legal fields[56].

"NLP technologies can significantly improve the efficiency and quality of legal research by allowing legal professionals to search for relevant information across different platforms using natural language."

3.1.5. NLP for legal research and analysis

One of the areas where the value of technology has been recognised for a long time is legal research. Legal research supports legal decision-making by finding information relevant to a specific case, and normally involves the review of a statute or case-law. Technologies in support of legal research have existed since the early 1970s, when the Lexis database was developed in the US. Since then, the legal profession has been relying on a wide range of web-based databases such as LexisNexis, ThomsonReuters Westlaw and Bloomberg Law (all of the above focusing on the USA law) in performing legal research, which has helped to significantly increase the efficiency and quality of legal research[57]. Similar databases containing national and European legislation exist also in Europe. For example, N-Lex[58] serves as a gateway to national law of EU Member States, while Légifrance provides access to French law.

While the creation of legal databases has significantly improved access to relevant information, manual search still remains a very labour-intensive activity, which is also subject to errors and omissions due to the limitations of conventional search systems. NLP technologies can significantly improve the efficiency and quality of legal research by allowing legal professionals to search for relevant information, such as relevant statutes, applicable legislation, case-law or doctrinal opinion, using natural language, across different platforms.

[52] https://mapa-project.eu/
[53] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[54] Ajausks, E., Arranz, V., Bié, L., Cerdà-i-Cucó, A., Choukri, K., Cuadros, M., Degroote, H., Estela, A., Etchegoyhen, T., García-Martínez, M. and García-Pablos, A., 2020, November. The Multilingual Anonymisation Toolkit for Public Administrations (MAPA) Project. In Annual Conference of the European Association for Machine Translation (pp. 471-472).
[55] García-Pablos, A., Perez, N. and Cuadros, M., 2020. Sensitive data detection and classification in Spanish clinical text: Experiments with BERT. arXiv preprint arXiv:2003.03106.
[56] See Ajausks, E., et al., 2020.
[57] Dale, R., 2019. Law and word order: NLP in legal tech. Natural Language Engineering, 25(1), pp. 211-217.
[58] n-lex.europa.eu

One of the approaches that facilitates the use of NLP in developing effective tools for legal research is the application of knowledge graphs to legal information. A knowledge graph represents a collection of interconnected information entities with semantic types and properties in machine-readable form[59]. Knowledge graphs are at the core of many AI and NLP applications, including context-dependent recommendation systems, data analytics and information search, and are used by major search engines, such as Google, as well as a wide range of other major technology companies, including Facebook, LinkedIn, Microsoft and Amazon, etc. Knowledge graphs can effectively integrate legal information (e.g. case-law and legislation) from different countries and in different languages using the Resource Description Framework (RDF), specified by the W3C consortium[60]. The RDF framework allows data to be represented in graph structures, thus making it more connected and interoperable. This knowledge graph-based approach was applied in a Horizon 2020 funder innovation project Lynx[61] in order to develop services to facilitate compliance in multilingual and multijurisdictional scenarios[62].

Approaches such as those used in the Lynx project, relying on a micro-services architecture and integrating semantic services, document manager (key building block of the knowledge graph) and workflow manager, can be effectively applied to develop solutions for legal research using mostly open source technologies and relying on open data in most cases[63].

A range of tools are already available on the market; however, most of these tools do not offer integration of a multilingual and multinational legal knowledge base[64]. Therefore, for multinational and multilingual use-cases, the development of custom solutions may be required. There are some ongoing initiatives seeking to implement AI tools in the field of legal analysis also in Europe[65].

In the context of legal research, one of the challenges is also to link the relevant case-law with the legislation in force during the time of the court decision. One of the solutions here is to develop the tool in a way which would highlight in the result the version of applicable legislation that has been referred to in the case-law.

