PAM
Pokrajinski arhiv Maribor

Moderna arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 8 (2025), št. 1 / Year 8 (2025), No. 1

Maribor, 2025

Moderna arhivistika
Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

ISSN 2591-0884 (online)
DOI: https://doi.org/10.54356/MA

Letnik 8 (2025), št. 1 / Year 8 (2025), No. 1
DOI: https://doi.org/10.54356/MA/2025/1

Izdaja / Published by:
Pokrajinski arhiv Maribor / Regional Archives Maribor

Glavna in odgovorna urednica / Chief and Responsible editor:
mag. Nina Gostenčnik., Pokrajinski arhiv Maribor, Glavni trg 7, SI-2000 Maribor,
telefon/ Phone: +386 2 228 5017; e-pošta/e-mail: nina.gostencnik@pokarh-mb.si

Recenziranje / Peer review process:
Prispevki so recenzirani. Za objavo je potrebna pozitivna recenzija. Proces recenziranja je anonimen. / All articles for publication in the conference proceedings are peer-reviewed. A positive review is needed for publication. The review process is anonymous.

Lektoriranje / Proof-reading:
mag. Boštjan Zajšek, mag. Nina Gostenčnik

Prevajanje:
mag. Boštjan Zajšek (slovenščina), mag. Nina Gostenčnik (slovenščina, angleščina)

Oblikovanje in prelom / Design and typesetting: *mag. Nina Gostenčnik*

Prejeto / Received: 14. 01. 2025

1.02 Pregledni znanstveni članek 1.02 Review article

https://doi.org/10.54356/MA/2025/TCGJ3655

EUROPEAN ARTIFICIAL INTELLIGENCE ACT: REGULATION OF ARTIFICIAL INTELLIGENCE IN DIGITAL ARCHIVES

Lana ŽAJA, Ph. D.
Archivist, Croatian State Archives, Zagreb, Croatia
lzaja@arhiv.hr

Abstract:

European Artificial Intelligence Act (EU AI Act) was adopted by the EU Parliament in 2024 as part of the EU's legislative process. In this article, the author proposes solutions based on AI tools and methodological approaches applied to Digitised and Born-Digital Archival Records from the point of view of computational approaches to analyze Digital collections. Computational Research methods include a wide range of approaches, including text and Data Mining, Data Visualization, Digital Data Analysis, and the use of Language Models in large numbers of Big Data. The application of Artificial Intelligence (AI) and its subsets, such as Data Mining, Machine Learning and Natural Language Processing, has expanded the scope of computational Research methods in Digital Archives, from the creator of Records to the Information Digital Specialist (Digital Archivist) and all the way to the end user. Digital collections such as eArchives have been increasingly recognised as an essential source for Multidisciplinary Research in recent decades.

Keywords:

EU AI Act, Artificial Intelligence, AI, Digital Archives, Big Data, Machine Learning, Language Models, Data Mining, Cybersecurity, Digital Archives, AI Systems, Digital Archivist

Izvleček:

Evropski akt o umetni inteligenci: regulativni okvir za umetno inteligenco v digitalnih arhivih

Evropski akt o umetni inteligenci (EU AI Act) je parlament EU sprejel leta 2024 v okviru zakonodajnega postopka EU. V tem članku avtorica predlaga rešitve na podlagi orodij umetne inteligence in metodoloških pristopov, ki se uporabljajo za digitizirane in izvorne digitalne arhivske zapise, in sicer z vidika računalniških pristopov k analizi digitalnih zbirk. Računalniške raziskovalne metode vključujejo širok spekter pristopov, vključno z rudarjenjem besedil in podatkov, vizualizacijo podatkov, analizo digitalnih podatkov in uporabo jezikovnih modelov v velikem številu velikih podatkov. Uporaba umetne inteligence (UI) in njenih podvrst, kot so podatkovno rudarjenje, strojno učenje in obdelava naravnega jezika, je razširila področje uporabe računalniških raziskovalnih metod v digitalnih arhivih, od ustvarjalca zapisov preko informacijskega digitalnega strokovnjaka (digitalnega arhivista) vse do končnega uporabnika. Digitalne zbirke, kot so e-arhivi, so v zadnjih desetletjih vse bolj priznane kot bistven vir v multidisciplinarnih raziskavah.

Ključne besede:

EU akt o UI, umetna inteligenca, UI, digitalni arhivi, veliko podatkovje, strojno učenje, jezikovni modeli, rudarjenje podatkov, kibernetska varnost, sistemi umetne inteligence, digitalni arhivist

1 Introduction

The European Union Artificial Intelligence Act[1] (hereinafter referred to as the EU AI Act) is the first comprehensive legal framework for Artificial Intelligence (hereinafter referred to as the AI) at a global level, which entered into force in 2024. The aim of this Act is to encourage responsible development and use of AI in the EU, while ensuring the protection of fundamental rights, security and ethical principles. The key provisions of the EU AI Act are risk-based approaches and this applies to AI systems, which are categorised according to the 4 levels of risk they pose. The first is unacceptable risk where systems that pose a clear threat to the safety, security or fundamental rights of people are prohibited. The second is high risk where systems may negatively affect the health, safety or fundamental rights of individuals and are subject to strict requirements before they can be placed on the market. The third one is limited risk where systems are subject to specific transparency obligations to ensure that people are informed when necessary. And the last one is minimal or no risk. Most of AI systems currently in use in the EU enter into this category and are allowed to be used freely.

The EU AI Act prohibits certain applications of AI that threaten citizens' rights, including Biometric Categorization Systems based on sensitive characteristics and the indiscriminate collection of facial images from the internet or from surveillance footage to create Databases for recognition (Lider Media, 2024). Before deploying a high-risk AI system, entities providing public services must assess the impact on fundamental rights. High-risk AI systems must be registered in the EU Database for high-risk AI systems, and users of emotion recognition systems must notify individuals when they have been exposed to such a system. Fines for non-compliance are financial amounts for companies that infringe the provisions of the EU AI Act, set as a percentage of their global annual turnover in the preceding financial year or as a pre-determined amount, whichever is the higher. As regards to Digital Archives, the EU AI Act does not contain specific provisions directly applicable to them. However, the application of AI in Digital Archives may be subject to the general rules and requirements of the Act, in particular if the systems used fall into the high or limited risk categories. This means that institutions managing Digital Archives should relate to the the risks associated with the application of AI and ensure compliance with the relevant provisions of the EU AI Act.

The first element of Machine Learning[2] is the AI algorithm. Open source Machine Learning libraries such as TensorFlow[3] i Theano[4] have removed many of the complexities of low-risk information systems involved in designing and building AI applications. These tools are free, well-described with Metadata, and supported by many active organizations. The availability of these tools has made building Machine Learning applications more popular with developers.

The second element is computing power, both in the form of raw power, Central Processing Unit – CPU, as well as large-scale Cloud Big Data storage solutions. Cloud storage services offered by Amazon Web Services, Google Cloud, Microsoft Azure and

[1] *The act is available at: https://Digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (09.01.2025.).*

[2] *More on Machine learning see: Yoosefzadeh-Najafabadi, M., Earl, H. J., Tulpan, D., Sulik, J., & Eskandari, M. (2021).*

[3] *More about TensorFlow: https://www.tensorflow.org/.*

[4] *Theano is a Python library (Original author is Montreal Institute for Learning Algorithms - MILA, University of Montreal) and optimizing compiler for manipulating and evaluating mathematical expressions, especially matrix-valued ones. In Theano, computations are expressed using a NumPy-esque syntax and compiled to run efficiently on either CPU or GPU architectures.*

many others, allow you to rent out their services with virtual machines and AI tools for Machine Training on Big Data sets (Herres, 2020, 130).

