

Integrating Human-in-the-Loop AI Systems in the Italian Judicial System

Samuele Yves Cerini^{1,2,*,†}, Nicola Deidda^{1,3,*,†}, Sara Lilli^{1,4,*,†}, Maria Vittoria Zucca^{1,4,*,†},
Giorgio Fumera³, Giorgio Giacinto^{2,3}, Paolo Prinetto^{2,*} and Salvatore Marsocci⁵

¹Scuola IMT Alti Studi Lucca, Piazza S. Ponziano 6, 55100 Lucca, LU, Italy

²CINI Cybersecurity National Lab, Via Ariosto 25, 00185 Roma, RM, Italy

³Università di Cagliari, Via Università 40, 09124 Cagliari, CA, Italy

⁴Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33, 56127 Pisa, PI, Italy

⁵Procura della Repubblica presso il Tribunale di Torino, Corso Vittorio Emanuele II 130, 10138 Torino, TO, Italy

Abstract

Despite ongoing efforts to digitize the Italian judicial system, challenges such as high case volumes, inconsistent digitalization and inefficient information management persist, in particular within Public Prosecutor's Offices. This research examines the potential of integrating Human-in-the-Loop Artificial Intelligence (AI) systems into the Italian criminal justice system, aiming to support prosecutors by automating repetitive tasks, improving document management and enhancing workflow efficiency. Three practical areas of concern and their respective potential solutions are identified: (i) the classification and registration of crime reports, (ii) the transcription of wiretaps, and (iii) structured data storage for analysis and crime prevention. Although rooted in the Italian national system, the research proposes technical solutions that could be transferable to other jurisdictions with similar prosecutorial frameworks.

Keywords

Italian Justice System, Artificial Intelligence, Crime Prevention, Human-in-the-Loop

1. Introduction

Within contemporary democratic legal systems, criminal law occupies a foundational role, as it safeguards individual freedom and security and constitutes the primary instrument through which the State protects the fundamental interests of the community, while maintaining a proper balance between crime prevention and the protection of individual rights. In the Italian context, this normative and institutional function takes concrete form in a criminal justice system structured as a network of entities, within which Public Prosecutors' Offices represent the first operational level of criminal proceedings, insofar as they receive crime reports (*notitiae criminum*) and carry out the preliminary organisation of investigative activities.

In recent years, several efforts have been undertaken to digitalise and modernise the justice system [1], with the aim of improving efficiency in case management and of reducing reliance on paper-based processes. Nevertheless, the system continues to face significant challenges due to a combination of structural and organisational factors. In particular, the high volume of cases, the continuous regulatory changes and the uneven levels of digitalisation across offices are increasingly hindering daily operations. These difficulties directly affect the efficiency of operational workflows and the ability of offices to process incoming information in a timely and effective manner.

Joint National Conference on Cybersecurity (ITASEC & SERICS 2026), February 09-13, 2026, Cagliari, IT

*Corresponding author

†These authors contributed equally

✉ samuele.cerini@imtlucca.it (S. Y. Cerini); nicola.deidda@imtlucca.it (N. Deidda); sara.lilli@santannapisa.it (S. Lilli); maria.zucca@santannapisa.it (M. V. Zucca); giorgio.fumera@unica.it (G. Fumera); giorgio.giacinto@unica.it (G. Giacinto); paolo.prinetto@cybersecnatlab.it (P. Prinetto); salvatore.marsocci@giustizia.it (S. Marsocci)

ORCID 0009-0001-4961-8915 (S. Y. Cerini); 0009-0005-7309-5703 (N. Deidda); 0009-0000-2796-3067 (S. Lilli); 0009-0004-0049-9611 (M. V. Zucca); 0000-0001-5300-226X (G. Fumera); 0000-0002-5759-3017 (G. Giacinto); 0000-0003-2400-8245 (P. Prinetto)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The consequences of such inefficiency are both numerous and far-reaching. In particular, they impact information management, which constitutes one of the most critical components of prosecutorial activity. A notable proportion of the incoming documentation consists of unstructured data, including PDF files, images, or scanned documents, which lack a standardised format suitable for automated processing. This operational context restricts access to information, hampers the correlation of related proceedings, and increases the risk of information loss or fragmentation.

One of the main critical areas within the informational domain concerns the management of *notitiae criminum*, namely criminal reports submitted to Public Prosecutors' Offices in the form of complaints, allegations, and communications from law enforcement agencies (LEAs) [2]. The registration and classification of such reports remain largely manual processes and require complex interpretative activities by office personnel, with evident negative impacts on processing time and information quality. In this context, three fundamental operational challenges emerge: (i) a marked imbalance between the volume of *notitiae criminum* received on a daily basis and the actual processing capacity of prosecutorial offices; (ii) the potentially inadequate information quality, from both a structural and a semantic perspective, due to the absence of standardised formats, heterogeneity in document drafting practices, and variability in the completeness and consistency of content; and (iii) the lack of automation tools to support the aggregated analysis of judicial workflows, which are essential for assessing workload distribution and understanding crime dynamics.

These critical issues can be further substantiated by examining the functioning of a medium-to-large Public Prosecutor's Office as a representative example of the operational "bottlenecks" within the Italian criminal justice system, characterised by informational pressure and labour-intensive workflows. In this context, the Public Prosecutor's Office of Turin (Piedmont, Northern Italy) stands out as one of the largest judicial offices in the country, as outlined by classifications issued by the Ministry of Justice, based on structural and organisational criteria [3]. Observational data drawn from the "Transparent Administration" system delineate a significant operational scenario for the Public Prosecutor's Office of Turin [4]: approximately 540 *notitiae criminum* per day, amounting to more than 200,000 reports annually. Of these, approximately 150,000 concern proceedings against unknown suspects, while around 50,000 relate to identified individuals. Each crime report typically consists of multiple pages, often digitised from paper-based or handwritten documents, and is processed on a daily basis by a limited number of operative personnel. In such context, one of the primary objectives is to ensure that each criminal report is assigned to the appropriate group of prosecutors in the shortest possible time and with the highest level of dependability. In this regard, AI-based systems are proposed as enabling technologies capable of automating specific stages of the workflow, improving document management, supporting preliminary analysis and facilitating a more efficient organisation of judicial information. As a result, the present papers proposes a *Human-in-the-Loop* paradigm, according to which Artificial Intelligence (AI) does not replace prosecutors or judicial officials, but rather assists them in informational and organisational tasks.