[59] Kroetsch, M. and Weikum, G. 2016. 'Special Issue on Knowledge Graphs'. Journal of Web Semantics, 37 (38): 53-54.
[60] World Wide Web Consortium.
[61] Lynx – Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe.
[62] Montiel-Ponsoda, E., and Rodríguez-Doncel, V. 2018. 'Lynx: Building the legal knowledge graph for smart compliance services in multilingual Europe'. In G. Rehm, V. RodríguezDoncel, and J. Moreno-Schneider (eds.) Proceedings of the 1st workshop on LREC (language resources and technologies for the legal knowledge graph) workshop, Miyazaki, Japan, pp. 19-22. http://lrec-conf.org/workshops/lrec2018/W22/pdf/book_of_proceedings.pdf #page=26.
[63] Moreno-Schneider, J., Rehm, G., Montiel-Ponsoda, E., Rodriguez-Doncel, V., Revenko, A., Karampatakis, S., Khvalchik, M., Sageder, C., Gracia, J. and Maganza, F., 2020. Orchestrating NLP services for the legal domain. arXiv preprint arXiv:2003.12900.
[64] Examples include: LexisNexis, ThomsonReuters Westlaw, Bloomberg Law (all mainly US-focused), LeReTo (AT & EU).
[65] See for example https://ec.europa.eu/info/sites/default/files/law/cross-border_cases/documents/transcription_webinar_18112021_-hans_suijkerbuijk.pdf

3.2. AI in forensic analysis and anonymisation of audiovisual media

3.2.1. Biometric recognition and forensic analysis

Despite the importance of textual evidence in criminal investigations, the available evidence is often not limited to text and contains a wide range of other media, including images, video and audio or voice media. Some of this data is recorded by CCTV cameras and is often of limited use for effective processing by computer vision systems in criminal investigations due to low quality[66]. In other cases, when the quality of video or image data is relatively high, computer vision systems can be a valuable tool in identifying victims in recorded or live-streamed media. To tackle the video quality issue, dedicated video/image enhancing algorithms have been developed[67]. It is important to mention here that this analysis does not concern any live biometric recognition and the issues that are addressed in the joint opinion of the EDPS-EDPB, where they call for a general ban on any use of AI for an automated recognition of human features in publicly accessible spaces.[68]

Prior to the introduction of computer vision systems relying on machine learning models, video and image media were mainly processed by law enforcement officers or investigators. The human processing of such data has a number of limitations:

- the **manual analysis of audiovisual media** is a very resource-intensive, requiring the allocation of large numbers of staff with specific expertise;
- the **manual processing** is often not effective due to the fact that humans are prone to errors (e.g. due to fatigue);
- the **manual analysis of media containing child sexual exploitation** and other types of violence are particularly challenging, resulting in significant psychological toll. This means that additional measures need to be put in place in order to address the psychological pressure, including the hiring of additional staff.

In that sense, machines are a nearly perfect replacement, as they can tirelessly perform such tasks, consuming relatively limited amounts of resources and in a significantly smaller amount of time. However, this does not mean that machines are not prone to error. Today, it is a widely accepted fact that machine learning algorithms, such as those used in different classification tasks, are only robust when used in narrowly defined applications (not easily generalisable beyond those), and are only as good as the data they are trained on (i.e. garbage in, garbage out). An increasing number of studies have proved that algorithmic systems are prone to gender, racial and age-related bias[69]. This can be due to the way the machine learning (ML) algorithm is constructed, as well as due to the biases contained in datasets used to train the ML models.

To tackle these challenges, in recent years significant work has been done in understanding sources of bias in biometric recognition systems. This work has been done by both academic researchers and industry to develop ways to improve performance of biometric recognition systems, for instance by improving training datasets, deploying effective performance evaluation measures with a specific focus on bias, or using novel approaches, such as de-biasing adversarial networks[70].

[66] Senan, M.F.E.M., Abdullah, S.N.H.S., Kharudin, W.M. and Saupi, N.A.M., 2017, January. CCTV quality assessment for forensics facial recognition analysis. In 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (pp. 649-655). IEEE.
[67] Xiao, J., Li, S. and Xu, Q., 2019. Video-based evidence analysis and extraction in digital forensic investigation. IEEE Access, 7, pp. 55432-55442.
[68] EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)
[69] Rathgeb, C., Drozdowski, P., Damer, N., Frings, D.C. and Busch, C., 2021. Demographic Fairness in Biometric Systems: What do the Experts say?. arXiv preprint arXiv:2105.14844' Fundamental Rights Agency, 2019. Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights.
[70] Gong, S., Liu, X. and Jain, A.K., 2020, August. Jointly de-biasing face recognition and demographic attribute estimation. In European Conference on Computer Vision (pp. 330-347). Springer, Cham.