The third and final important element is Big Data sets. Before starting to create an Information System Architecture for Data, renting servers for Data storage and installing open source machine learning programs for Digital Archives, Libraries and Museums, Big Data sets must be arranged in accordance with international standards. The quality and depth of Big Data sets determine the level of AI applications that can be achieved (Ghasemaghaei, M., & Calic, G., 2020, 108).

While an organization may not be at the stage where it is ready to start building AI applications, it can always plan for a future where Big Data sets are used to power smart solutions. In the initial phase, it is necessary to treat each new initiative or project as an opportunity to build the foundation for future Big Data Models. To generate and collect well-designed Big Data, it is necessary to start with preparatory questions to guide the given goal, such as: Does the organization have a cohesive Big Data collection policy? Is Big Data collected and stored in a common format? Is Big Data stored in a central location? When Big Data is collected, is it stored in a usable format at all Data collection touchpoints? Are the field names the same? Is the same level of validation and error checking applied at all levels of Information Systems and at all entities? Are there clear guidelines on what Big Data is collected and why when a new feature is developed in the system, in accordance with General Data Protection Regulation -GDPR?[5] Does all this Big Data that is being stored have a meaningful purpose?

Big Data must be automatically stored and accessible in real time within all Business processes of an organization. As AI applications in general become better and faster at connecting disparate sources of information, isolated Big Data sets that are difficult to access become an obstacle to finding value from an organization's Data. In the not-so-distant future, what we call AI today will be naturally embedded in our business processes, and ultimately we will no longer have to emphasize it, as happened with Optical Character Recognition software - OCR. It will become common for Business processes to function with all the described AI tools. What is under control today is the Data we create and own, and therefore we need to be ready for a future in which AI applications can learn on Big Data sets, but for this the quantity and quality of Data must be standardized.

The change in the Research paradigm in the Humanities and Social Sciencess and the practice of Digitalization finds a overlap space in Multidisciplinarity, at the connection with Technical Sciences with the common point on AI and is called Digital Humanities.

Digital Humanities is a rapidly moving Scientific Discipline, and its development has been driven by the paradigm shift in Research in the Humanities and Social Sciencess through the development of Information and Communication Technologies - ICT.

The terminological change from „Humanistic Computing“ to „Digital Humanities“ was created by John Unsworth, Susan Schreibman i Ray Siemens, as editors of the anthology „A Companion to Digital Humanities“ (2004.), tried to prevent this field from to called and being seen as „just Digitalization“. Consequently, this hybrid concept has created an overlap between fields such as rhetoric and composition, which use the methods of contemporary Humanities and Social Sciencess to study Digital objects, and Digital Social Sciences and Humanities, which use Digital Technology to study traditional humanities objects, such as Digital Archives, Libraries, and Museums. The application

5 *Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.*

of computer systems and the study of computer media within the Social Sciences and Humanities, has been called the „Computational Turn“ (Berry, 2011).

In the field of Social Sciences and in the field of Information and Communication sciences, the branch of Digital Archival Science (which is a technical term introduced for the first time in a Doctoral Dissertation, Žaja L. 2020.) This is a field of practice and Science that deals with the procedures and technologies used for the Preservation and Availability of Digital documents and information. Digital Archival Scienceis becoming increasingly important as more and more information is transferred from Analog media to Digital. Digital Digital Archivists are responsible for developing processes and technologies to store information, properly archive it, and make it available in an organized manner for Research and use. These professionals work with Digital documents, such as photographs, videos, audio recordings, and documents from Databases, must be professionally trained in managing this information, as well as distributing it to users. They must also develop and implement strategies for preserving and protecting Data Information. Digital Digital Archivists must know how to plan the structure of Information Systems, develop Digital Archival Technology, and ensure information accessibility to the general public, in accordance with regulations.

Digital Archives are Databases of Digital files and documents that are stored and maintained in Digital form in the Cloud. They are created to preserve and store Information and Big Data that is kept permanently for public and state bodies and for private companies or any other organizations. They store documents, photographs, audio and video, as well as other Digital Media. They are valuable resources for Research and provide faster online access to historically significant information that is also available in Traditional Analog Archives. Digital Archives are used to store, track and share documents, as well as copies, Research paper extracts and Digital books. In addition, they help preserve and provide access to Digital Media. Digital Archives provide a service for storing and managing large amounts of Big Data, and are also used to preserve Digital Artifacts. They provide professional and advisory assistance to state and public bodies under their jurisdiction, in keeping Records of their activities, for the purpose of preserving Cultural Heritage and Historical Archival Document Management System, for future generations.

2 Digital archives

In the segment: „Europe is ready for the Digital Age - strengthening the role of citizens with a new generation of Technologies“[6] „back in time“, in 2014, the European Commission adopted a series of measures to facilitate economic development and manage Big Data sets, some of which are: Regulation on the Free Flow of Non-Personal Data, Cybersecurity Act, Open Data Directive and General Data Protection Regulation – GDPR – EU 2016/679.[7]

In 2018, the AI Strategy was presented for the first time and a coordinated plan was agreed with the Member States. In 2019, the High-Level Expert Group on AI presented ethical guidelines for trustworthy and safe AI, building on the AI legislative framework presented in 2020. The European Commission's main activities related to the Digital Age are: European Chip Act, Artificial Intelligence, European Data Strategy,

[6] *European Commsion. URL: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-Digital-age_hr (12. 01. 2025.).*

[7] *Fn. 1.*

European Industrial Strategy, High Performance Computing, Digital Markets Act, European Digital Integrity, Contribution to European Defence, Space, Digital Services Act, Cybersecurity, Digital Skills, Connectivity and the EU-US Security and Technology Council. Other links to this international framework are: Digital Strategy, Recovery Plan for Europe, European Year of Youth, Cohesion Policy and the Digital Age and Supporting Reforms for the Digital Transition. With the Digital Strategy, the European Commission aims to achieve change that is in the interest of citizens and institutions and to contribute to the goal of a “Digital Decade” for Europe by 2050. With these programmes, Europe focuses on developing Data Governance and Technological infrastructure to strengthen its Digital sovereignty and set its own standards of performance, rather than following others.

