*European Review of Digital Administration & Law - Erdal*
2024, Volume 5, Issue 1, pp. 173-184
ISSN 2724-5969 – ISBN 979-12-218-1507-8 – DOI 10.53136/9791221815078 13

# *Digitalization of Italian Administrative Justice: the Use of AI Technologies**

**Brunella Bruno**
(State Counsellor, Head of the Administrative Justice Information Technology Service)

**Domenico Franco Sivilli**
(Director-General for Information Resources and Statistics of Administrative Justice)

**ABSTRACT The contribution illustrates the experience of Italian administrative justice in the digital evolution, with particular reference to the implementation of recent projects for the introduction of AI technologies. The approach is operational and concrete and contains, in the initial part, a concise exposition of the goals achieved so far, with declination of the conditions that have allowed the start-up of the most innovative projects, in a framework in which the centrality of data has been strongly emphasised through numerous interconnected initiatives. The next development concerns the strategy for the use of AI technologies - in line with European regulation and inspired by a vision that has even anticipated the choices emerging from the initiatives of national discipline -, from which specific application spin-offs follow. The use cases developed will be indicated and, above all, the centrality of the judge's role in defining the architecture of the products, in identifying methodologies and in training and supervision activities will be highlighted. The exposition concludes with a focus on cyber profiles pertaining both to security understood in a broad sense, including not only the aspects related to the growing cyber threat but also those directly related to the explicability of AI systems, and to environmental sustainability. In terms of content, this work provides operational guidance on a state-of-the-art model, shaped on real needs, with a view to concrete and measurable benefits and reliable, credible and sustainable future developments.**

**KEYWORDS: Explainability - IA system - Administrative Justice - LLM, NER, NLP - Deep learning – Quantization – Cybersecurity – Ethical - Privacy**

**TABLE OF CONTENTS: 1. State of digitisation: realised developments and ongoing projects. – 2. AI in Italian Administrative Justice: deployment strategy and application effects. – 3. Selection and implementation of use cases: central role of the judge. – 4. Security implications of using AI. – 5. Relevance of environmental sustainability.**

## 1. *State of digitisation: realised developments and ongoing projects*

The awareness of the strategic role of technological evolution, which is necessary for the streamlining of processes and activities, has long been consolidated by the Italian administrative-justice system, also facilitated by the particular vision that derives from the performance of its functions. The administrative judge, in fact, is a very sensitive sentinel of modernity, having to provide, in a context characterised by technological progress so fast that it often does not allow *ex ante* intervention by the legislator, immediate responses to the protection needs connected to the use of new technologies.[1]

With this in mind, the administrative-justice system was the first in Italy among all jurisdictions to launch, as of 1 January 2017, the telematic process, digitised in its entirety and not in individual parts. The filing of deeds and documents relating to appeals takes place in digital form, according to the rules of the telematic administrative process (PAT) and the digitisation of the past has also taken place.

This was an epochal stage, but certainly not a point of arrival, due to the incessant adaptation and refinement required. However, this stage marked a clear dividing line between a 'before', made up of papers and people, and an 'after', marked by the intermediation of increasingly refined technological systems capable of providing complete coverage of activities. It cannot be disputed that the availability of all-digital records is one of the preconditions without which it is not possible to think of introducing tools that make use of artificial intelligence (AI).

The Information System of Administrative Justice (SIGA) constitutes the heart of the activities and is characterised by not being a container of disconnected information but an integrated platform that allows all categories of users (judges, secretariats, parties,

*Article submitted to double-blind peer review.

[1] L. Carbone, *L'algoritmo e il suo giudice*, in *www.giustizia-amministrativa.it*, 2023; V. Neri, *Diritto amministrativo e intelligenza artificiale: un amore possibile*, in *www.giustizia-amministrativa.it*, 2021; G. Gallone, *Riserva di umanità e funzioni amministrative*, Milan, Cedam, 2023, 114 ff.

defenders) to operate and to have a digital database that constitutes the valuable and secure resource also used to train artificial-intelligence models.

The developments that have taken place have made it necessary to reconsider the infrastructural choices made, which have also been determined by other factors, essentially identifiable in security requirements and the need to ensure compliance with the regulations that have arisen.

In November 2020, in compliance with the provisions of the Three-Year Plan for IT in Public Administration, the migration process to the cloud of the SIGA system[2] was planned. AgID established, in fact, that the data centres (such as that of the Administrative Justice) classified as 'type B' according to the census conducted in 2019, could not be further enlarged, with the consequent need to migrate to qualified cloud-service providers meeting certain requirements, given also the limited residual 'capacity' of the internal data centres.

The migration of SIGA to the cloud took place last year, following the acceleration of this delicate and complex process, and is a fundamental prerequisite for the achievement of further objectives, including the management of AI projects and their subsequent evolutions with adequate computing resources.

The clouds used, the same ones that are also used by important European institutions, are qualified for the type of data of the Administrative Justice and issued by the National Cybersecurity Agency.[3]

Indeed, one of the pillars of the national digital transformation strategy is the cloud-first approach, which includes the creation of the National Strategic Hub (PSI), a cloud infrastructure with data centres located in two Italian regions, with the aim of providing the public sector with an efficient, secure and reliable cloud.