In light of the aforementioned context, the present research study examines a series of use cases, each of which is representative of the activities carried out during the so-called *preliminary investigation* phase. In this phase, AI can intervene as a support tool, automating tasks that are often repetitive, albeit simple, yet nevertheless inevitably necessary for the proper management of judicial organizational processes. In particular, the analysis is structured along three main directions: (i) the legal classification and registration of *notitiae criminum*; (ii) the automated transcription of wiretaps and intercepted communications; (iii) the structured storage of information and its exploitation for purposes of statistical analysis and crime prevention. While these case studies are grounded in the specific features of the Italian justice legal system, they outline a technical and organizational paradigm that is potentially transferable to other jurisdictions, characterised by comparable prosecution systems.

2. Regulatory and Technical Foundations

2.1. Legal Framework

The introduction of AI systems within the administration of justice in Italy takes place within a complex and multilayered regulatory framework, shaped by national reforms on digitalisation, constitutional guarantees, personal data protection obligations, and emerging standards at the level of the European Union.

At the domestic level, the digital transformation of criminal proceedings has undergone a significant acceleration in recent years, particularly following the COVID-19 pandemic, which made it necessary to introduce telematic and digital tools to ensure the continuity of judicial activity. Decree-Laws No. 137/2020 and No. 44/2021 introduced transitional provisions aimed at expanding the use of digital technologies during the preliminary investigation phase and in communications between judicial offices, parties, and law enforcement agencies (LEAs). On a structural level, the “Cartabia Reform” (Law No. 134 of 27 September 2021) consolidates the Telematic Criminal Trial (*Processo Penale Telematico*, PPT) as the ordinary infrastructure for the filing, transmission, and preservation of procedural and investigative documents. The reform assigns ministerial information systems – among them the Criminal Case Information System (*Sistema Informativo della Cognizione Penale*, SICP) and the TIAP platform used by Public Prosecutors’ Offices – the role of official and exclusive repositories of judicial information, requiring that document flows be managed in digital form according to principles of authenticity, integrity, and traceability. An essential component of the national regulatory landscape is the recent “*Recommendations on the Use of Artificial Intelligence in the Administration of Justice*”, adopted by the High Council for the Judiciary (CSM) on 8 October 2025, which represent the first comprehensive institutional position on the topic [5]. The CSM acknowledges the potential of AI to support legal research, document analysis, and the management of informational workflows, while also emphasizing risks concerning data protection, algorithmic transparency, biases, and the variability of generated outputs. For this reason, CSM stresses the need to safeguard judicial autonomy, ensure the verifiability of AI outputs, and maintain continuous human supervision – requirements aligned with Articles 24 and 111 of the Italian Constitution, which guarantee the right to defense and due process.

A key contribution of the Recommendations is the introduction of a clear distinction that shapes the Italian model of AI integration:

- **Judicial activities “*stricto sensu*”**, where AI systems may not be used unless they comply with the high-risk requirements established by the EU Artificial Intelligence Act (Regulation 2024/1689) and where AI cannot influence the judge’s reasoning or the assessment of facts and law;
- **Organisational, administrative or procedural activities**, where AI may be employed as a supportive tool, provided it operates in a secure, traceable manner and under human supervision. This category includes tasks such as document summarisation, thematic classification of publicly accessible decisions, statistical reporting, registry analysis, automated document comparison, scheduling and routine low-complexity activities.

The CSM further highlights the significant organizational heterogeneity across judicial offices, which affects data quality and underscores the need for technological architectures ensuring reliability, non-discrimination, and auditability of AI systems. In general, the national framework requires full compliance with the data protection law (GDPR and Legislative Decree 196/2003) and with the EU Artificial Intelligence Act (Regulation 2024/1689). Beginning in 2026, magistrates will be allowed to use only certified and registered high-risk AI systems. For this reason, the Recommendations urge the development of an internal AI platform for the justice domain, capable of ensuring data sovereignty, cybersecurity, and full traceability. Within this regulatory structure, AI is allowed exclusively as *Human-in-the-Loop* technology, assisting preparatory and organizational tasks, without replacing or influencing judicial decision-making.

Transposing the discussion to the supranational regulatory framework, the pertinent provisions can be found in Article 6 and Annex III of the Artificial Intelligence Act (Regulation EU 2024/1689) [6].

Article 6, “*Classification rules for high-risk AI systems*”, states in paragraph 2 that “AI systems listed in Annex III are also considered high-risk”. Annex III, in turn, defines high-risk AI systems as those operating in the “administration of justice and democratic processes” sector, providing the definition in point 8: “AI systems intended to be used by or on behalf of a judicial authority to assist in the identification and interpretation of facts and law and in applying the law to a specific set of facts, or to be used similarly in alternative dispute resolution”. In paragraph 3, however, the same Article 6 introduces exceptions to the high-risk classification in the sectors listed in Annex III. These exceptions use broad formulations: “[...] An AI system listed in Annex III is not considered high-risk if it does not present a significant risk of harm to the health, safety, or fundamental rights of natural persons, including the lack of material influence on the outcome of the decision-making process.” It further specifies that: “Paragraph 1 applies when at least one of the following conditions is met: (a) the AI system is intended to perform a limited procedural task; (b) the AI system is intended to improve the result of a previously completed human activity; (c) the AI system is intended to detect decision patterns or deviations from previous decision patterns and is not designed to replace or influence the previously completed human assessment without adequate human review; or (d) the AI system is intended to perform a preparatory task for an assessment relevant to the use cases listed in Annex III”. However, there remains a need for clarification. In particular, the notion of a “preparatory task” is open to multiple interpretations, especially concerning the presence-absence of human review. Similarly, the concept of a “limited procedural task” may refer to tasks that are purely statistical or organizational in nature.

European courts have also intervened on this matter. The Court of Justice of the European Union (CJEU) approved the “*Artificial Intelligence Strategy*” in 2023, introducing the concept of “smart courts” within the e-Justice domain [7]. Smart courts utilize technologies like AI to improve efficiency, transparency, and accessibility in judicial processes (e.g., case management systems, the analysis of judicial documents and AI-powered assistants that support routine tasks such as scheduling and document preparation). Relatedly, the “*European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*”, developed in 2018 by the EU Commission for the Efficiency of Justice (CEPEJ) outlines five key principles: (i) respect for fundamental rights; (ii) non-discrimination; (iii) quality and security; (iv) transparency, impartiality, fairness; and (v) user control [8]. These principles offer a valuable reference for national legal systems, ensuring ethical AI integration while upholding human rights and fairness. In this context, from a legal-comparative perspective, the evolving frameworks for the use of AI in judicial systems across several European jurisdictions can be summarized as depicted in Table 1. These frameworks illustrate how different countries are structuring governance mechanisms to promote the responsible integration of AI in courts while addressing legal, ethical, and operational considerations.

2.2. Technical Framework

Although AI presents an opportunity to enhance judicial workflows, its deployment is constrained by structural, legal and cultural challenges that must be explicitly acknowledged and addressed, as these are not merely implementation details but involve constitutional values (e.g., fair trial under Art. 111, judicial independence under Art. 104 of the Italian Constitution), the architecture of the Italian justice system, and the technical characteristics of judicial data and workflows.