The scope of Digital Archives is to accelerate and strengthen the Digitization process and to develop the Architecture of eArchive Information Systems. By designing a Workflow, the basic goals related to the activity of Digital Archives as a whole are determined. The main task of eArchives is to make Digital archival materials available online to users, by accelerating the Digitization process such as Document Management System - DMS. Digital materials should be ensured to be permanently Preserved and usable in the long term according to the principles of Authenticity, Integrity, Readability, Confidentiality and Credibility. Of particular importance is the introduction of modern Departments, by establishing new uniform systematizations in Digital Archives. The quality of the development of the Architecture of the Information Infrastructure of eArchives, in order to avoid as much as possible the existing dangers of slow progress in Digitization and storage of materials in the computer cloud, is possible by organizing three sectors in Digital Archives (Žaja, 2023). In the first sector for Disruptive Technologies in the work process, it foresees the integration of Disruptive Technologies based on AI tools and new Business models of eArchives, creating a faster process of continuous Digitization. Although technology often dominates the transformation process, ultimately, success always depends on Digital Digital Archivists and teamwork. Developing new Business models of eArchives requires knowledge and competence in working with Big Data, Analytics and AI tools, but also professional knowledge of Archival funds in terms of organizing materials, knowledge of Archival norms and standards, and the creation of unique and harmonized norms and specifications related to the procedure and Digitization parameters. In the second sector, computing and Information and Communication Technologies - ICT in the work process foresee virtual Big Data transfer and the availability of all Digital information in the computer Cloud (Digital Records and Born Electronic Records) and constitute the most pervasive generic Technology of today. As such, it is the foundation of 21st century society. The aforementioned Department enables the solution of complex Engineering problems for the design of complex Systems and the Management of Research and Development teams. The main purpose of the Department is to work on computers for Information processing with scanners for Digitization, as well as methods of its application in Archival activities. The Department of Computing encompasses Archival theory and practice, methods of Publishing organized material, design and construction of Chatbots, application and operation of computer Systems in eArchives. In this segment, complex Engineering problems are solved, Information Systems are designed, teams are coordinated to improve the Archival esystem, and opportunities that keep pace with new AI Technologies are explored and developed. In the third sector for Promotion, Scientific Research projects and publishing, the work process envisages the promotion of eArchives as the information center of Big Data. The organization of Educational programs and Workshops is encouraged and carried out, Cultural activities are carried out in cooperation with other Heritage and Scientific Institutions, and participation in the planning of Scientific and professional eArchive projects is carried out. It also envisages Multidisciplinary cooperation and publishing activities with the task of organizing, unifying and monitoring the overall Scientific activity of the eArchievs. Activities are particularly focused on Research, Scientific evaluation and publication of Archival materials and Scientific papers of an Multidisciplinary nature, as well as organizing and participating in World Scientific Conferences. It encourages and implements Scientific projects, cooperates with domestic and international institutions of an Multidisciplinary scope, and initiates and coordinates the publication of a future Multidisciplinary AI Scientific Journal (Žaja, 2021).

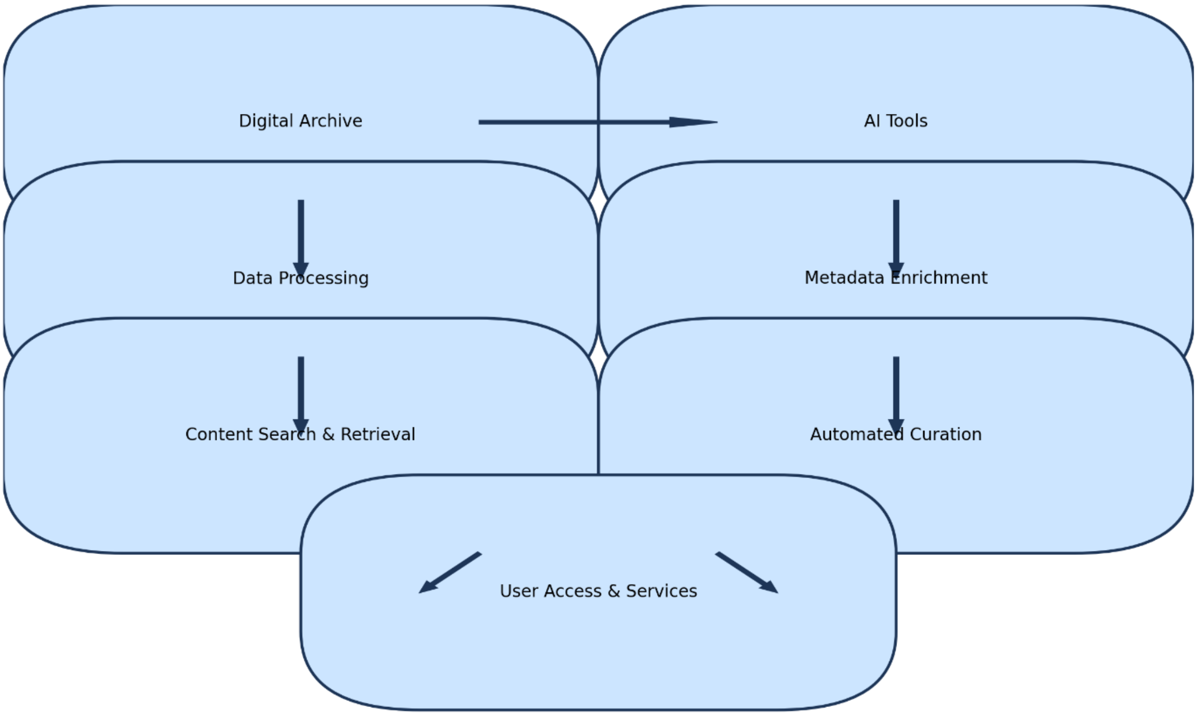
The establishment of modern Departments through cooperation with Universities on Scientific Research Projects, professional student internships, lifelong education of Archival workers and educational workshops contributes to the stable development of eArchives. Due to the accelerated changes in Technological processes, on Social and Political plans that are dynamically implemented at the EU level and in the world in general, the strategic directions of Scientific Research development of AI in eArchives must necessarily continuously adapt to Technological progress in the field of Archival activities, both in Archival theory and practice, and in Archival Science as a Scientific discipline that has its foothold in the legislative framework. In order to improve eArchives, an approach and expectations exist in the context of using modern technologies and AI tools.

3 AI in digital archives

AI in Digital Archives refers to a range of Technologies that can operate separately, but in accordance with the needs of Digital Archives, and together. From pattern recognition as part of computer vision technology, through the recognition of named entities from text (personal names, names of organizations, names of locations, dates, times, financial amounts or meters long) to issues of Classification, Natural Language Processing – NLP, Machine Learning and its subtype Deep Learning. A good part of these technologies uses Artificial Neural Networks. Advanced technologies are gradually becoming an integral part of Digital Archive applications, because the perception of Digital Archives in the world implies the management of virtual Databases and the delivery of requests in the shortest possible time. With the help of AI tools, large amounts of Big Data can not only be searched faster, but also in-depth Content Analysis can be carried out, which is an ideal Data source for visualizing groups of Content Semantic Networks. For example, a virtual tour of the object and representative examples of Digital Archive materials to an interactive guide for studying Archival materials. All of the above are mandatory functionalities that become available to Digital Digital Archivists/Information Experts and, together with Researchers and members of the academic community, can be developed, adapted to user requirements and modern Archival sources, but also to those materials that still need to be edited in accordance with standards before the start of Digitization. Digital sources and original Electronic Records are used by Scientific Researchers, especially historians, just as they are used with paper (Analog) materials and materials. Modern AI tools adapted to Digital Archival materials are necessary today, in order to ensure the greatest possible use and faster availability of Analog and Digital Documents (Russel, Norvig, 2021).

For the comprehensive functioning and achievement of the set goals of Digital Archives, Human Resources should be provided, which are set according to the Multidisciplinary principle, in accordance with the financial plan in accordance with national and international frameworks. Employees should be permanently encouraged to undergo professional and higher Education Training in order to meet the accelerated Technological development in achieving the set goals with their knowledge, skills and competencies.

The fundamental vision of Multidisciplinary Scientific Research project cooperation is reflected in the Technological Development of eArchives in cooperation with universities in the fields of Technical and Social-Humanities Sciences. Scientific contribution to the academic community is brought by including Technological-theoretical and practical knowledge in the Multidisciplinary environment of eArchive Business processes. Research into Information and Communication solutions based on AI improves existing eArchive systems. The purpose of designing lectures and workshops is to combine Technological and Humanities and Social Sciences multidisciplinarity in Lifelong Education. Special attention should be on students whose final and diploma theses are based on the Development of Information Systems and the implementation of AI in eArchives, because quality studies with concrete content solutions that scientifically contribute to the understanding of Humanities and Social Sciences and Technical aspects and have implementation solutions of AI tools in the Business processes of eArchives can be an integral part of future Scientific papers in AI Multidisciplinary journals.