In this regard, it should be pointed out that it has not escaped the attention of attentive operators that the Court of Justice of the European Union, in its document relating to its strategy for the use of AI, published at the beginning of this year, identifies a best practice in AI solutions that are to be installed and used on-premise, with a non-positive assessment of cloud technologies in the judicial sphere.[4] Evidently, this position reflects the consideration of possible riskier uses of AI technologies in the judicial sector, currently excluded in the Italian administrative-justice system, with the result that there are no factors pushing in a different direction, unsupported by the strategy followed at national level.

Along the outlined evolutionary path, a series of initiatives, which are strongly interconnected, are inserted. On the one hand, in fact, the areas already subject to digitisation are affected by complex interventions, including a substantial re-engineering activity of internal portals (of the magistrate, the president, the secretariats) and external portals (of the citizen, the lawyer, the judge's auxiliary), which go beyond the mere graphic restyling, introducing partly new logics and state-of-the-art technologies, with the aim of greater efficiency and the achievement of increasingly high qualitative and quantitative standards.

From another point of view, the perimeter of digitalisation is progressively and significantly increasing its extension, both with regard to the functions performed by the Italian administrative-justice system (the reference is, in particular, to the digitalisation of the cycle of affairs pertaining to the Consultative Sections of the Council of State, which include the opinion rendered in the context of the justicial remedy of the extraordinary appeal to the Head of State) as well as in relation to administrative activities (in addition to the realisation of the magistrate's virtual file, it is worth mentioning the project, currently at an advanced stage of realisation, concerning the digitalisation of the budget).

It is in this context that a nucleus of innovative projects is to be found, which have been made possible by the opportunities offered by the National Recovery and Resilience Plan for Italy (PNRR), submitted to

[2] The three-year plan for IT in the public administration constitutes the reference document for the individual PAs, in accordance with Article 14-*bis* of Legislative Decree No 82 of 7 March 2005, which entrusts the Agency for Digital Italy (AgID) with the task of drafting it and verifying its implementation.

[3] For an illustration of the profiles related to the definition of the perimeter of national cyber security and the certification system, see B. Bruno, *Cybersecurity tra legislazioni, interessi nazionali e mercato: il complesso equilibrio tra velocità, competitività e diritti individuali*, in *www.federalismi.it*, 14, 2020.

[4] https://curia.europa.eu/jcms/upload/docs/application/pdf/2023-11/cjeu_ai_strategy.pdf.

the European Commission on 30 April 2021, pursuant to Article 18 of Regulation (EU) No. 2021/241, the approval of which was notified by the General Secretariat of the Council in a note dated 14 July 2021. These projects, for the purposes that are relevant here, include the implementation of a business-intelligence and artificial-intelligence platform. The core of these projects is the centrality of secure and quality data, with a view to the management and exploitation of information assets.

Ahead of schedule, the Administrative Justice Data Warehouse (DWH) was implemented, a project destined to radically change the GA's statistical analysis capabilities. The DWH is able to historicise the data present in the information system (SIGA) and make thematic analysis areas (data marts) available to stakeholders (magistrates, secretariats, IT officials, etc.) that allow data to be represented in usable information for decision-making processes. The system will make it possible, for example, to know in real time the number of pending cases as a whole and in each individual court seat, the time taken to settle disputes, and quantity and quality size of litigation.

This development is the basis for the project to create the business-intelligence platform, with the adoption of software that enables access to customisable graphical user interfaces. This project also includes the Open Data portal of the administrative-justice system, through which the complete fulfilment of the relevant obligations deriving from EU and national regulations will be ensured, the promotion of transparency through the publication of datasets freely accessible from the institutional website, the facilitation of the conscious participation of citizens, the promotion of innovation and research, and the implementation of efficiency, as these data can be used to optimise internal processes, improve planning and make informed decisions based on concrete data.

The strong correlation between Open Data and AI engagement should not be overlooked. Underlying both are data that must have certain characteristics and be of quality. Open data can be a reservoir that can be drawn upon to use AI tools. From another point of view, certain uses of AI can help to respect specific constraints (think of privacy regulations) in the direction of also preserving an indispensable quality that is that of the comprehensibility of the data.

Hence, important pieces of a much more complex picture of ongoing developments come to the fore, with respect to which the organisational and management set-up plays a decisive role. A strong integration of competences is, in fact, indispensable and is ensured by the presence within the competent Service of magistrates who work in close synergy with the engineering-informatics component, committing themselves to the implementation, as accurate and calibrated as possible on the specific concrete needs, of all projects. Often, in fact, especially in relation to the most complex and innovative activities, it is not only a need for customisation of products available on the market that comes to the fore, but, as will be highlighted below with regard precisely to AI-related projects, a much more demanding activity that includes the very definition of product architecture, as well as the instruction of AI tools, their training and supervision and control.