From a technical perspective, most AI methods assume, as their input, the use of large volumes of relatively homogenous, well-annotated data. Judicial data in Italy is typically fragmented across multiple case management systems, courts, and offices, and is often produced over decades using changing templates and software, with many records locked in legacy formats or scanned images. This heterogeneity complicates the construction of reliable training sets and undermines the portability of AI models across different jurisdictions or courts. Furthermore, the confidentiality and security requirements that apply to criminal case files limit the possibility of centralizing data or utilizing cloud-based infrastructures without robust anonymization and governance. AI systems inevitably reproduce the statistical structure of the data on which they are trained. In the judicial domain, those data embody historical inequalities, local cultures, and resource constraints. Training a model on decisions from a

Table 1

European governance frameworks for AI integration in judicial systems

Jurisdiction	Governance Framework
France	The <i>Tribunal des activités économiques de Paris</i> , on September 2025, adopted the “ <i>Charte d’utilisation de l’intelligence artificielle au service de la justice</i> ”, affirming that: “Artificial intelligence is considered and used as a complementary tool to assist in decision-making, not as a substitute for human decision-making” [9]. Afterward, the <i>Court de Cassation</i> , in April 2025, adopted the report “ <i>L’Intelligence artificielle au service de la justice</i> ” [10], which sets out guiding principles and prerequisites for AI projects in the judicial field.
Netherlands	The <i>Raad voor de Rechtspraak</i> in the Netherlands adopted the “ <i>Artificial Intelligence Strategy for the Dutch judicial system</i> ” which defines acceptable use cases, temporarily excludes high-risk applications (“The Dutch judicial system considers that the use of AI could be very beneficial in low-risk proceedings and therefore focuses on this use”) until appropriate safeguards are in place. The strategy also mandates human oversight and includes a 10-point plan to guide the responsible adoption of AI [11].
Spain	Since 2024, the <i>Comité Técnico Estatal de la Administración Judicial Electrónica</i> (CTEAJE) has implemented the “ <i>Política de uso de la IA en la Administración de Justicia</i> ”, which plan a national governance framework for AI use within the judicial sector [12]. This policy, aligned with Royal Decree-Law 6/2023, focuses on ensuring ethical, legal, and responsible AI applications, while safeguarding judicial independence and data protection.
Germany	The <i>E-Justice-Rat / Bund-Länder</i> is finalizing a coordinated national strategy for the use of AI in the judicial sector. The <i>E-Justice Council</i> notes: “particularly because the potential for discrimination associated with AI is not always obvious, I would like to draw your attention especially to making the use of AI exclusively free of any form of discrimination” [13].
United Kingdom	In October 2025, the UK Courts and Tribunals Judiciary issued “ <i>Artificial Intelligence Judicial Guidance for Judicial Office Holders</i> ”. [14] It points out the need for judicial personnel to understand the capabilities and limitations of AI. Indeed, the guidance advocates for responsible AI use, particularly for administrative tasks, while cautioning against relying on AI for legal research or analysis/reasoning without independent verification.

subset of courts may encode biases, which can then be propagated when the model is used elsewhere. Even seemingly neutral tasks such as *notitiae criminum* classification and prioritization of possible procedural paths may have disparate impacts on vulnerable groups if the underlying data reflect unequal policing or reporting patterns. In the judicial context, the requirement that decisions be reasoned and subject to scrutiny extends, at least normatively, to any technological tools that influence them. Models that cannot provide intelligible explanations for their outputs sit uneasily with this requirement. Yet, simpler models that are more explainable may be less accurate or less flexible in capturing complex legal patterns. This creates a structural tension between performance and transparency. Moreover, it is not sufficient that the system be explainable to technical experts; explanations must be usable by parties who might question the role played by AI in their case. Designing interfaces and documentation that enable genuine contestability, rather than post hoc rationalizations, is a major open challenge.

Additionally, it is important to notice that AI components are not intended to operate in complete isolation. These systems must be integrated with the existing case management systems and document repositories used by courts and prosecutors’ offices. Within the Italian judicial system, such systems often exhibit heterogeneity across different jurisdictions and court levels, and may not provide standard interfaces for data access or real-time interaction. Introducing AI support (e.g., for intelligent search, document classification, or workflow optimization) therefore demands significant work on interoperability, data schemas, and process redesign. Without this integration layer, AI tools risk remaining experimental prototypes disconnected from everyday practice.

Nevertheless, judicial decision-making is a deeply professional practice, shaped by legal education, shared interpretative traditions, and strong notions of independence and responsibility. Introducing AI tools into this environment can trigger both resistance and problematic forms of reliance. Training, co-design, and gradual, transparent piloting are essential to ensure that AI is perceived as an aid rather than an intrusion or a constraint.

Finally, the use of AI in justice is unavoidably a matter of “democratic legitimacy”. Even if tools are technically sound and formally compliant with legal constraints, they may still be perceived by citizens and practitioners as undermining the human nature of adjudication or as introducing opaque technological intermediaries into the courtroom. In a context where trust in institutions is fragile, missteps in the communication or governance of AI projects can have long-lasting effects. The Italian judiciary will therefore need not only technical solutions, but also a strategy for public engagement: clear narratives about what AI is (or is not) capable of doing, explicit guarantees about the central role of the judge, and transparent channels for raising concerns and complaints.

3. Methodology

The methodological analysis that follows adopts a scenario-based approach grounded in the activities carried out during the preliminary investigation phase. The analysis focuses on a set of operational scenarios that are representative of everyday prosecutorial and law-enforcement practices, namely: (i) the classification and registration of *notitiae criminum*; (ii) the automated transcription and processing of intercepted communications; and (iii) the structured storage of investigative information for purposes of statistical analysis and crime prevention.

For each scenario, the analysis unfolds in two successive stages. First, it examines the practical and operational criticalities that characterize current practices, with particular attention to inefficiencies, fragmentation of information and procedural bottlenecks. Second, building on this assessment, the study explores potential AI-enabled solutions that could be implemented to address the identified shortcomings, with a focus on assisting preparatory and organizational tasks and enhancing the overall effectiveness of investigative workflows.