Picture 1. Implementation of AI Workflow in Digital Archives 2025.

The main goals of Scientific Research and project cooperation related to the implementation of AI tools in eArchives are virtual improvement and creative design of websites that meet the know-how needs of Information and Communication activities, which especially refers to the accessibility of materials to the end user. Analysis of Big Data sets is necessary to achieve the goals and lay the foundation for their full use for browsing and searching Archival funds. The development of a robust analytical solution must meet the criteria of Business processes for Research analysis and large organized Big Data sets, such as Digitized funds. Also important is the creation of interactive dashboards related to virtual previews of the desired content and the creation of predictive models of user behavior of eArchive websites, which implies measuring successful visits to the website and Social Networks on the website. Validation of the created solution and Analysis of the implemented approach with reference to the project goal is a solution that meets the criteria of the user request. Support and maintenance of virtual solutions includes solving unforeseen challenges and providing support throughout the entire project development cycle, which are carried out by Technical Universities. Designing and implementing custom Chatbots[8] solutions include multiplatform and multilingual virtual Chatbots that respond quickly to targeted user queries. Identifying the „pain points" that the Chatbot solves and defining the values that eArchive users gain serves to set precise expectations and recognize the limitations of the Chatbot. It is important to detect the „personalities" of the Chatbot with respect to the target user group of the eArchive and to determine the functionality of the Chatbot and the Big Data it collects. It is necessary to introduce translation of program functionalities for euser queries and prediction of possible questions and appropriate answers in as many languages as possible. This segment also includes the human factor working on writing alternative and different endings of conversations, which would be incorporated into AI Systems. The conversation flow must be implemented by converting the diagram into a AI Workflow, integrating it with Business applications, testing the created Chatbot, and creating a final solution in the cloud, which will enable insight, monitoring, and execution of conversations with the Chatbot, as well as adapting the virtual knowledge base to the Chatbot (Gupta, 2023).

AI is an extremely useful tool in STEM-u[9] and Multidisciplinary Sciences, with which we can detect emerging problems in eArchives faster and easier and offer constructive and quick solutions. Indicators of results are reflected in the increased number of educated Digital Archivists/Information Experts with specific knowledge and skills, as well as the number of students of History, Information and Communication Sciences (Archival studies, Library Science and Informatics) and system and Software

[8] *Modern Chatbots are typically online and use generative artificial intelligence systems that are capable of maintaining a conversation with a user in natural language and simulating the way a human would behave as a conversational partner. Such Chatbots often use deep learning and natural language processing, but simpler Chatbots have existed for decades. Although Chatbots have existed since the late 1960s, the field gained widespread attention in the early 2020s due to the popularity of OpenAI's ChatGPT, followed by alternatives such as Microsoft's Copilot and Google's Gemini. Such examples reflect the recent practice of basing such products upon broad foundational large language models, such as GPT-4 or the Gemini language model, that get fine-tuned so as to target specific tasks or applications (i.e., simulating human conversation, in the case of Chatbots). Chatbots can also be designed or customized to further target even more specific situations and/or particular subject-matter domains.*

[9] *STEM: Science, Technology, Engineering, and Mathematics – science in STEM it usually refers to two of the three main branches of science: the Natural Sciences, including Biology, Physics, and Chemistry, and the Formal Sciences, exemplified by Mathematics, along with Logic and Statistics. Despite this definition, Psychology is, however, considered a major part of STEM. The movement was adopted by the National Science Foundation in the United States in 2001.*

Engineers as participants in the design and strengthening of Information Systems Architecture and the implementation of AI in eArchives.

The planned Information Architecture of Data Centers enables Big Data extraction, loading, transformation and storage (Belov, Kosenkov, Nikulchev, 2021). Visualization and Analysis of all Data using Machine Learning should be based on AI tools. By establishing and upgrading the Big Data Lake Digital Archives and creating the Data WareHouse Data Storage Platform, it is necessary to form six key segments. The first is the storage of (classified) inaccessible Data, which can be approved, in accordance with the law, to natural and legal persons and specialized state administration bodies, where analytics based on Data Analytics, processing and visualization tools would be carried out for the purpose of making decisions and policies based on Big Data Driven Policy Making. The second is the storage of (unclassified) accessible Big Data, with implemented analytical tools and elements of AI, equally accessible to both the private and public sectors and the academic community for the development of new services that create added value and economic benefit for society. The third is that the roadmaps on the obligation of all Digital Archives to create e-catalogues of Big Data for users and Metadata for Information professionals, for each individual process and software solution, as well as the obligation to publish this Big Data on the Open Data Portal should be implemented in a disciplined manner. The fourth is that in accordance with the EU recommendations for the free flow of non-personal Data in the EU and the European Interoperability Framework, it is necessary to regularly publish Records and lists in Machine Readable Formats, with protected proprietary rights and with appropriate Metadata on the central Open Data Portal, thus contributing to interoperability and simplification of Data Science, including cross-border. The fifth is the continuous monitoring of the possible adoption of regulations that allow the transfer and processing of Data in accordance with the EU directive: „Framework for the Free Flow of Non-Personal Data“[10] thus contributing to the achievement of the EU's goal: „Creating a single Data market“. And the last, sixth, is that Open Data Policy should encourage more intensive cooperation between public administration and the private sector in the field of Information Technologies, improve Digital Public Services, and increase the transparency of public administration.

Digital Archives include Digital Preservation, additional strategies for preserving Digital Objects, and Digital Preservation challenges. Digital Preservation involves a formal effort to ensure that Digital Information of lasting value remains accessible and usable. Such assurance includes planning and various strategies to ensure access to reformatted and original Digital content. Furthermore, the key planning parameters are: Appraisal, then Data Integrity, defining what Data is, describing Data accuracy checks in which Data must not change its Authenticity and Authentication. Standards for Digital Preservation containing preservation Metadata and implementation strategy are. Preservation Metadata: Implementation Strategies – PREMIS[11] and Open Archival Information System – OAIS.[12] Specification of the PREMIS function based on the OAIS reference model that emphasizes the significance Open Archives Initiative Protocol for

[10] *Available at: https://Digital-strategy.ec.europa.eu/en/policies/non-personal-Data.*

[11] *See more: https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.*

[12] *See more: Lee, C. A. (2005). Defining Digital Preservation Work: A Case Study of the Development of the Reference Model for an Open Archival Information System. Thesis. The University of Michigan. hdl:2027.42/39372. Available at: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/39372/dissertation_callee.pdf?sequence=2&isAllowed=y.*

Metadata Harvesting – OAI-PMH[13], that is, a protocol developed for the collection of Metadata with Descriptions of Records in Digital Archives, based on international standards. Additional strategies for Digital are: Refreshing Data, Migration Data, Replication Data, Emulation Data, Encapsulation Data, Persistent Archives and Metadata. Research in the field of Authenticity of Records in Electronic Systems is based on a development project called International Research on Permanent Authentic Records in Electronic Systems – InterPARES.[14] The challenges of Digital Preservation are defined by the technical term Digital Obsolescence, which means that a Digital resource is no longer readable due to its archaic format. By pointing out other challenges related to the Preservation of Digital content that are found in broad issues such as, for example, the economic challenges of Digital Preservation, it is possible to ensure stable conditions for Information Systems with the necessary programming knowledge. In order to build stable Preservation Programs, it is necessary to ensure significant financial investments, together with the accompanying operational running costs for Data entry and transfer, Data Management System - DMS, Big Data storage and quality Multidisciplinary professional staff. Professionally trained employees inevitably entail issues of education of Digital Archivists/Information Experts for Digital Preservation who must be able to create the Architecture of Information Systems and, in cooperation with technical associates, create it and put it into stable operation.