## 2. *AI in Italian Administrative Justice: deployment strategy and application effects*

The guideline followed in launching AI projects was that of laying solid, secure and reliable starting bases, which, without binding choices referring to later -though largely- foreseen developments, would allow to grasp the opportunities offered by new technologies, but with AI being given a clear instrumental role, supporting the judge in research, analysis and updating activities, without any interference with the decision-making activity, which is entrusted exclusively to the judge.

This approach makes it more pertinent to speak, in relation to the use of these technologies in Italian administrative justice, of 'accelerated intelligence' rather than 'artificial intelligence', emphasising the auxiliary role entrusted to it.

Machine-learning models, in particular deep-learning models, are complex to handle and can present problems such as hallucination phenomena or overfitting to the data (overfitting and overgeneralisations). These risks are particularly insidious in a sensitive context such as the judiciary.

The difficulties of AI in performing complex legal reasoning are widely known to practitioners, necessitating a prudent and well-supervised approach.

People therefore remain at the centre, since the judicial function is entrusted, by the

Constitution, to the judge, who is pre-established, third to the parties, impartial and bound only to the law, and since the protection of rights and ethical values cannot be delegated to technology. Technical solutions have followed this requirement, while maintaining full control of the decision-making process by the judge.

In line with the approach taken, targeted decisions were taken on the implementation and concrete level.

First of all, the use of AI for the generation of texts of any kind was ruled out as it was not considered a solid starting point on a finalistic level to ensure the best quality, transparency and explainability of AI systems. Moreover, none of the use cases developed by the Italian Administrative Justice are among those defined as high-risk under the IA Act (EU Regulation 2024/1689 of 13 July 2024), as they do not concern the decision-making activity of the concrete case. The choice made by the Administrative Justice anticipated those initiated at a national level; the reference is to the draft law, No. 1146AS, concerning 'Rules for the development and adoption of artificial intelligence technologies', approved by the Council of Ministers on 23 April 2024, which allows the use of these technologies in the judicial sector only for the organisation and simplification of work and for case law and scholarship research.[5]

In order to ensure stronger security guarantees, the training data of the models adopted are those in the internal database, without contamination by external sources and especially without synthetic data generated by other AI systems.

The most innovative aspect in the implementation of these projects must be identified in the conception of an exclusive architecture, through the elaboration of specific pipelines for each use case, which was designed jointly by the judges and the administration's engineering-informatics team, with the involvement of prominent external partners, selected through a public procedure. This method made it possible to accurately calibrate the technologies used in relation to the purposes pursued, with the aim of guaranteeing a high level of reliability and security, also through the use of a supervised approach, which is valuable in avoiding the so-called black-box phenomenon. Therefore, the instruments that come to the fore are exclusive in that they were designed and implemented by the Italian administrative-justice system.

Moreover, the solutions adopted are never centred on a single technology, instead resort to the combination of several methods and models, in combination with each other. This approach stems both from security requirements, as the use of more than one technology reduces the risk of attacks, at least those likely to have a greater impact, and from the aim of achieving better performance.

No less important is the solution adopted for training, as it is not only the initial training but also the subsequent training that is supervised. This means that the feedback provided by the users when using the products will not be taken up automatically by the models, but will first be scrutinised by a unit made up of the judges and the computer scientists who implemented the project and are continuing its development.

Particular consideration was also given to profiles related to environmental impact, which must constitute a guiding factor in the choices of a modern administrative justice, attentive to all implications and capable of balancing the different values involved. To this end, recourse was made to quantization algorithms that make it possible to transform high-dimensional data into a more compressed space, without any significant alteration in performance; in this way - as will be better specified on a technical level in section 5 -, we obtain: lower energy consumption; lower latency, with greater speed; the occupation of fewer resources, with a consequent positive impact on costs as well.

## 3. *Selection and implementation of use cases: central role of the judge*

The identification of the use cases of AI technologies was the subject of careful analysis, dating back to the initial phase, of presentation, in 2020, of the financing project within the PNRR which was kept firm, despite the integration consisting of a more detailed illustration of the objectives within the framework of the strategy followed.

The use cases developed, in fact, were selected on the basis of the concrete utility to be derived from them and the expected benefits in relation exclusively to support activities in management and organisation,

[5] https://www.senato.it/service/PDF/PDFServer/BGT/01418921.pdf.

research and study activities, as well as with reference to tasks delegated to the secretariats. In fact, it is not experiments with the purpose of speculative and theoretical knowledge that come to the fore, but projects oriented towards concretely measurable benefits according to which they have been financed, with great commitment on the part of the administration, which has also allocated its own resources, both human and economic.

There are five use cases with the following contents and purposes.