3.1. Critical Issue 1 – Automated Registration and Management of a Notitia Criminis

3.1.1. Problem Analysis

The management of a *notitia criminis* within Italian prosecutorial offices is increasingly strained by a persistent overload of incoming reports, heterogeneous documentation formats, and strict procedural deadlines. Prosecutors must process a high and often growing volume of initial crime reports, many of which arrive simultaneously and require immediate formal registration. This scenario generates structural pressures: bottlenecks in the intake phase, uneven allocation of cases among prosecutors, and a non-negligible risk of inaccuracies. Each *notitia criminis* is accompanied by a diverse set of documents – most commonly the preliminary report drafted by the judicial police, the complaint or notification originating the report, and any accompanying explanatory memorandum with annexed acts – followed by the formal registration form to be completed and entered into the so called *Sistema Informativo della Cognizione Penale* (SICP). The complexity and variability of these materials necessitate substantial manual review, during which personnel must extract key information, reconcile inconsistencies, and ensure coherence across the entire document set. This labor-intensive process not only slows the intake workflow but also makes it challenging to maintain uniform quality standards under conditions of time pressure and workload saturation. In this context, an automated system for the registration and management of *notitiae criminum* could provide concrete operational support by accelerating the preparation of the registration form, enhancing the accuracy and consistency of collected data, and facilitating a more balanced distribution of case files among public prosecutors. Nevertheless, the diversity of document types, the variability of writing styles, and the presence of domain-specific terminology pose significant challenges for AI-based solutions, which must be capable of interpreting

complex, non-standardized legal and investigative texts while operating within the constraints of a judicial information system.

3.1.2. Proposed Technologies

The registration and initial management of *notitiae criminum* is the first structured contact between raw incoming information and the criminal justice system. In current practice, this phase is characterized by a high degree of manual effort: operators open each incoming report, interpret its content, identify the parties and the alleged offences, assign one or more legal qualification codes, fill in the mandatory fields of the case management system, and finally route the file to the competent prosecutor or organizational unit. As highlighted in the introduction, this process is time-consuming, highly dependent on individual expertise, and particularly vulnerable to informational overload and local variability in practices. Automating this workflow, within a Human-in-the-Loop paradigm, aims not to replace the interpretative role of judicial staff but to support them in transforming unstructured, heterogeneous input into structured, reusable, and auditable data. From a technical perspective, the automated registration pipeline can be devised as a sequence of modular AI components operating on the incoming documentary flow.

First, a document ingestion layer [15] is responsible for acquiring files from existing channels (e.g., certified e-mail, police information systems, digital desks), normalizing formats, and performing Optical Character Recognition (OCR) [16, 17] and layout analysis on scanned or image-based documents. This step yields a machine-readable representation that preserves basic structural cues such as sections, headings, tables, and signatures, which are useful for downstream analysis. On the textual result, Natural Language Processing (NLP) [18] and Large Language Models (LLMs) [19] are employed to carry out core judicial information extraction tasks [20]. These include: (i) identification and normalization of the main actors involved (suspected persons, victims, reporting parties, institutions); (ii) extraction of key factual elements, such as time and place of the alleged conduct, *modus operandi*, and referenced objects or assets; and (iii) detection of procedural information, such as the origin of the report (police force, private individual, public administration) and any prior references to existing case numbers. When available, domain ontologies and controlled vocabularies can guide these models, constraining outputs toward legally meaningful categories and increasing consistency across offices.

A central component of the pipeline is the automated proposal of legal classification and registration metadata. The system can associate each *notitia criminis* with one or more offence categories and procedural qualification codes, following known classification schemes. Rather than committing to a single deterministic label, the system generates a ranked list of candidate classifications, each with a confidence score and a brief explanation that highlights the textual passages supporting the prediction. This representation serves as a decision support tool for the human operator, who retains full responsibility for the final classification and can easily compare alternative proposals.

Once the core metadata has been extracted and validated, an automated routing and management phase can be activated. Based on configurable criteria (e.g., offence category, geographical competence, specialized prosecutorial groups, or workload balancing) the system suggests the competent unit or prosecutor to whom the case should be assigned. At the same time, it can perform preliminary checks for potential correlations. For instance, by matching entities (e.g., persons, addresses, or vehicles), *modus operandi* patterns, or previous case identifiers against existing records, the system can flag reports that may relate to ongoing investigations or previously archived proceedings. These correlation suggestions are again submitted to human review, allowing operators and prosecutors to accept, refine, or discard them.

The Human-in-the-Loop [21] dimension is embedded at multiple stages of this workflow. Each automated step output should be presented to the operator via an interactive interface. These human interventions are not merely post-hoc adjustments, as they can be logged and fed back into the models as supervision signals, enabling continuous improvement through online or periodic retraining. This allows the system to adapt to local drafting styles, emerging crime phenomena, and regulatory updates.

Organizationally, the introduction of automated registration tools reshapes the role of administrative

staff and prosecutors in the early stages of criminal proceedings. Administrative operators shift from performing repetitive, low-level data entry to supervising and validating higher-level proposals generated by AI systems, focusing attention on ambiguous or complex cases where human judgment is most needed. Prosecutors, in turn, benefit from more timely and structured information: incoming files can arrive already enriched with preliminary classifications, correlation suggestions, and basic analytical views (e.g., distributions by offence type or origin), which support early strategic decisions on prioritization and investigative directions.

Finally, the systematic structuring of *notitiae criminum* at the point of entry enables new forms of monitoring, evaluation, and policy support. Since each report is converted into a rich, machine-actionable representation, it becomes possible to perform aggregated analyses over time and across offices, supporting dashboards for workload monitoring, early detection of anomalous trends, and more informed resource allocation. At the same time, explicit logging of all automated suggestions and human decisions fosters transparency and accountability, as each registration step can be reconstructed *ex post*, along with the interplay between machine recommendations and human choices.

3.2. Critical Issue 2 - Automated Transcription of Intercepted Communications

3.2.1. Problem Analysis

In investigative contexts, particularly those involving the lawful interception (LI) of communications, transcription accuracy is of paramount importance. The resulting reports, whether complete verbatim transcriptions or structured summaries of interpersonal communications, play a critical role in prosecutorial decision-making and are repeatedly consulted by prosecutors, defense and the judiciary throughout criminal proceedings. The need for transcription arises not only from judicial and procedural requirements, but also from practical considerations related to evidence management. Compared to extensive raw audio or video recordings, precise and comprehensive transcripts are substantially easier to organize, archive, and analyze.

In light of the central role that accurate transcription plays in investigative and judicial processes, the legal frameworks governing lawful interception further underscore its importance. Within the Italian criminal justice system, the use of lawful interception activities is subject to stringent regulation, as outlined in the Code of Criminal Procedure (Art. 266 et seq.). These provisions define the specific circumstances under which wiretapping is permissible and establish the procedural requirements that must be strictly observed. Regardless of whether communications are obtained through spyware or environmental surveillance devices, the output of lawful interception activities in Italy typically consists of audio and/or video recordings. Once their relevance and evidentiary value have been assessed within the context of a given case, these recordings must be transcribed, thereby converting multimedia material into textual form suitable for judicial use.