4 Data mining

Data Mining is the process of Analyzing large amounts of Big Data. It is used to discover patterns and trends in Big Data sets to help companies make informed decisions. The analysis involves the application of various AI techniques such as Machine Learning and Statistical Analysis. It can be used to discover customer behavior, improve customer service, and help companies make more accurate decisions (Fayyad, Piatetsky-Shapiro, Smyth, 1996). For example, a multinational interdisciplinary project InterPARES Trust AI[15], for the period from 2021 to 2026, aims to design, develop and use AI for the availability and accessibility of reliable public records by forming a sustainable, permanent partnership that supports original Research, teaches students and other highly qualified experts and creates feedback loops between the Academic community, Archival institutions, State experts for Records and Industry. The goals of ITrust AI are the identification of specific AI Technology that can respond to the critical challenges of Archival Records, determining the advantages and risks of using AI Technology on Archival Documents, ensuring that Archival concepts and principles develop AI tools, and confirming the achieved goals through Analytical presentation studies.

The actual task of Data Mining is the semi-automatic or automatic analysis of large amounts of Big Data to extract previously unknown, interesting patterns such as groups of Data Records (Cluster Analysis), Fuzzy Records (Anomaly Detection) and Dependencies (finding association rules with sequential pattern mining). It usually involves the use of Database search techniques, such as spatial indexes. Spatial index patterns can then be viewed as a kind of summary of the input Big Data and can be used in further analysis. For example, in Machine Learning and Predictive Analytics, a single

---

13 *See more: Lynch, C. A. (2001). Metadata harvesting and the Open Archives Initiative“. URL: A Bimonthly Report (217). URL: https://www.cni.org/wp-content/uploads/2001/08/Metadata-Harvesting-and-the-Open-Archives-Initiative.pdf (13.01.2025.).*

14 *See more: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_complete.pdf.*

15 *See more: InterPARES Trust AI (2021. – 2026.). URL: https://interparestrustai.org/trust.*

step of Data Mining can identify multiple groups in the Big Data, which can then be used to obtain more accurate prediction results using Information Systems for rapid AI decision support. Neither the collection of Big Data, nor the preparation of Big Data, nor the interpretation of the results are part of the steps of Data Mining, although in some way they belong to the overall process, as additional steps.

Data Optimization is the process of improving the efficiency of Data Storage and Data access by applying various methods such as Indexing, Compression, and Normalization. Data Optimization helps reduce the amount of Big Data stored in a Database and speeds up Data retrieval and manipulation. It also helps improve Big Data quality by ensuring that Data is accurate and consistent. Data Optimization AI techniques can also help reduce the costs associated with Big Data storage and retrieval (Papalambros, Wilde, 2017).

Today, AI transforming entire industries, conglomerates and institutions. Ever cheaper, but very often freely available and easy-to-use Technological tools offer Digital Information Professionals/Digital Archivists almost unlimited creative potential in terms of challenges of evaluating, selecting and reviewing the sensitivity of born (authentic) Digital Records that can only be solved with the help of AI. Digital Archives are just one segment of the „Technological Revolution", whose improved capabilities are triggered by online access to large amounts of Big Data and Disruptive Technologies and AI tools that change the way we think about everyday Business tasks and Social Communication. Far from the hustle and bustle around big news headlines that promise new AI Algorithms every day for example: driving a car, automatic writing (by pronouncing words by voice) or machine translation of texts into any world language, and even the creation of artistic masterpieces, the existing „AI revolution" is already happening before our eyes in an increasingly present and subtle way.

In this way, „AI revolution" will transform Traditional archives and present the challenges and opportunities of Open Source Digital Archives to the information market, which will have to introduce new ethical and Legal Norms of Digital Legislation and standards for such a role, implementing them mandatorily in the Legislation on AI.

Access to Digital Archives can be optimized by semantically representing documents and artifacts, making it possible to more efficiently find information and explore it in the eArchive. Supporting user interaction with the system in more meaningful and simple ways by introducing new AI Algorithms that are capable of extracting and combining Big Data from structured and unstructured documents will provide new levels of knowledge and scientific discoveries. Semantic Web Technologies and Ontologies can help represent the level of confidentiality of Documents/Text/Data and support access and Authorization for automated Records Management. The next step is Technological visualization and intuitive virtual associations that can provide greater support to users during the exploration of Born Digital Collections, both Digital Information Professionals/Digital Archivists and those who manage the Decision Making Process. And the last stage is Information-Communication interaction based on conversations in Natural Language with Chatbots, and not through fixed boundaries for searching (typing key terms in the provided toolbars), will enable greater online usefulness of Digital Archives to the virtual user, far from the standard setting of a dry desktop, paving the way for a more natural use of these collections in virtual augmented reality.

The application of AI tools to the analysis of Big Data sets that need to be optimized has opened up a new field with a whole range of interesting possibilities, but it also has pitfalls, such as: one-sidedness, subjectivity, privacy and security issues, as well as a lack of transparency and explanation. These pitfalls are just some of the very predictable risks associated with AI, which usually arise from unpredictable problems when introducing any new Technologies. For example, Data Generalization can make it easier to analyze and identify trends and patterns in the Big Data. Data Generalization is the process of reducing the amount of Big Data by removing or aggregating information that is not relevant to the analysis. It involves summarizing the Big Data so that the most important information can be used without having to consider all the details. For example, a Big Data set from accounting purchase orders can be generalized by aggregating the Big Data into categories such as what type of document it is, how much of the documents we have, and what is the cost of such an extensive Research.

For example, Data Generalization in Digital Archives can be achieved through various AI methods. These include Data Compression, Transformation, and Data Summarization. Data Compression is the process of reducing the size of a file or Data set while preserving the original content. Data transformation is the conversion of Data into a different format or structure and is performed for Preservation and easier use. The aforementioned Data Summarization is the process of extracting the most important information from a Big Data set and presenting it in a concise and easy-to-understand form. These methods can be applied to reduce the amount of space required to store Data in Digital Archives, making them more efficient in terms of greater capacity for storing Electronic Records and Digitized materials. Additionally, Data Generalization can aid in the Analysis and Analytics of Big Data sets, as it simplifies the complexity of the Data and makes the results easier to interpret.

The success of AI Generalization relies on translating human knowledge of Manual Generalization techniques into explicit rules and logic, in such a way that they can be encoded in AI computer language. There are very few manual guidelines for Generalization in textbooks, and these are usually vague and incomplete. The concept of AI Generalization has wide application in many related disciplines, and sometimes has special context-specific meaning. AI Generalization can consolidate Data and display the most important information. For any two related concepts, e.g., concept A and concept B, concept A is considered a Generalization of concept B if and only if, every instance of concept B is also an instance of concept A. However, there are instances of concept A that are not instances of concept B. For example, an Electronic Record may be a AI Generalization of a Digital Archive because every Record is stored in an Archive Information System, but there are also those Electronic Records that are not part of the whole Digital Archive, such as, for example, the financial-accounting e-invoices of a private company. The difference between Optimization and Machine Learning stems from the goal of AI Generalization is: while Optimization AI Algorithms can minimize loss on a Machine Learning Data set, Machine Learning is concerned with minimizing loss on unseen samples of a Big Data set. Characterizing the AI Generalization of different learning AI Algorithms is an active topic of current Research, especially for Deep Machine Learning AI Algorithms.