A need that has long been strongly felt is the ability to identify appeals pending in the individual chambers of the judicial offices and to be set for decision. Being able to intercept similar appeals to be decided on makes it possible to achieve several objectives: 1) optimisation of the study and analysis; 2) evaluation for the purpose of discussion in the same hearing or in "thematic" hearings; 3) avoiding conflicting decisions in the individual sections; 4) achieving a better distribution of workloads; 5) ensuring faster decisions. In this direction, various interventions based on more 'traditional' computer systems have been implemented, which are useful for tracing connections using classification criteria but not for detecting similarities through semantic-analysis methods. The use of AI technologies in relation to this application case will allow a more complete mapping, overcoming the current limitations and providing more efficient tools in terms of management. This is, therefore, a useful use case for the staff of the trial office, the Secretariats and the Presidents, in order to improve the process of identifying cases to be set for hearing.

Another use case concerns the search of case-law precedents and is functional to speed up the study of files by judges. Currently, the search engine we use is focused on logical connectors and does not consider semantic connections. The result of a search does not always reflect an adequate level of relevance to the case under study. The objective pursued through the use of AI is to enable the search of case-law precedents, using natural language; the system processes the input semantically and the input vector is compared with that of the other elements in the database, with the subsequent display of a list of decisions, based on the level of relevance, on which the user can provide feedback.

Two other use cases may be examined together, since they are similar both in terms of the purposes pursued and the technologies used. The phase concerned is the one in which the judge is proceeding with the examination of a defence document that contains references to rules and case-law pronouncements. The current state of available tools implies the need for the judge to interrupt the analysis in order to search an external database for the norm or precedent. With the realised AI tool it will be possible to overcome this step, making the work smoother and faster. The pipeline in this case consists of two main segments: an extraction segment for identifying references on an external database and a verification and enrichment segment because once the reference has been identified, a link must be created for each element. In this way, each reference can be viewed by the judge with immediacy. Also for this use case, as for all others, there is the collection of feedback to detect errors or incompleteness. Not to be underestimated, also in relation to this use case, are complexity factors that may emerge in practice: the references contained in the acts, in fact, are not always explicit but often are implicit, making a semantic analysis necessary; moreover, to indicate a normative reference, lawyers may use different forms of writing (abbreviations, acronyms, etc.), thus also emerging as a result of the fact that the references are not always explicit but often implicit. In addition, lawyers may use different forms of writing (abbreviations, acronyms, etc.) to indicate a legal reference, thus highlighting the importance of uniformity, which, indeed, would be useful to consider not only with regard to citations of rules and precedents but, more generally, for the formats used in the preparation of documents.

Finally, another use case relates to a particularly delicate and relevant activity: the anonymisation of measures. The objective in this case is the availability of a tool that, without prejudice to the verification, modification and validation activities of the staff of the secretariats, provides an anonymisation proposal that is as compliant as possible with the reference legislation and at the same time avoids the excesses, often recorded in the performance of this activity by the secretariats, that compromise the very intelligibility of the measures and then make it difficult if not impossible to search for them.

A brief description of the process followed makes the essential role of the judge particularly evident. In fact, a distinction must be made between the decision-making process and the drafting of the written text to convey the information aimed at communicating and explaining the decision. The process of anonymisation must aim at a selective removal of information, with the need to identify and remove only information that is actually sensitive, as it entails privacy risks, and not also data that do not have that connotation and is essential to understanding the content of the text. This implies a balancing act that reaches the best possible trade off between the need to ensure the clarity of the text and the need to protect personal data; such a balancing act requires a careful selection of sensitive information and a consideration also of the context in which, within the act, this information fits. In order for the IA to perform a balanced anonymisation that complies with the relevant legislation, it is necessary to have precise operating instructions from the judges that guide the algorithm in the selection and processing of sensitive information, adapting to the specific contexts. Careful analysis of feedback is also essential to avoid distortions induced by user error and to allow the AI to evolve in a manner consistent with the above requirements.

The realised use cases are fully consistent with the above-mentioned criteria and the correctness of the selective choice made was confirmed in the practical implementation phase, in which all the complexity of a grounding involving multiple profiles emerged, starting with the instruction of models and the identification of corrective and refinement measures to refine the expected outputs, also in relation to applications that aim to streamline and speed up specific support and auxiliary activities.

The difficulties associated with the development of semantic analysis, which is essential to ensure the effective identification of afferents by specific processes, have highlighted the need to change the approach to how systems are interrogated. These processes, in fact, are based on different logics from those used hitherto, thus requiring specific training of users in the formulation of inputs and in the understanding of the operating limits of these systems.

The decision, moreover, to exclude, in this first phase, solutions aiming at the synthesis of texts and chatbots was reinforced, in the consideration that the work of the judge cannot and must not stop at the surface, as each element must be considered going deep into the formation of a conviction whose completeness and integrity must be preserved. These applications, which are certainly possible, can be more effectively introduced at a later stage and with less risk, when the projects implemented to date are integrated with the IACS, thus making them usable for all users, and when they are further implemented with the developments already identified. Crucial, both in the conception phase and in the implementation phase, is, as already pointed out, the laborious work of integrating competences, with an essential role of the judge in identifying the operational implications arising from the application of legal rules, to which all information systems and tools must conform.

This is an incessant activity that requires continuous refinement, as the judge has to propose but also simplify legal concepts, which are not always clearly defined, to allow their translation into computerised rules.