This type of operation is typically executed through the active listening of the material collected by specialized judicial police personnel. Until fairly recently, no explicit guidelines addressing the modalities and procedures of the transcription process of audio and video recordings obtained were ever drafted. However, recent regulatory provisions, such as the “*Decreto Giustizia*” (d.l. No. 105 of 2023, converted into Law No. 137/2023), and the “*Riforma Nordio*” (Law No. 114 of 2024), have instituted certain precautions and guarantees in favor of the individuals intercepted. For instance, to date, it is no longer possible to transcribe content deemed irrelevant to the investigation, even in draft form (known in jargon as “*brogliacci*”). Moreover, the reports must be devoid of expressions that could potentially compromise the reputation of individuals or disclose details pertaining to the private lives of the interlocutors. Furthermore, all expressions that allow the identification of persons other than the parties involved must be discarded in the same way, unless such information is deemed relevant to the investigation.

Consequently, any potential automation of the transcription operations must be subjected to the same rules and procedures that govern the current, purely manual, transcription scenario. However, it is clear that any blind and indiscriminate application of the latest rules could limit the capabilities enabled

by contemporary or future AI transcription technologies, potentially diminishing their operational effectiveness and efficiency.

3.2.2. Proposed Technologies

The adoption of AI-based tools for the automated transcription of audio recordings has the potential to significantly enhance the operational capacity of lawful interception units. By enabling a greater volume of wiretapping material to be processed within the same temporal constraints, such technologies can substantially improve workflow efficiency. Moreover, emphasizing transcription quality through AI-assisted solutions can reduce the time burden on human operators, allowing them to focus on higher-value tasks, such as the interpretation of acoustically challenging or ambiguous recordings.

Accordingly, law enforcement agencies are increasingly exploring automation to streamline and optimize the transcription of the large volumes of investigative interviews and intercepted communications generated each year, particularly through AI models such as Automatic Speech Recognition (ASR) systems [22, 23]. To be effective in a forensic context, automated transcripts must support rapid evidence exploitation through features such as full-text search, speaker attribution, and, where applicable, automatic translation, while ensuring traceability of actions to preserve evidential integrity. The motivation for adopting ASR systems, such as advanced deep encoder-decoder architectures exemplified by OpenAI's Whisper [24], is centered on managing the immense volume and velocity of intercepted data. These systems promise an operational efficiency that traditional, purely manual transcription methods simply cannot provide. However, successful implementation of such models requires a careful and exhaustive evaluation of both the legal and technical risks inherent in applying generalized AI solutions to domain-specific, high-stakes forensic evidence [25].

Indeed, the introduction and utilization of such AI models is not without technical challenges. The analysis of intercepted communications within Italian prosecutorial offices is often hindered by a combination of structural, linguistic and contextual factors that significantly burden investigative workflows. Wiretapping byproducts are typically produced in large volumes and processed under strict time constraints, placing substantial pressure on human analysts and translators. Moreover, the communicative practices observed in many criminal milieus – particularly within organized crime contexts – are characterized by linguistic heterogeneity, frequent code-switching between regional dialects and standard Italian [26], and the use of implicit, allusive, or context-dependent references intended to obscure meaning. For instance, interlocutors may alternate between Neapolitan, Sicilian, or Lombard dialects within a single conversation, inserting idiomatic expressions or slang terms that are rarely documented in standard language corpora. Messages often employ metaphorical or coded language, such as referring to illicit activities using innocuous everyday terms (e.g., “moving merchandise” to indicate smuggling operations, or “meeting the accountant” to denote a clandestine financial transaction). Elliptical constructions, abrupt topic shifts, and incomplete sentences further complicate parsing and semantic analysis. These linguistic and structural features complicate comprehension even for native speakers and make the material difficult to standardize for analytical purposes. Moreover, despite Italian's status as a richly diverse Romance language – marked by significant dialectal variation – publicly available large-scale datasets have remained comparatively underrepresented when contrasted with those of major world languages such as English or Mandarin [27]. Challenges to automatic speech transcription may arise both from purely technical domains (like overall poor audio quality recording) to semantic-related domains, like issues brought by non-linear conversational structures, heavy usage of dialects and fragmented or elliptical speech, all of which increase the cognitive load required for accurate interpretation. Within this framework, a carefully-designed automated transcription system could contribute to improving efficiency by providing preliminary linguistic normalization and multilingual support. At the same time, these same features – dialectal variability, intentional obfuscation, conversational ambiguity – represent substantial obstacles for AI-based systems, which must be designed to operate robustly under conditions that diverge markedly from the clean, standardized data on which most language models are trained.

In light of the aforementioned challenges, it is essential to address the problem of accurate tran-

scription through an appropriately structured system design. Rather than treating the transcription of multi-speaker, multi-dialect audio recordings as a monolithic task, the problem can be decomposed into a set of smaller, more manageable subtasks. A single-model architecture is unlikely to robustly handle the full range of complexities present in real-world intercepted communications, which may include all of the aforementioned complexities (overlapping speech, heterogeneous acoustic conditions, multiple speakers, and dialectal variation). To overcome these limitations, a multi-layered architecture, conceptually similar to a processing pipeline, can be envisaged, in which the output of each model determines the activation and configuration of subsequent components.

In such an architecture, an initial model would operate as a preprocessing and routing layer, analogous to a reverse proxy. Its primary function would be to segment a continuous audio or video recording into multiple shorter audio excerpts based on speaker identity, leveraging speaker diarization techniques that exploit timbral features and speaker-specific characteristics observable in representations such as mel-spectrograms. Each resulting segment would be annotated with metadata, including a unique speaker identifier and, where possible, an estimate of the language or dialect being spoken.

These annotated audio segments would then be forwarded to specialized Automatic Speech Recognition (ASR) models, for example Whisper-like architectures, each optimized or fine-tuned for a specific language or dialect. By dynamically selecting the most appropriate ASR model for each segment, the system can significantly improve recognition performance, thereby maximizing transcription accuracy and minimizing systematic and language-specific errors. Each ASR component would output a verbatim transcription corresponding to its input segment.

Finally, once all segments have been transcribed, a Large Language Model (LLM) could be employed as a post-processing layer to aggregate and contextualize the individual transcripts. For segments exceeding a predefined length (τ_L) or informational threshold (τ_I), the LLM could generate concise summaries that are explicitly linked to their corresponding audio excerpts. This final step would support both detailed examination and high-level analysis of intercepted communications, while preserving traceability between the original audio evidence and its textual representation. Figure 1 provides a visual representation of the described architecture.