Data Statistics in AI is a branch of mathematics and applied statistics that deals with the Collection, Analysis, Interpretation, and Presentation of Data. This field provides a wide range of AI techniques for summarizing and analyzing Data, such as Descriptive Statistics, Probability Distributions, Regression Analysis, and Hypothesis Testing. Data Statistics can be applied to determine the nature and extent of relationships between variables, as well as to identify patterns and trends in Data sets. It can also be used to predict future events and trends. Data Statistics is an essential AI tool in many fields, including economics, finance, and scientific Research. Data Statistics in Digital Archives may include the number of Digital and Analog documents, Metadata, Entities, Electronic Records, the size (quantity) of the Data set, the number of users, the average time users spend Searching and Researching information, and the average number of items viewed or downloaded from the eArchive. Depending on the Archival Information System Workflow AI Model, other statistics may also be available, such as the number of images, videos, or audio files, the number of visitors to a website, the number of items viewed or downloaded by each user individually, and the average number of items viewed or downloaded in a day, week, month, and year. In addition, Data Statistics may include the number of searches performed and the average time it takes to find a specific item in the Archival Information System.

Machine Learning and Data Statistics are closely related fields in terms of AI methods, but different in their main goal. Statistics results in conclusions about the population from a given sample, while Machine Learning finds predictive patterns that can be AI Generalized across a given Big Data set. According to Michael Irwin Jordan, (1956) the ideas of Machine Learning, from methodological principles to theoretical tools, had a long history in statistics. He was also the first to introduce the term „Big Data", which is now adopted as „Data Science"[16] as the official term for a Scientific field. (Jordan, 2014) Some statisticians have adopted Machine Learning methods, resulting in a complex Scientific field called "Statistical Machine Learning" (Gareth, Witten, Hastie, Tibshirani, 2013).

5 Machine learning

Machine Learning is a subfield of Computer Science that uses AI algorithms to learn from Big Data and predict future events. It allows computers to quickly and accurately process large amounts of Data and recognize patterns that are often not visible. With the help of Machine Learning, computers can „autonomously" learn from their mistakes and improve their own capabilities and performance over time. This technology is most widely used in many areas, most often in Finance, Healthcare, and Robotics. Machine Learning programs can perform tasks without being explicitly programmed to do so. This involves computers learning from Big Data sets provided to them to perform specific tasks. For simple tasks assigned to computers, it is possible to program AI algorithms that "tell" the machine how to perform all the steps needed to solve the problem. In this case, the computer does not need to learn. For more advanced tasks, it can be challenging for programmers to manually create the necessary algorithms. Today, practical examples have shown that „speaking" to a computer in developing its own algorithms is more effective than human programming (Ethem, 2020, pp. 1-3, 13-18).

The discipline of Machine Learning uses various approaches to teach computers to perform tasks where no fully satisfactory AI Algorithm is available. When there are a large number of potential answers, one approach is to mark some of the correct answers as valid. The correct answers are then used for Machine Learning to improve the AI Algorithms used to determine the correct answers. For example, to train an Information

[16] *Data Science is an Interdisciplinary field that uses Scientific Methods, Processes, Algorithms, and Systems to extract knowledge and insights from structured and unstructured Big Data and apply that knowledge.*

System for a specific task of Digital Character Recognition, MNIST[17] is often used as a Big Data set of Handwritten Digits.

Some implementations of Machine Learning use Data and AI Neural Networks in a way that mimics the workings of the biological human brain (Zhou, 2020). In its application to Business problems, Machine Learning is also called Predictive Analytics.[18] Supervised Machine Learning is a Machine Learning method that uses Data with known outcomes. Using known inputs and expected outputs, it is possible to build or train a AI Computer Model, which can then be used to predict similar problems. The model is then tested on new Data to see how accurately it predicts outcomes. Supervised Learning AI algorithms can be applied to a wide range of tasks, such as classification, regression, and image recognition.

Supervised Machine Learning in AI can be applied to Digital Archives in a variety of ways. For example, automating tagging can be used to streamline the Indexing, Categorization, Classification, and Cataloging processes. In addition, supervised Machine Learning can be used to detect unnecessary duplicates or similar items approved for destruction in a Digital Archive, ensuring that there is no duplication of entries and that cloud storage space is not wasted. Supervised Machine Learning can also be applied to detect and remove malicious content in a Digital Archive, helping to protect against potential cyber security threats. Unsupervised Machine Learning in AI is a type of Machine Learning in which algorithms independently find features of patterns. The goal is to discover previously unknown patterns and group similar groups of Big Data. Examples of Unsupervised Learning include clustering, anomaly detection, and dimensionality reduction. Clustering AI Algorithms group similar patterns, while anomaly detection AI Algorithms detect deviations from a set norm. Dimensionality reduction algorithms reduce the number of variables in a Big Data set while preserving important information.

In Digital Archives, AI can be applied to discover patterns and clusters within Digital collections. It can also be used to analyze Big Data, discover patterns, and make predictions. It is a good tool for discovering hidden relationships between Big Data and making recommendations for future Digital collections, as well as for detecting trends and anomalies in Digital Archives, which can be used to extract meaningful information from Big Data and further applied to activities such as comparative analysis and content analytics, Metadata extraction, and text mining. Unsupervised Machine Learning can be used to build new Digital collections from existing Digital Archives by automatically grouping Digital resources into significant collections. In addition, this type of AI Learning can be used to program Automated Metadata for categorizing Digital resources, which can help improve the accuracy of searching and retrieving records from Digital Archives. Unlike Supervised Learning, where a AI computer Model is built (learned) based on known Data, Unsupervised Methods exhibit self-organization that evaluates patterns through Statistical Methods or by training AI Neural Networks. Semi Supervised Machine Learning is a form of Machine Learning that is typically applied when limited Data Sources are available. In this type of AI Learning, a model is trained on known Data and its outcomes to learn relevant features of the problem, and then used to learn examples from the problem domain, without a known outcome, usually for the purpose of AI

[17] *Modified National Institute of Standards and Technology – MNIST is a Database of handwritten digits that is commonly used for training various image processing systems, and it is also widely used for training and testing in the field of Machine Learning.*

[18] *Predictive analytics encompasses a range of statistical techniques from Data Mining, Predictive Modeling, and Machine Learning that analyze current and historical facts to predict future or otherwise unknown events.*

Generalization. Partially supervised Machine Learning can help reduce the time and cost associated with Big Data labeling (associating input and output quantities/terms), as well as improve the accuracy of the AI Model.

This type of AI Learning, using AI tools, is increasingly popular in Aigital Archives because it helps to accurately process Big Data, which involves using algorithms to detect patterns in the Data and then make predictions based on those patterns. Partially Supervised Machine Learning in Digital Archives can be used to automatically categorize Data, identify relevant information, and predict which types of records might be most useful for a particular task. For example, it can be used to automatically tag documents and images, classify and retrieve images, and extract text from documents. It can also, like other Machine Learning, be used to detect anomalies or deviations in Data, which can help alert Digital Archivists and Information Professionals when deviations from common Protocols in Information Systems occur. It is also used to improve the accuracy of information searches in Digital Archives and provide more efficient search results, help identify upward and downward trends in Data searches, and suggest new ways to organize Records in Databases. It refines the accuracy of Data analysis, enabling Digital Archivists and information professionals to better understand the applications in which Digital collections are used and to make more effective decisions for the overall development of Information Systems based on feedback from the system itself.