With farsightedness, the Administrative Justice has structured the organisation of the internal organisation responsible for digitisation by providing for the presence within it of judges who are able not only to convey more directly the needs related to digitisation but also to follow its developments, analysing the implications and supervising the results. This is even more necessary in the current phase of accelerating technological development.

The implementation of the IA projects, at the prototype level and which will be the subject of the subsequent integration with the SIGA, has shown that the judge's intervention in the implementation of these projects is not an option but is indefectible, intervening in the definition of the initial ground rules, i.e. in the technically defined process of prompt engineering, of instructing the AI; in the design for each use case of an architecture of the tools to be employed in such a way as to make the process complete, coherent and 'explainable' at every stage; in the verification of the results, which are also analysed with a view to perfecting the rules, requiring in many cases the resolution of conflicts between opposing requirements, since it is not possible to rely on generically trained AI systems,

which present limitations and risks; in supervision and subsequent fine-tuning; in the mapping and collection of the rules that are to be defined, in order to train the users who will have to use these tools.

**4.** ***Security implications of using AI***

Since 2020, Italian Administrative Justice has implemented advanced cybersecurity systems based on artificial intelligence. These systems, together with traditional cyber-defence tools, have proven effective in protecting judicial-information assets from cyber-attacks, which have occurred but have been adequately managed by this hybrid security ecosystem.

The application of AI was, on a technical level, a natural progression as a result of the knowledge gained and the results achieved in the cybersecurity sector. In that sector, AI makes it possible to quickly analyse large volumes of data, identify anomalous activities with precision and present them clearly to the end user. This increases responsiveness to cyber incidents and reduces the response time to intrusions.

In particular, the ability to analyse large amounts of data allows, through real-time monitoring tools, to obtain information on key malware. AI can provide significant help by suggesting actions to be taken to: a) prevent an attack by a specific malware; b) intervene immediately after detecting a compromised resource, or predict the malware's next step. In this case, AI enables specific ex post remediation and ex ante prevention, implementing an effective prevention policy.

In general, AI makes it possible to: easily detect threats by monitoring activities within the network; prevent security incidents by preventing unauthorised intrusion and protecting sensitive data; and implement corrective and adaptive measures following attacks.

It is important not to overlook the two-way relationship between AI and cyber security. On the one hand, AI can help detect and prevent cyber threats, verify the identity of users and monitor networks. On the other, AI can be used to create new forms of cyber-attack that are more sophisticated and difficult to counter, in a context of constantly increasing threats.

Several attack techniques targeting AI-based computer systems, known as adversarial machine learning, have been identified. These techniques fall mainly into three categories:

- data poisoning attack: this type of attack aims to compromise the data used for training or re-training an AI model. Attackers insert contaminated data into the training set, causing the model to learn incorrectly. Contamination can occur through the association of wrong labels or biasing towards a specific input category. As a result, malicious data are misclassified as legitimate, leading to the rejection of legitimate data. An example is the compromise of e-mail's spam filter, where attackers massively flag legitimate e-mails as spam. Once the filter is compromised, attackers can send malicious e-mails, including malware, bypassing security filters;
- evasion attack: this attack uses malware to compromise the predictions of an AI model. By introducing imperceptible changes to the original input, the model is induced to make classification errors. An interesting example is the use of special clothing (such as shirts that do not cover the face) to evade facial recognition techniques. Similar approaches can also be applied in malware analysis;
- extraction attack: the objective of this attack is to retrieve the training set on which the AI model was trained. The problem becomes more serious if the training set contains sensitive or confidential data, thus violating privacy legislation.

Adversarial techniques therefore pose a danger to defensive systems, as they bypass alarms and access control tools based on available data. However, they can also become a defensive tool, helping to design more robust anomaly detection and authentication systems against this type of attack.

The constant focus on IT security in Administrative Justice received a further boost with the start of projects to deploy AI technologies. Designing the AI architecture with the security-by-design approach made it possible to carefully prevent threats from the outset, ensuring a secure and reliable architecture.

In line with national and international analyses, there is an awareness that the concept of cybersecurity is no longer sufficient. It is necessary to adopt a broader and transversal approach, that of cyber resilience: the ability to continue to deliver the expected results despite the occurrence of

cyber-attacks.[6]

Specific strategies are constantly being adopted to defend AI systems against internationally recognised adversarial machine-learning techniques. These strategies include:

- reinforcement training: by including adversarial examples during training, the model is exposed to perturbed input, forcing it to generalise better without depending on specific features;
- the ensemble of models: combining the predictions of several models makes it more difficult for an attack to influence all models simultaneously. Ensembles increase the diversity of responses, improving the overall resistance to attacks;
- data verification: implementing verification checks to detect adversarial examples in the training dataset helps reduce the likelihood of such examples influencing the model. Removing or accurately labelling these examples preserves the integrity of the training;
- input transformation: applying slight transformations to inputs during the inference phase can reduce the impact of adversary attacks. These transformations, such as adding random noise to the input, make it more difficult for attacks to generate effective perturbations;
- sensitivity analysis: conducting sensitivity analysis to identify the part of the input where the model is most vulnerable to attack can guide the design of targeted defences and improve robustness;
- continuous updates: since adversarial machine learning is a constantly evolving field, it is crucial to keep defences up-to-date to ensure the security and effectiveness of the system.