The basic enabling technologies in the forensic analysis of visual media (images and video) are exactly the same as those used for facial recognition technology by border control authorities or law enforcement. Therefore, biometric identification algorithms can be relatively easily implemented in forensic video and image analysis systems. Similar technologies, but trained on datasets containing objects instead of faces can be used to identify specific objects, which can also be relevant for investigative purposes. In addition, similar technologies can be applied to assessing the authenticity of video recordings or images (e.g. identification of deep fakes).

"Only systems relying on stable, reliable and well-understood algorithms can be deployed, even if the current state-of-the-art systems based on deep neural networks may be more effective in terms of their performance."

Other biometric modalities, such as gait, can also be used where facial identification is not possible (e.g. due to low resolution)[71]. Although gait recognition today is not very precise, it can still be used to determine a probability if a person appearing on a video has appeared elsewhere; however, the results of using such behavioural biometrics as gait should be treated with caution, considering their relatively low precision.

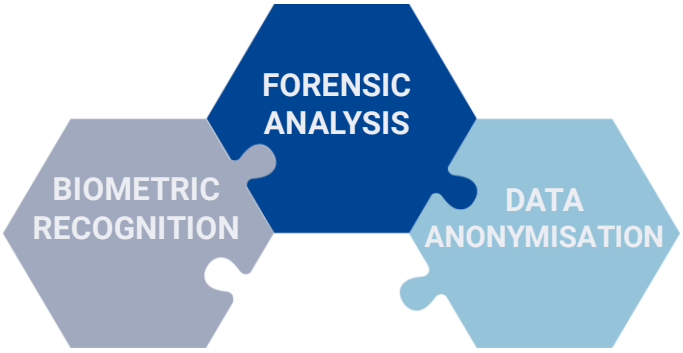
In order to enhance accuracy of biometric recognition systems, multimodal recognition methods can be deployed, which may be relevant in forensic analysis. For example, facial recognition in video analysis can be combined with voice recognition from the audio stream, further enhancing the capabilities of forensic analysis systems to identify specific individuals, whether they are criminals or victims of criminal activities. Voice or speaker recognition technologies have been used in forensics since the late 1990s when a Gaussian Mixture Model-Universal Background Model speaker identification system was developed[72]. Since then, voice or speaker recognition technologies have developed significantly, in particular by incorporating deep learning methods, which today are the state-of-the-art in speaker verification and identification[73]. The introduction of deep learning approaches in developing speaker recognition technologies allowed some of the challenges to be addressed that could not be addressed with earlier methods, such as noise or reverberation[74].

As is the case with facial recognition systems[75], the standardisation and evaluation of speaker recognition technologies is an important component in developing reliable systems. The National Institute of Standards and Technology (NIST)in the US has been carrying out Speaker Recognition Evaluations since 1996. The purpose of this international evaluation is to measure the state-of-the-art in technology and identify the most effective algorithmic approaches[76]. One of the particular challenges of using speaker recognition systems in forensic work is that audio is often recorded in an uncontrolled environment (e.g. background noise), which makes it especially challenging for automated systems. Standardisation and accreditation of the systems used in forensic applications is especially important to ensure that the system is robust and suitable for application in the conditions in which it will be used. This means that only systems relying on stable, reliable and well-understood algorithms can be deployed, even if the current state-of-the-art systems based on deep neural networks may be more effective in terms of their performance[77].

[71] M. Almeida, P. L. Correia and P. K. Larsen, 'BioFoV – an open platform for forensic video analysis and biometric data extraction', 2016 4th International Conference on Biometrics and Forensics (IWBF), 2016, pp. 1-6, doi: 10.1109/IWBF.2016.7449693.
[72] See e.g. Reynolds, D.A., Quatieri, T.F. and Dunn, R.B., 2000. Speaker verification using adapted Gaussian mixture models. Digital signal processing, 10(1-3), pp. 19-41.
[73] Sztahó, D., Szaszák, G. and Beke, A., 2019. Deep learning methods in speaker recognition: a review. arXiv preprint arXiv:1911.06615.
[74] Bai, Z. and Zhang, X.L., 2021. Speaker recognition based on deep learning: An overview. Neural Networks.
[75] See NIST Face Recognition Vendor Test https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt
[76] See NIST Speaker Recognition Evaluation: https://sre.nist.gov/
[77] Kockmann, M., Farrell, K., Colibro, D., Vair, C., Alexander, A. and F. Kelly (2021) 'Voice biometrics: perspective from the industry' in García-Mateo, C. and G. Chollet (Eds) *Voice Biometrics: Technology, trust and security*. The Institute of Engineering and Technology: London.