Reinforcement Machine Learning is a type of Machine Learning that allows machines and software agents to find solutions to problems by learning from their own mistakes and successes. Unlike supervised learning, in which an agent is given a set of Big Data about the problem to be solved and instructions on how to solve it, reinforcement learning involves the agent continuously exploring and developing its „behavior" based on an assessment of its success in approaching the solution. This type of Machine Learning technique uses feedback loops to learn and adapt to a given problem. It can be applied to optimize Digital Archives by AI Learning from user feedback and automatically adapting the system to better serve given search queries. For example, it can improve search results, more accurately recommend relevant content, and optimize Digital content delivery to users. In optimizing the workflow of Digital Archives by AI Learning from the interactions of given queries and automatically adapting the workflow, the efficiency of the Information System can be improved. Finally, in order to improve the security of Digital Archives, by AI Learning from user requests and automatically adopting the Information System to queries, better protection of Digital content can be automatically programmed. Reinforced Machine Learning focuses on finding a balance between exploring unexplored territory and exploiting existing knowledge (Kaelbling, Littman, Moore, 1996).

6 Language models

Language Models are a type of AI Statistical Model used in Natural Language processing to predict the probability of occurrence of certain words or phrases in a given context. They can be used to help machines understand the meaning of human language, as well as to improve the accuracy of Machine Learning AI algorithms. Language models are commonly used in applications such as Natural Language Processing, Machine Translation, and Speech Recognition. They can also be used to generate text, as well as to generate new words or phrases that are similar in meaning to existing ones. The Scientific definition is that: „The Language Model is a Probability Distribution of Strings of Words." (Jurafsky, Martin, 2021)

Digital Archives use AI Language Models to improve the Retrieval of Records. Language Models are AI Algorithms used to analyze Natural Language and understand the context of documents as well as the general terms of a query. This helps the information system better match queries to relevant documents. For example, Language Models can identify synonyms, dialects, and other related terms that can help the search engine better understand the intent of a given query. In addition, Language Models can recognize and correct misspelled words, which results in improved accuracy of search results. Language Models also help reduce the amount of „noise" (Unnecessary Language Patterns) in search results by filtering out irrelevant content. These Models enable Digital Archives to provide accurate and relevant search results.

AI Language Models generate probabilities by training text corpora in one or more Languages. Since Languages can be used to express an infinite variety of valid sentences, which is also a feature of AI Digital Infinity. Language Modeling faces the problem of assigning probabilities to Linguistically valid sequences that may never be found in the training Big Data. Several modeling approaches have been devised to overcome this problem, such as applying the Markov assumption or using Artificial Neural Architectures such as Artificial Recurrent Neural Networks or AI Transformers. These Models are also useful for a variety of problems in Computational Linguistics. From initial applications in Speech Recognition, to ensure that meaningless strings of words are not included in the Training Set - TS, to broader uses in AI Machine Translation. For example, in scoring Language translations, generating speech that is as human-like as possible, part-of-speech labeling, parsing the grammar of a given Language, Optical Character Recognition - OCR, Handwriting Recognition, Grammatical Induction, Information Retrieval, and other application solutions.

All AI Machine Learning Models require a large amount of reliable Big Data to make accurate predictions about the meaning of words. When training a Machine Learning or Language Learning Model, Machine Learning engineers and programmers need to collect a representative sample of Big Data. The Data in the Training Set can be as diverse as large Big Data sets collected from individual users requesting a service. Overfitting is something to watch out for when training a Machine Learning Model. Trained Models derived from biased or under-estimated Data can result in distorted or unwanted predictions or interpretations. Biased models can have harmful outcomes and thus deepen negative effects on expected outcomes. Given the issue of excessive and unnecessary amounts of Data, for fear of missing something important, the ethics of Machine Learning is becoming a separate area of study and is being integrated into AI Systems, with carefully selected expert teams of Machine Learning engineers, programmers and AI experts.

Building a AI Machine Language Model in Digital Archives with AI tools, requires several steps. First, a specific set of Data needs to be collected from the Digital Archive. This Data can include text, images, audio, and video. Once the Data has been collected, it needs to be edited and prepared for analysis. Data preprocessing includes tokenization, lemmatization, and normalization. Once the Data has been preprocessed, it can be used to build a Machine Language Model. Depending on the type of Data, different types of models are applied. For example, if the Data is text, a Neural Network Model is applied to generate a vector representation of the text. If the Data is images, a Convolutional Neural Network – CNN model is most often used. Once the model is built, it is ready to generate predictions on new Data. The accuracy of the predictions will depend on the quality of the Data and the complexity of the model. Finally, the model can always be further trained, validated, and upgraded to ensure that it works at its highest potential.

Many Business Processes have become simpler as a result of the introduction of new and inventive software thanks to advances in technology. The creation of AI transcription software is one of the more significant results of Machine Learning and Language Models combined with Natural Language Processing – NLP (Goldberg, 2016). This model allows computers to understand Natural Language in the way humans understand and experience it. Whether it is spoken or Written Language, Natural Language Processing uses AI tools to take input of signs and symbols from the real world, process them, and give a meaningful way for the computer to process them. Just as humans have various sensors, or senses, such as ears for listening and eyes for seeing, computers have algorithms for reading and microphones and speakers for producing sound. And just as humans have a brain for processing input of experiences and knowledge, computers have a program for processing incoming information. At a certain point in the processing, the input is converted into programmed code that the computer can understand. Transcribing audio and video recordings has become much easier and more productive thanks to transcription tools based on AI. In the past, transcribing audio recordings required a lot of time and effort, especially when it came to large audio files. After listening to a certain segment of the audio file, the recording would be stopped and then restarted, and the spoken words would then be written down (or typed on a typewriter). However, this process, in addition to being time-consuming, was laborious and error-prone. Transcription software uses information gathered by analyzing the audio recording, isolating different speech patterns (idioms, repeated words, pauses, etc.), before transcribing the spoken words into text. The level of precision achieved by AI tools is equivalent to human transcription.

NLP combines Computational Linguistics, modeling human language based on grammatical rules, with Statistical, Machine Learning, and Deep AI Learning Models. Together, these technologies enable computers to process human language in the form of text or voice Data, understanding its full meaning, along with the intent and sentiment of the speaker or writer. NLP powers computer programs that translate text from one language to another, respond to spoken commands, and quickly summarize large amounts of text, even in real time. Today, we all interact with NLP in some way, in the form of voice-controlled GPS systems, Digital assistants, speech-to-text dictation software, customer service Chatbots, and other conveniences. But NLP is also playing an increasingly important role in Business Solutions, helping to streamline Business Processes, increase employee productivity, and streamline key business functions and activities. Today, some of the most popular speech-to-text recognition tools include: Rev, Otter.Ai, Airgram, GoTranscript, SpeakAI, Deskript, Amberscript, Fireflies, Sonix, Scribie, Notta, Transcribe, Verbit.Ai, OTranscribe, Trint etc.