3.3. Critical Issue 3 - Statistical Analytics for Prosecution Workflows

3.3.1. Problem Analysis

Notitiae criminum, investigative acts, procedural decisions and ancillary communications flow through multiple organizational units, often under significant time pressure and with limited visibility over system-wide dynamics. While individual crime cases are handled with legal rigor, the broader documentary and operational flow remains largely underexploited as a source of strategic information. For instance, seasonal or contextual factors - such as increases in property crimes during holiday periods or spikes in cybercrime following major data breaches - can lead to sudden overloads in particular prosecutorial units, with downstream effects on processing times and overall efficiency. Without timely, data-driven insights, these dynamics are often addressed reactively rather than proactively. This limitation constrains the ability of judicial organizations to move beyond a purely repressive and case-by-case perspective on criminal phenomena. A more effective approach requires the capacity to extract and aggregate structured information from individual cases in order to identify emerging patterns, anticipate pressure points, and support preventive and organizational responses. Statistical analysis of document flows and procedural events can enable early detection of systemic stress, inform resource reallocation and reduce the risk of operational bottlenecks that undermine both effectiveness and procedural timeliness. AI-enabled systems, as will be discussed in the following section, offer a critical opportunity to address this gap by transforming unstructured and semi-structured prosecutorial data into actionable knowledge. By capturing, classifying and aggregating information embedded in case files, such tools can support a shift toward evidence-based, anticipatory governance of prosecutorial workflows.

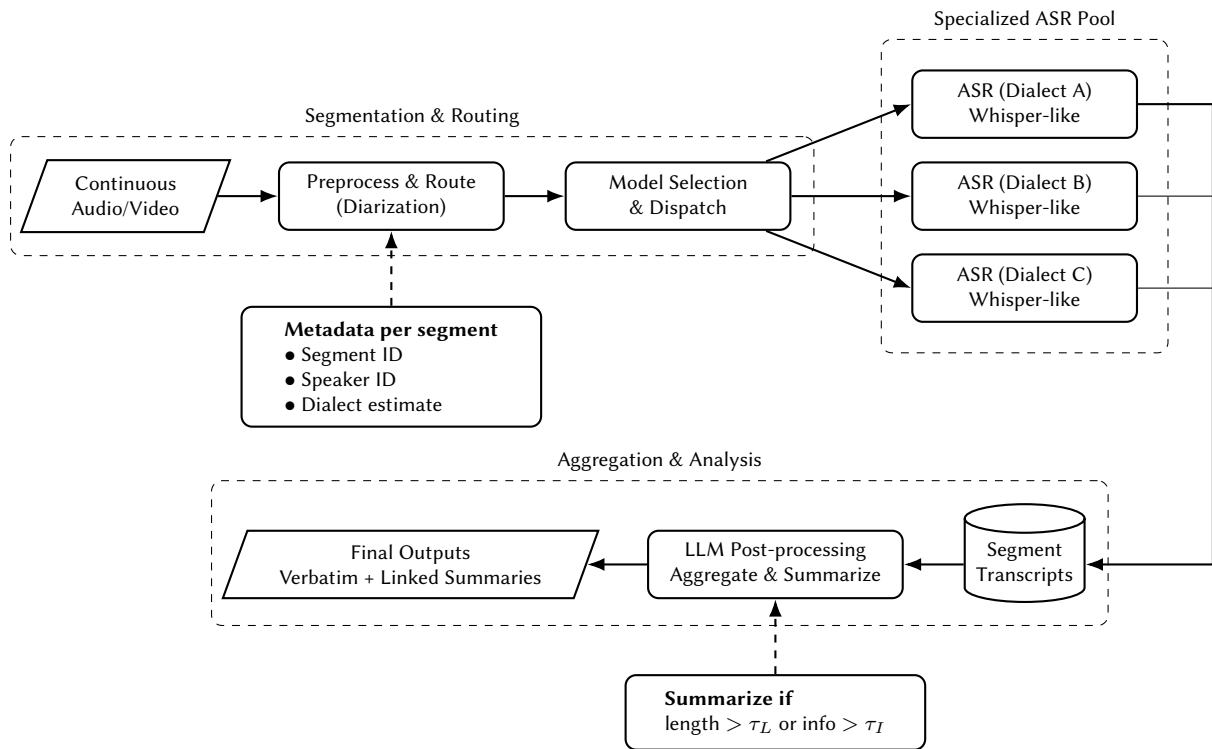


Figure 1: Multi-layer architecture for multi-speaker, multi-dialect transcription: diarization-based segmentation and metadata annotation, dialect-aware ASR routing, and LLM-based aggregation with optional summarization and traceability.

3.3.2. Proposed Technologies

Beyond the support to individual case management, AI-enabled systems make it possible to derive systematic, data-driven insights on prosecutorial workflows as a whole. Once *notitiae criminum* and related case events are registered in a structured and consistent manner, they can be aggregated along multiple dimensions – offence type, source of the report, territorial origin, assigned unit, processing stage, and temporal trends – to support statistical analysis and strategic decision-making. This type of analytics is not limited to traditional crime statistics but also extends to measuring internal organizational dynamics, such as workload distribution, processing times, and bottlenecks in the progression of cases through the various phases of proceedings.

From a technical perspective, the availability of structured data enables the construction of analytical models and dashboards that track key performance indicators (KPIs) for Prosecutors’ Offices. Examples include distributions of incoming reports by offence category and origin, time-to-registration and time-to-assignment for new files, transition times between procedural stages (from registration to the opening of investigations, from investigation to request for dismissal or indictment), and backlog indicators segmented by crime type or organizational unit. Statistical and machine learning techniques can be used to identify latent patterns in these indicators, such as systematic delays associated with specific offence classes, recurrent overload in certain units, or seasonal variations in incoming flows.

Large-scale analytics also enable prospective, rather than purely descriptive, views on prosecutorial activity. Predictive models can estimate future inflows of *notitiae criminum* based on historical trends and contextual variables, thus supporting capacity planning and resource allocation decisions. Scenario analysis tools can simulate the impact of organizational changes on key workflow metrics, such as the introduction of specialized groups, adjustments to routing rules, or the adoption of new digital channels. In a Human-in-the-Loop setting, these models serve as decision support instruments, presenting results through transparent visual interfaces accompanied by confidence intervals and explanations of the

most influential factors, which remain subject to the critical evaluation of judicial managers.

A further dimension concerns the monitoring of consistency and equity in prosecutorial workflows. By systematically analyzing how similar cases are classified, assigned and processed across time and organizational units, statistical analytics can help detect undesirable divergences (e.g., significant differences in processing times for comparable offence categories, or repeated re-routing of particular types of cases). While such patterns do not in themselves prove bias or dysfunction, they provide an evidence base for targeted audits, training interventions, or organizational reforms. Here, AI systems can support the identification of clusters of comparable cases and the computation of fairness-oriented metrics, while leaving the interpretation and normative evaluation of these findings to human decision-makers.