3.2.2. Anonymisation of visual data

Considering the sensitive nature of biometric data, and in particular of facial images, particular attention needs to be dedicated to the way this data is processed and used, to ensure compliance with the relevant legal acts, such as the GDPR, LED, Regulation 2018/1725. One of the well-recognised challenges in developing well-performing biometric recognition systems is the bias embedded in the matching algorithms[78].



▲ *Figure 2: AI applications in audiovisual data processing*

Addressing this challenge requires the use of training and testing datasets that are representative of the population to which the biometric recognition system will eventually be applied. However, creating such datasets is not trivial. First, it is expensive, as such datasets need to contain very large amounts of images that are correctly labelled. Images for such datasets cannot be simply scraped off the internet, as in this case the dataset will likely not be representative. Second, creation of such datasets will be legally challenging, in particular from the perspective of privacy and personal data protection.

Recent developments in the use of generative adversarial networks (GANs), however, suggest that there may potentially be a possibility to create biometric matching algorithms where personal data protection and privacy are ensured by default, which of course does not mean that the data protection aspects would not need to be carefully assessed. First, GANs can be used for deep anonymisation or de-identification of facial images, which can then be used for training and testing of biometric matching algorithms[79]. Contrary to the anonymisation techniques used previously (e.g. blurring or pixilation), anonymisation techniques based on GANs allow for the preservation of the utility of image data (e.g. as training datasets), while at the same time ensuring privacy protection. Second, GANs can be used to generate synthetic biometric data for algorithm training and testing[80]. Nevertheless, the accuracy of biometric recognition systems trained on synthetic data, when applied to the real-world environment, is insufficient. Therefore, in those use cases where biometric recognition systems are used for the identification of persons with potential legal consequence, such systems need to be developed using real representative datasets.

In addition to using biometric anonymisation techniques for training and testing datasets, similar techniques can be used to effectively anonymise images or video materials containing biometric identifiers (i.e. human faces). Such applications may be relevant for anonymisation of evidence to protect victims, for example. Similar solutions can also be used for removing other identifiers from visual media, such as license plate numbers from cars, where this is relevant. Similarly to fingerprints and facial images, voice can also be used as a biometric identifier. Like other biometric identifiers, voice suffers from similar constraints, in particular as regards the need to preserve privacy and protect personal data. As described before in the context of facial images, anonymisation techniques can also be used in order to remove identifying characteristics from voice recordings[81]. Such techniques can be useful when it is necessary to anonymise audio evidence (e.g. recorded telephone conversations) to protect victims or witnesses.

---

[78] See, for example, the report of the National Institute of Standards and Technology 'Face Recognition Vendor Test (FRVT). Part 3: Demographic Effects', available here: https://doi.org/10.6028/NIST.IR.8280. See also Pawel Drozdowski, Christian Rathgeb, Antitza Dantcheva, Naser Damer, Christoph Busch. 2020. Demographic Bias in Biometrics: A Survey on an Emerging Challenge. IEEE Transactions on Technology and Society, IEEE doi: 10.1109/TTS.2020.2992344.

[79] See e.g. Hukkelås, H., Mester, R. and Lindseth, F., 2019, October. Deepprivacy: A generative adversarial network for face anonymization. In International Symposium on Visual Computing (pp. 565-578). Springer, Cham.; Undirwade, A. and Das, S., 2021. Image Anonymization Using Deep Convolutional Generative Adversarial Network. Machine Learning Algorithms and Applications, pp.305-330.; Lindseth, E., 2020. An Encoder-Decoder based Deep Learning Approach for Anonymization of Visual Surveillance Media with Preservation of Utility (Master's thesis, University of Agder).