Among the technologies underpinning the AI infrastructure of administrative justice are foundation models, sophisticated artificial intelligence systems based on neural networks trained with deep learning algorithms. A significant example of these systems are large language models (LLM).

The decision to use LLMs, which are considered basic pre-trained and multi-tasking models, was guided by the principle of application reuse in software engineering. These models can be adapted to specific needs with shorter training periods than conventional artificial-intelligence systems, which would require large amounts of data and specific training sessions, leading to higher costs in terms of time and required computing power. The application-functional architecture is designed to counter adversarial machine-learning attacks. This architecture is hosted in a European cloud region dedicated to Italian Administrative Justice, protected with advanced cryptographic techniques and encryption keys guarded by Administrative Justice. Training is carried out using selected and controlled data residing in dedicated and protected databases. The platform is structured in elementary 'macro-functions'. Lighter' language models (such as NERs, applications of natural language processing or NLP techniques) perform entity recognition and classification functions within a text (e.g. people, places, organisations, dates), while LLM models perform more complex tasks such as classification and semantic recognition of text. The output subsystem makes all the information produced by the AI accessible to external systems, using cloud services that allow the creation and execution of specialised applications for individual use cases, technically classifiable as containers.

After being collected from the sources, the data undergo pre-processing. In this phase, data from different sources are cleaned, transformed and prepared for model training. This process may include operations such as removing missing values, normalising the data and coding categorical variables.

Next, we move on to the artificial-intelligence processing layer, where relevant features in the data are identified. These features are used to train the model and are stored in the persistence layer. This step is crucial for improving model performance and reducing training time.

Model training takes place at this level, using supervised, unsupervised or reinforcement learning algorithms, depending on the specific problem to be addressed. Within the AI processing level, there is a class of models called sentence-transformers, which specialise in transforming text into a vector representation, a function commonly known as embedding.

Currently, all available templates have a maximum limit of language units (so-called tokens) that can be processed per single

[6] I. Ahmad, A. Amelio, D. Domenico and D.F. Sivilli, *Deployable artificial intelligence: cybersecurity, usability, metaverse and legal challenges*, in *www.federalismi.it*, 6, 2024.

instruction. Therefore, when working with long texts, such as court documents, specific strategies must be adopted to enable the models to process the entire document. One such technique is chunking, which aims to divide the text into smaller parts, or blocks, so that the model can handle them one at a time.

Due to the importance of text splitting, in the first phase of technical analysis it was necessary to study and compare different sentence transformer models, analysing their behaviour with different input chunking configurations. During training, the model learns from the data and optimises its parameters to minimise errors in the predictions.

Each use case is implemented by creating detailed processing chains, known as pipelines, organised in several steps with clearly defined logical and functional tasks. The controlled approach adopted in the project allows complex processing algorithms to be broken down into elementary and specific logical-functional steps, exploiting the principle of modularity within the pipelines. This allows the use of specialised models for certain functions, without abandoning generalist models that can be used for more complex steps or as validators of specific pipeline tasks.

Through various verification steps, the outputs of the various steps are examined to ensure the regularity of the processing. In addition, the overall control and management system monitors the flow from data acquisition to the return of the result, ensuring process management, data control and system security.

Once trained, the model is released for the inference phase and its accuracy and performance are evaluated using validation methods based on the feedback collected. Based on the results, the model can be further optimised. Monitoring of AI solutions is essential to ensure that the system continues to provide reliable and high-quality predictions and information. The system of collecting feedback from users makes it possible to monitor, correct and improve the performance of the AI solution. A 'guided' rather than 'automatic' feedback system was chosen. Although it requires considerable human effort on the part of the AI team, this approach allows for better control over the quality of feedback, directing corrections towards improving model training even without the need for large amounts of feedback. In contrast, an automatic feedback-collection system would have required a large amount of feedback for fine-tuning and could have introduced cognitive bias, thus worsening the model. The feedback-collection component is implemented through structured functionality in the front end of the system, which allows users to submit feedback for each use case. Users can provide feedback either through predefined classifications or through free-text fields. All these data are collected in structured tables in the system's database and subsequently used by the analysis components. In the analysis and validation phase, the feedback is examined and classified. For these operations, both classical techniques based on SQL queries and specific AI algorithms for classifying this type of data are used. The output of this analysis is fundamental both to feed a control report on the feedback trend and to identify any corrective actions needed in the event of negative feedback. With respect to human centrality, the decision to apply corrective actions or develop new functionalities always lies with the legal-informatics control team, which can, however, be supported by AI in proposing corrective actions.