In conclusion, integrating statistical analytics into daily prosecutorial practice necessitates effective governance, interpretability and feedback mechanisms. Dashboards and reports must be aligned with the information needs of different roles, and updated at a cadence that matches operational decision cycles. Crucially, the same logging and audit structures that record the interaction between humans and AI at the micro-level (e.g., corrections to automated classifications or routing suggestions) feed into macro-level analytics, enabling offices to monitor the actual impact of AI tools on efficiency, quality and workload. In this way, statistical analysis of prosecution workflows becomes both a beneficiary of AI-driven information structuring and a key instrument for governing and refining the deployment of AI in the judicial context.

4. Discussion and Implications

The analysis demonstrates how a modular Human-in-the-Loop architecture, integrating various technical solutions (such as OCR, NLP, LLMs and graph-based correlation engines), can be seamlessly incorporated into existing prosecutorial information systems. This approach enables the transformation of unstructured documentary inflows into machine-readable formats, all while preserving the operational integrity of criminal proceedings. The direct implications of the proposed framework can be articulated across three main domains, as follows:

- *Implications for operational practice:* the systematic structuring of *notitiae criminum* and the automation of routine tasks (e.g., registration, preliminary classification, and routing) have the potential to reduce processing times and alleviate the cognitive and administrative burden on operative staff. Office personnel may re-orient their activities from manual data entry toward supervision of automated outputs, while prosecutors gain earlier access to more coherent and information-rich case files, supporting timely prioritization and investigative orientation.
- *Implications for technological infrastructure:* the proposed approach requires a technological stack built on a shared, structured data layer and loosely coupled AI services rather than monolithic applications. Concretely, this entails (i) the introduction of an ingestion and preprocessing layer capable of operating with heterogeneous documents; (ii) the deployment of NLP and LLM-based services that enrich incoming reports with entities and legal classifications, and (iii) the adoption of a persistence model that combines different storage paradigms (e.g., document-oriented, relational) to store the results of the automated pipeline. Around these components, a logging infrastructure is required to monitor performance and drift, as well as to maintain complete audit trails of inputs, model suggestions, and human corrections. Moreover, this architecture should support interoperability with existing case management systems, gradual roll-out across offices, and provide a controlled environment in which AI components can be updated or replaced without disrupting core judicial workflows.
- *Implications for policy and system-level governance:* the integration of AI solutions into prosecutorial workflows necessitates robust governance to ensure transparency, fairness, and accountability. Policies must ensure human oversight of AI-generated outputs, particularly in *notitiae criminum* classification (see par 3.1). Furthermore, the management of judicial data requires compliance

with applicable legal standards concerning data protection (e.g., the GDPR and the Law Enforcement Directive 2016/680, as implemented in Legislative Decree No. 51/2018). Specifically, for the processing of *notitiae criminum*, the adoption of pseudon/anonymization techniques is crucial, not only to ensure compliance with privacy regulations, but also to enhance the effectiveness and performance of AI systems by safeguarding sensitive data during processing.

Taken together, these considerations bring to the fore that the responsible adoption of AI in Public Prosecutors' Offices is not merely a technical exercise but a multi-stakeholder endeavour involving operators, system architects, and policymakers alike. For *judicial practitioners*, the proposed framework offers a concrete pathway to alleviate informational overload and improve the timeliness and quality of prosecutorial action. For *technologists*, this represents both a technical challenge and a research opportunity, that is, developing compliant infrastructures to support the AI deployment in the judiciary. For *policymakers*, it underscores the urgent need for clear governance instruments, coordinated with the EU AI Act's regulatory framework for high-risk AI (Arts. 6 and Annex III), and informed by EU judicial governance models, thereby ensuring that AI deployment remains anchored in the constitutional principles of legality and due process.

5. Conclusions

Starting from an empirical analysis of current practices and structural constraints in Italian Public Prosecutors' Offices, the present paper has highlighted how informational overload, unstructured documentation, and limited analytic capacity jointly contribute to delays, fragmentation, and even suboptimal use of available information. Against this background, we have proposed a set of practical scenarios in which AI does not replace prosecutorial judgment, but augments it by transforming raw documentary inflows into structured, navigable, and analyzable knowledge.

At the operational level, the proposed analysis covers three phases of the preliminary investigations. First, automated registration and management of *notitiae criminum* aim to reduce bottlenecks at the entry point of the system by leveraging OCR, NLP, LLMs and configurable classification pipelines to support the extraction of key entities, facts and legal categories. Second, automatic speech recognition models are employed to automate the transcription and management of wiretapped communications, streamlining the collection of transcribed material and reallocating human effort toward higher-complexity transcription and analytical activities. Third, statistical analytics over prosecutorial workflows enable an aggregated view on how cases are received, processed, and resolved, providing evidence for capacity planning, organizational reform, and monitoring of consistency.

A key contribution of this work lies in articulating these components within a coherent Human-in-the-Loop architecture. Rather than treating automation as a self-contained objective, we have emphasized interaction points where judicial officers and prosecutors review model outputs, correct classifications, and interpret analytic indicators. These interactions are not merely safeguards, but central design elements: they provide continuous supervision signals for the models, anchor AI-supported decisions in legal expertise and help build trust by making uncertainties, rationales, and limitations explicit. In this sense, the human role shifts from low-level data entry and ad hoc pattern recognition toward higher-order oversight, prioritization, and interpretation.

This study is necessarily limited in several respects. The scenarios have been instantiated within a specific national and institutional context, and further work is needed to validate and adapt the proposed models across different offices and legal systems. The quality and representativeness of training data, the robustness of models to distributional shifts (e.g., new types of offences or drafting styles), and the long-term impact of AI tools on prosecutorial strategies and resource allocation remain open empirical questions. Future research should combine technical evaluation (e.g., extraction and classification performance, transcription accuracy and word error rate, predictive accuracy of workflow models) with socio-legal studies examining how these tools influence decision-making practices, perceived legitimacy and organizational outcomes.