[80] See e.g. Murphy, T.M., Broussard, R., Rakvic, R., Ngo, H., Ives, R.W., Schultz, R. and Aguayo, J.T., 2016. Use of synthetic data to test biometric algorithms. Journal of Electronic Imaging, 25(4) and for more recent developments Zhang, H., Grimmer, M., Ramachandra, R., Raja, K. and Busch, C., 2021. On the Applicability of Synthetic Data for Face Recognition. arXiv preprint arXiv:2104.02815.

[81] See e.g. the recently launched VoicePrivacy initiative, focusing on developing and evaluating privacy-preserving solutions for speech technology; see also Yoo, I.C., Lee, K., Leem, S., Oh, H., Ko, B. and Yook, D., 2020. Speaker Anonymization for Personal Information Protection Using Voice Conversion Techniques. IEEE Access, 8, pp.198637-198645.

CONCLUSIONS

The field of justice is undergoing digital transformation, and artificial intelligence, as a set of different technologies, has great potential to contribute to and further enhance this process, allowing for a significant improvement in both the efficiency and effectiveness of operation of the judicial authorities. Similar to digitalisation, the use of AI may, in the long term, help significantly reduce the costs for the judicial authorities. Efficiency gains, improvements in effectiveness, cost reduction: all of these performance improvements can eventually result in improved access to justice and reduced time to render judicial decisions. These benefits, however, will largely depend on the robustness and reliability of the tools deployed. As one of the well-known AI applications – the now infamous SyRi system used by the Dutch government to detect welfare fraud[82] – has shown, blackbox algorithms that are neither robust nor reliable, lack adequate human oversight and where the users are overreliant on the output of the system, can have significant implications, including severe harm for the affected citizens and diminished trust in government. Therefore, a balanced approach is necessary in order to ensure the protection of fundamental rights while enhancing the digital transformation.

"Prior to the implementation of AI technologies in practice, authorities considering such applications should perform a robust risk assessment, as well as assess the legal, ethical and fundamental rights implications of the use of such technologies."

Most of the technologies that have potential for a substantive impact in the field of justice, such as NLP and biometric recognition, are already available on the market. Nevertheless, the implementation of such solutions in the judicial field, cannot be considered as a purely technical challenge. As argued in the report, some of the applications of AI in the field of justice may have significant legal, ethical and fundamental rights implications. The use of certain techniques is already called to be banned by the privacy and data protection regulators[83]. In such cases, the development and deployment of AI-driven solutions should be evaluated with special care in order to avoid situations where fundamental rights and freedoms of individuals are negatively affected and persons are subjected to various forms of discrimination, limitations of rights, unfair treatment, etc. For this purpose, a considerable amount of legal and ethical research has been conducted in order to analyse the potential impact of AI tools, both in private and public spheres and this research provides input into the analysis of possible AI use cases in the field of justice. There are, however, numerous use-cases where AI can be applied effectively and deliver significant gains without major ethical or legal implications. In this report, we have tried to focus specifically on such use-cases in order to show where such technologies can deliver benefits already in the near term.

In most cases, such technologies are already available on the market, but may need adaptation to the specific context. However, prior to the implementation of these technologies in practice, authorities considering such applications should perform a robust risk assessment, as well as assess the legal, ethical and fundamental rights implications of the use of such technologies. It is also important to conduct the testing and evaluation of such technologies in real-life conditions, in order to ensure that their performance meets the relevant standards, in particular as regards accuracy and bias or any kind of discrimination. Therefore, we suggest that a risk-based approach (integrating cost benefit analysis, fundamental rights and data protection assessments) shall be developed and applied to determine the safeguards and the deployment model for AI systems in practice.

[82] https://iapp.org/news/a/digital-welfare-fraud-detection-and-the-dutch-syri-judgment/

[83] EDPB and the EDPS call for a general ban on any use of AI for an automated recognition of human features in publicly accessible spaces - such as of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioural signals - in any context.

RELEVANT RESEARCH AND INNOVATION PROJECTS FUNDED BY THE EU (NON-EXHAUSTIVE LIST)

THE ROXANNE PROJECT: although the project focuses on **developing solutions for law enforcement authorities**, they may potentially be of relevance in criminal investigations conducted by the judicial authorities.

THE LOCARD PROJECT aims **to develop a solution to automate the collection of digital evidence.** The project's goal is also to provide a comprehensive management approach to the handling of digital evidence to be presented in a court of law, thus overcoming some of the current issues in this process. Specifically, LOCARD aims to increase trust in the handling and processing of digital evidence and the management of chain of custody by providing transparency and using an immutable chain of custody stored using blockchain technology.