Among the difficulties in approaching AI systems, also in the field of cybersecurity, is the complexity of understanding how these systems work, often perceived as 'black-boxes'. A new field of research, called Explainable Artificial Intelligence (Explainable AI or XAI), aims to make complex and often obscure machine-learning algorithms transparent. Despite the numerous approaches that have emerged to explain the decisions of black-box classifiers, these are not always easy and intuitive for end users. Even for AI experts, understanding how output is generated in highly complex, data-driven models and algorithms involving millions of interacting parameters is a challenge. The inherent complexity of these processes makes it difficult to fully understand how the models produce their final results. Developments in the field of XAI are the subject of constant attention in order to assess the introduction of the most suitable methodologies to ensure the explainability of the AI systems envisaged for administrative justice. Recent investigations have highlighted two fundamental aspects in clarifying a deep neural network (DNN) model: interpretability

and explainability. Interpretability allows developers to explore the model's decision-making process, enhancing understanding of how the model produces its results. Unlike a simple prediction, interpretation provides additional information or explanations crucial to understanding how an artificial-intelligence system works, offering significant help to those with the appropriate skills to understand black-box models.

Explainability, on the other hand, allows the end user to gain confidence in the accuracy and objectivity of the AI by providing details of the DNN decision-making process. This dual focus on interpretability and explainability contributes to a better understanding of the complexities involved in the operation of DNN models used in administrative-justice AI applications. The overall goal is to enable human users to understand, trust and competently supervise the evolving artificial-intelligence communities. On this basis, the developed AI architecture strongly integrates the concepts of supervision and control of machine-learning models. Although the implemented platform offers advanced artificial-intelligence services and uses highly specialised pre-trained models to perform complex functionalities, the approach taken was to ensure a high level of control and configurability of the solution.

Each use case is implemented through very granular processing pipelines, organised in several steps with logically and functionally well-defined tasks. During various verification phases, the outputs of the different steps are examined to ensure their correctness. In addition, the overall control and management system monitors the entire flow, from data acquisition to the return of the result, ensuring process management, data control and system security.

The high security standards implemented in AI projects by the administrative justice system, together with the general protection measures of the systems in use, require an essential complement in user behaviour, with precautions being indispensable from two perspectives.

On the one hand, on a technical and general level, all protection systems, starting with multi-factor authentication (MFA), must not be perceived as an obstacle, but as ground rules to be complied with. On the other hand, with specific reference to AI technologies, in spite of the numerous system checks and fine-tuning of training, the control by each user in relation to the activities performed remains fundamental, in accordance with the principle of the reservation of humanity.

Although the use cases outlined above are not among those considered high-risk under the IA Act and do not present similar problems to those associated with more advanced use of AI, such as text generation, the responsibilities involved in carrying out activities, especially in a sensitive area such as justice, still apply.

## 5. *Relevance of environmental sustainability*

In recent years, all transformations have been digitally underpinned, making digital transformation the true enabler for addressing ecological-sustainability issues. The strongly interconnected ecological and digital transitions are often considered 'twins' due to their strategic relevance at EU and national level. This is evidenced by the fact that the European Union has set its sights on both achieving climate neutrality and ensuring a sustainable, fair and globally competitive future. To achieve this dual transition, investments and expertise are needed, as the European Commission has emphasised in several contexts. Italy has been committed to digitisation and sustainability for several years, promoting initiatives and programmes within global and European policies. Today, there is a greater awareness that the combined digital and ecological transitions can offer countless advantages and opportunities. This cultural evolution involves both businesses and public administrations, which have realised how sustainability can become an opportunity.[7] There is in fact a growing awareness that offering green products, made in sustainable factories and supported by an ecological supply chain, represents an element of attraction for private and public customers, who are increasingly sensitive to the values of sustainability and in search of products that respect the environment. In the Italian entrepreneurial fabric, but also in the public administration, the need is felt to manage the two transitions jointly and in a coordinated manner: the digital one and the one towards sustainability. Digital is no longer only aimed at improving the performance of companies or public administrations, but is also a great

[7] https://www.agendadigitale.eu/smart-city/green-digital-transizioni-gemelle-futuro-sostenibile.

enabler for environmental sustainability. However, finding digital skills is already difficult and the challenge increases when looking for professionals capable of working in teams that address both dimensions: digitalization and sustainability. Sustainability, understood as technology, requires complex skills because it involves designing and implementing products and services with minimal environmental impact. The engineering know-how required to tackle sustainability is very high. Among the key strategies to achieve the dual transition to sustainability and digitisation, it is crucial to address the growing demand for skills for a sustainable future. The availability of professionals with appropriate transversal technical skills will be crucial for EU's global competitiveness. Therefore, there is a need to intensify efforts in the development of digital and sustainable skills, both through upskilling and reskilling. In Italian administrative justice, efforts have been made to foster the synergy between legal and information-engineering skills. Digital infrastructures, such as data centres and communication networks, require a lot of energy, especially when they host AI applications. Thinking about the energy impact of disruptive technologies, such as artificial intelligence, is crucial. In general, choosing cloud solutions has a positive impact on the environment and this is the choice that the Italian administrative-justice system has made. Cloud providers are more energy efficient than small and medium-sized private data centres. This is because they have the financial capacity to invest in state-of-the-art power and cooling technologies designed to consume less energy, reducing the impact of the large number of servers they manage.