Despite these limitations, the results suggest that AI-driven structuring, correlation, and analysis of judicial information can play a meaningful role in addressing some of the systemic bottlenecks affecting Public Prosecutors' Offices. By embedding automation within a Human-in-the-Loop framework, it is possible to harness the strengths of modern AI technologies while preserving, and in some respects amplifying, the centrality of human legal judgment. In this perspective, the approach outlined in the present paper can be seen as a blueprint for a gradual, accountable, and evidence-based digital transformation of prosecutorial activity, with potential benefits not only for efficiency and organizational resilience, but ultimately for the quality and timeliness of criminal justice as experienced by citizens and institutions.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-5 and DeepL in order to: Grammar and spelling check, text translation and to improve the writing style. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] M. della Giustizia, Ricognizione della digitalizzazione del processo civile e penale e della transizione digitale del ministero della giustizia, https://www.giustizia.it/giustizia/page/it/digitalizzazione_del_processo, 2021. A cura della Direzione generale per i sistemi informativi automatizzati. [Accessed: 2025-12-01].
- [2] A. N. Magistrati, Dettaglio proposta ANM sui "carichi di lavoro" dei magistrati italiani, <https://www.associazionemagistrati.it/allegati/carichi-lavoro-metodo-e-proposta-operativa-anm.pdf>, 2019. Documento approvato dal Comitato Direttivo Centrale, 9 marzo 2019. [Accessed: 2025-12-01].
- [3] Ministero della Giustizia – Direzione Generale di Statistica, Procure della repubblica con indicazione dei parametri più significativi, 2013. https://www.giustizia.it/resources/cms/documents/Elenco_uffici_procura_15_parametri.pdf [Accessed: 2025-12-01].
- [4] P. di Torino, Trasparenza – Procura di Torino, 2024. <https://procura-torino.giustizia.it/trasparenza.page> [Accessed: 2025-12-01].
- [5] Consiglio Superiore della Magistratura, Raccomandazioni sull'uso dell'intelligenza artificiale nell'amministrazione della giustizia, 2025. Delibera plenaria dell'8 ottobre 2025.
- [6] E. Parliament, the Council of the European Union, Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828 (artificial intelligence act), 2024. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689, [Accessed: 2023-12-14].
- [7] Court of Justice of the European Union, Artificial intelligence strategy of the court of justice of the european union, 2023. URL: https://curia.europa.eu/jcms/upload/docs/application/pdf/2023-11/cjeu_ai_strategy.pdf, [Accessed: 2023-12-14].
- [8] E. C. for the Efficiency of Justice (CEPEJ), European ethical charter on the use of artificial intelligence in judicial systems and their environment, Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018), 2018. URL: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>, [Accessed: 2023-12-14].
- [9] Tribunal des activités économiques de Paris, Charte d'utilisation des systèmes d'intelligence artificielle et des données personnelles et sensibles au sein du tribunal des activités économiques de paris, 2025. https://cdn.prod.website-files.com/601987a724bdae251872ed2c/68c1480bfd1cce0f32afdf28_CHARTE_IA_1757489895.pdf [Accessed: 2025-12-01].
- [10] C. de Cassation, Préparer la cour de cassation de demain: Cour de cassation et intelligence artifi-

- cielle, 2025. https://www.courdecassation.fr/files/files/Publications/IA%20-%20Rapport%202025/Rapport_IA_2025_Web.pdf [Accessed: 2025-12-01].
- [11] R. voor de Rechtspraak, Responsible and innovative: AI for a fair dutch judicial system, 2025. <https://www.rechtspraak.nl/Organisatie-en-contact/innovatie-binnen-de-rechtspraak/Paginas/AI-Decree.aspx> [Accessed: 2025-12-01].
- [12] S. G. del CTEAJE, Politica de uso de la inteligencia artificial en la administración de justicia, 2024. Available online: <https://www.administraciondejusticia.gob.es/documents/7557301/7558184/CTEAJE-NOR-Politica+de+uso+de+la+IA+en+la+AJ+v1.0.pdf/ddc0eda1-950b-e926-b367-be511b16f2f9?t=1733925140122>.
- [13] E-Justice-Rat / Bund-Länder, E-justice-rat / bund-länder - official website, 2024. URL: https://en.justiz.de/laender-bund-europa/e_justice_rat/index.php;jsessionid=55C6A661489BBDB468CC7B7274E981CB, [Accessed: 2024-12-12].
- [14] Court, T. Judiciary, Artificial intelligence (AI) guidance for judicial office holders, 2025. URL: <https://www.judiciary.uk/wp-content/uploads/2025/10/Artificial-Intelligence-AI-Guidance-for-Judicial-Office-Holders-2.pdf>, [Accessed: 2025-12-12].
- [15] N. Gopalakrishnan, R. A. Andrews, H. D. Rajeshbhai, V. H. Ramurs, R. A. Bangalore, Current state of document ingestion and understanding, in: 2023 2nd International Conference on Futuristic Technologies (INCOFT), 2023. doi:10.1109/INCOFT60753.2023.10425676.
- [16] G. Nagy, At the frontiers of OCR, Proceedings of the IEEE (1992). doi:10.1109/5.156472.
- [17] T. T. H. Nguyen, A. Jatowt, M. Coustaty, A. Doucet, Survey of post-OCR processing approaches, ACM Comput. Surv. (2021). doi:10.1145/3453476.
- [18] L. Qin, Q. Chen, X. Feng, Y. Wu, Y. Zhang, Y. Li, M. Li, W. Che, P. S. Yu, Large language models meet NLP: A survey, 2025. doi:10.48550/arXiv.2405.12819.
- [19] J. Pan, G. Li, A survey of LLM inference systems, 2025. doi:10.48550/arXiv.2506.21901.
- [20] M. Siino, M. Falco, D. Croce, P. Rosso, Exploring LLMs applications in law: A literature review on current legal NLP approaches, IEEE Access (2025). doi:10.1109/ACCESS.2025.3533217.
- [21] S. Natarajan, S. Mathur, S. Sidheekh, W. Stammer, K. Kersting, Human-in-the-loop or AI-in-the-loop? Automate or collaborate? (2025). doi:10.1609/aaai.v39i27.35083.
- [22] M. Malik, M. K. Malik, K. Mehmood, I. Makhdoom, Automatic speech recognition: a survey, Multimedia Tools and Applications 80 (2021) 9411–9457. doi:10.1007/s11042-020-10073-7.
- [23] H. Kheddar, M. Hemis, Y. Himeur, Automatic speech recognition using advanced deep learning approaches: A survey, Information fusion 109 (2024) 102422. doi:10.1016/j.inffus.2024.102422.
- [24] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, Robust speech recognition via large-scale weak supervision, in: International conference on machine learning, PMLR, 2023, pp. 28492–28518.
- [25] R. Stoykova, K. Porter, T. Beka, The AI Act in a law enforcement context: The case of automatic speech recognition for transcribing investigative interviews, Forensic Science International: Synergy 9 (2024) 100563. doi:10.1016/j.fsisy.2024.100563.
- [26] M. La Quatra, A. Koudounas, E. Baralis, S. M. Siniscalchi, Speech analysis of language varieties in Italy, arXiv preprint arXiv:2406.15862 (2024). doi:10.48550/arXiv.2406.15862.
- [27] M. Giordano, C. Rinaldi, A survey on spoken italian datasets and corpora, IEEE Access (2025). doi:10.1109/ACCESS.2025.3538952.