THE COMPRISE project defines a fully private-by-design methodology and tools that will reduce the cost and increase the inclusiveness of voice interaction technology through research advances on privacy-driven data transformations, personalised learning, automatic labelling and integrated translation. This leads to a holistic easy-to-use software development kit interoperating with a cloud-based resource platform.

THE LYNX PROJECT provided effective ways of accessing huge amounts of digital regulatory compliance documents, including legislation, case-law, standards, industry norms and best practices. In particular, this solution envisages an ecosystem of smart cloud services to better manage compliance documents, based on a Legal Knowledge Graph, which integrates and links heterogeneous compliance data sources. This ecosystem enables smart search, smart assistance and smart referencing of case-law, as well as Artificial Intelligence technologies and the machine translation of regulatory compliance documents.

THE MANYLAWS PROJECT'S overall objective is to enable access to legal information across the European Union and improve the efficacy of decision-making in legislative procedures operated by public bodies. Specifically, the Action will set-up a platform (ManyLaws) which will deliver a set of services for citizens, businesses and administrations built upon text mining, advanced processing and semantic analysis of laws of the European Union, Austria, and Greece. The addition of laws from other Member States will be explored and, if possible, implemented. Moreover, these services will be tested within at least two law-making procedures of the Greek and Austrian Parliaments.

DEFINITIONS

ALGORITHM – an unambiguous specification of how to solve a particular problem.

ARTIFICIAL INTELLIGENCE SYSTEM (AI OR AI SYSTEM) – software that is developed with one or more of the techniques listed in Annex I of the AI Act and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations or decisions influencing the environments they interact with[84].

APPLICATION PROGRAMMING INTERFACE – a connection between computers or between computer programs. It is a type of software interface, offering a service to other pieces of software.

COMPUTER VISION – enables machines to analyse, understand and manipulate video and image media using machine learning models and other techniques[85].

DEEP LEARNING – an area of machine learning that attempts to mimic the activity in layers of neurons in the brain to learn how to recognise complex patterns in data. The 'deep' in deep learning refers to the large number of layers of neurons in contemporary ML models that help to learn rich representations of data to achieve better performance gains.

GENERATIVE ADVERSARIAL NETWORK – a class of machine learning framework. Given a training set, this technique learns to generate new data with the same statistics as the training set. For example, a GAN trained on photographs can generate new photographs that look at least superficially authentic to human observers, having many realistic characteristics.

MACHINE LEARNING – a subset of AI that often relies on statistical techniques giving machines the ability to learn from data without being explicitly given the instructions on how to do so. This process is known as training a model using a learning algorithm that progressively improves model performance on a specific task.

ML MODEL – the output of the process of training an ML algorithm on data. ML models can be used to make predictions.

NATURAL LANGUAGE PROCESSING – enables machines to analyse, understand and manipulate human language using machine learning models and other techniques.

NATURAL LANGUAGE GENERATION – allows machines to respond in natural language – either text or speech – providing information on the basis of a certain dataset.

NEURAL MACHINE TRANSLATION – an approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modelling entire sentences in a single integrated model.

PROOF-OF-CONCEPT – a demonstration of the practical feasibility of an idea to prove that the method, idea, or technology works.

PILOT PROJECT – an implementation of a technology that has been tested in a PoC in real-life conditions, but on a limited scale, in order to test the performance of a technology in real-life conditions.

SUPERVISED LEARNING – the most common type of ML algorithm currently in use. In supervised learning, a model attempts to learn to transform one kind of data into another kind of data using labelled examples.

UNSUPERVISED LEARNING – a process in which a model attempts to learn a dataset's structure often seeking to identify latent groupings in the data without any explicit labels. The output of unsupervised learning is often used as the input for supervised learning algorithms.

[84] COM/2021/206 final.
[85] Definitions provided above are based on definitions included in the State of AI Report 2021 (stateof.ai).

Publications Office
of the European Union

ISBN 978-92-95227-17-0
ISSN 2022.2822
doi: 10.2857/364146
Catalogue number: EL-09-22-215-EN-N

eu-LISA

EUROJUST