In addition, the use of cloud services relies on shared resources, such as networks, computing machines and physical facilities, which are optimised to be fully utilised, reducing waste. For instance, the servers used for the cloud are optimised for dynamic scalability, saving resources. In short, cloud-service providers can achieve great economies of scale, maximising efficiency and utilisation of hardware and processes. Doing the same thing themselves, in their own on-premise IT infrastructure, would be much more costly. Accelerating the migration of data centres to greener technologies is essential to reduce the negative impact on the environment. The adoption of renewable-energy sources to power digital infrastructures is an important step, especially if it is accompanied by the use of decentralised smart grids that can manage and distribute the increase in renewable energy as required. Action is also needed to reduce the demand for computing capacity by switching off unused services and providing so-called serverless services, i.e. charging the user for applications only when they are required. In this way, services are not always up but delivered only when needed. Action must also be taken on the development side, reducing the oversized computational capacities that applications use through delicate tuning work.

Digitalisation can improve efficiency in various areas, but it must be guided by ethical and responsible principles. Considering the social and environmental impact of digital technologies during design and implementation means ensuring responsible digitisation. From this point of view, we should go back to thinking as we did forty years ago, when saving on the use of certain memory locations was indispensable to ensure the functioning of the developed applications. Particular attention must be paid to AI applications which, if left unsupervised and uncontrolled, can prove to be extremely energy-intensive. Training complex models (e.g. LLM models) requires intensive computing resources, resulting in significant energy consumption. One technique used to green AI applications is quantisation, which reduces the precision of data without significantly compromising its usefulness, thereby improving processing efficiency.

Quantization consists of converting the weights and activations of neural networks (the parameters that can be valued by training in the model) from floating-point numbers to integers with a smaller bit width. Although it can affect the accuracy of machine-learning models, especially those based on deep learning, quantization is often necessary to reduce the amount of memory and computing power, and thus the energy impact. The quantization technique has been successfully tested in the AI applications of Italian administrative justice related to PNRR projects. The predictions or classifications obtained from the models used in the previously described use cases are processed, converting the results into a comprehensible and energy-efficient format (quantified LLM) for end users or supplementing them with

other information. In this specific case, the LLAMA2 model, used in some pipelines, represents its basic information in a 32-bit format.

With specific mathematical processing, it is possible to create a version of the model with a lower number of bits, e.g. 4 bits. From the queries made and their comparison, in terms of required resources and response performance, it was observed that the difference in performance is in the order of a few percentage points, while resource consumption is very low. However, it is important to strike a balance between efficiency and accuracy in order to ensure optimal performance in artificial-intelligence (AI) applications, making general use of optimisation of applications and machine-learning algorithms. In the overall context, the production of digital devices, which generates waste and requires resources, should also not be underestimated.

Promoting repair, recycling and reuse of devices can reduce environmental impact. Companies should design digital products with a circular-economy perspective, promoting durability and reducing waste. Users can also contribute by reducing the environmental impact of their digital activities, e.g. by switching off devices when not in use, using more energy-efficient devices or reducing the use of unnecessary digital resources (e.g. by sending fewer emails with heavy attachments). In conclusion, balancing digital growth with environmental impact requires a holistic and collaborative approach involving companies, governments and individuals.

Public and private organisations have a duty to monitor the environmental impact of their digital operations and publicly report on progress. Transparency is key to identifying areas for improvement and taking corrective action. Technology can be a positive force if it is managed in a responsible, sustainable and shared way, and Italian administrative justice has implemented this paradigm for all its projects, including those on artificial intelligence.

Finally, a thought must be given to the future of artificial intelligence, which is today considered a disruptive technology. Although not recent, as we are experiencing it today, AI can be considered an emerging technology and, like any emerging technology, it can be both a replacement and an innovation. In 1970, traditional X-rays replaced older imaging techniques, but did not represent true innovation. In the 2000s, magnetic resonance imaging (MRI) transformed imaging, offering much more precise details and without the use of ionising radiation.

Today, many applications of AI are substitutes: interesting and useful, but not necessarily innovative. An AI that summarises a text or compiles an instruction is a replacement technology, while innovations such as Marconi's radio or space rockets have unlocked new human capabilities.

We are currently investing in mostly replacement AI. For innovative AI, we will need open minds and a scientific and technological breakthrough, not just financial resources. We should distinguish between SAI (Substitutive Artificial Intelligence) and IAI (Innovative Artificial Intelligence). Some groups are working on IAI, and we have to wait for one of them to succeed before we can take another step forward.

In Administrative Justice, we have developed use cases that, although they can be categorised as SAI (there are no implemented use cases that do not do what humans can do in other ways today), they represent innovative support due to the unique technology used and are useful for improving the justice ecosystem. We continue to monitor research and innovation in order to integrate emerging signals from the IAI in a timely manner.