This advancement in AI tools has enabled not only the conversion of speech to text and vice versa, but also the use of AI tools to translate interviews from one language to another. By applying NLP AI algorithms which can interpret and understand the spoken words in an interview, and then generate a translation in the target language. AI based translation tools help speed up the interview translation process, reducing the time and cost of manual translations. In addition, AI can detect errors in translations, assisting in ensuring accuracy and consistency. AI will continue to develop and become increasingly sophisticated in the next few years. AI systems are likely to become more present in our daily lives and will serve to automate many routine tasks and processes. AI driven Technologies will become more advanced and powerful, making them even more useful. As AI continues to develop, it may become increasingly difficult to define the boundaries between human and machine intelligence (Human Brain – Computer). AI can also be used to solve difficult and complex problems, such as climate change and earthquakes. In the near future, AI may even be capable of making decisions that exceed the capabilities of human intelligence. Ultimately, the potential applications of AI are still unknown today and continue to be explored.

7 Conslusion

The application of AI in Digital Archives brings revolutionary possibilities for the management, organization and analysis of large amounts of Big Data. Technologies such as Machine Learning, pattern recognition and Natural Language Processing enable more precise indexing, automation of routine tasks and improved accessibility of information to users. In this way, AI not only increases the efficiency of Archival Systems, but also contributes to the preservation and contextualization of Cultural and Historical Heritage. However, the application of AI tools comes with a number of challenges. It is crucial to ensure respect for Ethical principles, protection of privacy and transparency in the use of AI tools. In the context of regulations, such as the EU AI Act of 2024, Archival institutions need to align their practices with prescribed standards to ensure the responsible and lawful use of AI Technologies. The future of Digital Archives lies in balancing Technological innovation with the preservation of fundamental Human Rights. With a responsible approach, AI has the potential not only to improve the way we manage information, but also to make historical resources more accessible and relevant to a wider audience. Further Research and Multidisciplinary collaboration will be key to maximizing the benefits of this Technology and minimizing potential risks.

8 Sources and literature

Belov, V., Kosenkov, A. N., & Nikulchev, E. (2021). Experimental Characteristics Study of Data Storage Formats for Data Marts Development within Data Lakes. In Applied Sciences (Vol. 11, Issue 18, p. 8651). MDPI AG. https://doi.org/10.3390/app11188651.

Berry, D. (2011). The Computational Turn: Thinking About the Digital Humanities. Culture Machine, vol 12. Available at: https://web.archive.org/web/20131012010142/http://www.culturemachine.net/index.php/cm/article/download/440/470.

Ethem, A. (2020). Introduction to Machine Learning. (Fourth ed.). Massachusetts : Institute of Technology - MIT.

European Commission. URL: https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_hr?utm_source=chatgpt.com (06.01.2025.).

European Commission. (s. d.). AI Act. Available at: https://artificialintelligenceact.eu/ai-act-explorer/.

Regulation (EU) 2016/679 of the European parliament and of the Council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016). Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

Fayyad, U., Piatetsky-Shapiro, G,., Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence. Available at: https://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf.

Gareth J., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer. URL: https://hastie.su.domains/ISLRv2_website.pdf (10. 01. 2025.).

Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. In Journal of Business Research (Vol. 108, pp. 147–162). Elsevier BV. https://doi.org/10.1016/j.jbusres.2019.09.062.

Goldberg, Y. (2016). A Primer on Neural Network Models for Natural Language Processing. Journal of Artificial Intelligence Research, 57. AI Access Foundation. Available at: https://web.archive.org/web/20180421110114/https://www.jair.org/media/4992/live-4992-9623-jair.pdf.

Gupta, A. (2023). GPT-4 takes the world by storm - List of companies that integrated the Chatbot. MINT. Available at: https://www.livemint.com/technology/tech-news/gpt4-takes-the-world-by-storm-list-of-companies-that-integrated-the-Chatbot-11679293885004.html.

Hardy, P., Lee, D., (2005). GIS-Based Generalization and Multiple Representation of Spatial Data Proceedings, ISGI 2005, International Symposium of Generalization of Information. Available at: https://www.Researchgate.net/publication/228695100_GIS-based_generalization_and_multiple_representation_of_spatial_Data.

Herres, D. (2020). Oscilloscopes: A Manual for Students, Engineers, and Scientists. Springer International Publishing. https://doi.org/10.1007/978-3-030-53885-9.

Duranti, L., Preston, R. (2008). InterPARES 2 Project. Padova: Associazione Nazionale Archivistica Italiana. Available at: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_complete.pdf.

InterPARES Trust AI (2021. – 2026.). Available at: https://interparestrustai.org/trust.

Jurafsky, D., Martin, J. H. (2021). N-gram Language Models. Speech and Language Processing (3rd ed.). Available at: https://web.stanford.edu/~jurafsky/slp3/.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. arXiv. https://doi.org/10.48550/ARXIV.CS/9605103.

Lee, C. A. (2005). Defining Digital Preservation Work: A Case Study of the Development of the Reference Model for an Open Archival Information System. Thesis. The University of Michigan. . Available at: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/39372/dissertation_callee.pdf?sequence=2&isAllowed=y.

Lider Media (2024). Parlament donio revolucionarni propis o umjetnoj inteligenciji. Available on: https://lidermedia.hr/tehno/parlament-donio-revolucionarni-propis-o-umjetnoj-inteligenciji-156078?utm_source=chatgpt.com.

Lynch, C. A. (2001). Metadata harvesting and the Open Archives Initiative. A Bimonthly Report (217). Available at: https://www.cni.org/wp-content/uploads/2001/08/Metadata-Harvesting-and-the-Open-Archives-Initiative.pdf.

Michael I. Jordan (2014). Statistics and machine learning. Reddit.com. Available at: https://www.reddit.com/r/MachineLearning/comments/2fxi6v/comment/ckelmtt/?context=3.

Papalambros, P. Y., Wilde, D. J. (2017). Principles of Optimal Design: Modeling and Computation. Cambridge University Press.

PREMIS Data Dictionary (full document), Version 3.0. Availabel at: https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.

Russell, S. J., Norvig, P. (2021.). Artificial Intelligence: A Modern Approach. (4th ed.). Hoboken: Pearson.

Yoosefzadeh-Najafabadi, M., Earl, H. J., Tulpan, D., Sulik, J., & Eskandari, M. (2021). Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean. In Frontiers in Plant Science (Vol. 11). Frontiers Media SA. https://doi.org/10.3389/fpls.2020.624273.

Zhou, V. (2020). Machine Learning for Beginners: An Introduction to Neural Networks. Medium. Available at: https://towardsDatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9.

Žaja, L. (2020). Razvoj arhivistike prema suvremenoj znanosti - analiza časopisa Arhivski vjesnik od 1899. do 2017. godine. Doctoral Dissertation. Faculty of Humanities and Social Sciences. Zagreb. Croatia. Available at: https://repozitorij.ffzg.unizg.hr/islandora/object/ffzg:3319/Datastream/PDF/view.

Žaja, L. (2021). Tranzicija Digitalne znanosti u virtualno okruženje umjetne inteligencije. Split: Redak.

Žaja, L. (2023). Povijest umjetne inteligencije – informacijsko – komunikacijski aspekti u umjetnoj inteligenciji. Zagreb: Croatian State Archives.

POVZETEK

EVROPSKI AKT O UMETNI INTELIGENCI: REGULATIVNI OKVIR ZA UMETNO INTELIGENCO V DIGITALNIH ARHIVIH

Dr. Lana ŽAJA
Arhivistka, Hrvaški Državni arhiv, Zagreb, Hrvaška
lzaja@arhiv.hr

Dolgo pričakovani Akt o umetni inteligenci je bil sprejet v Evropskem parlamentu kot del zakonodajnega postopka EU. Zakon nalaga razvijalcem ter informacijsko-tehnološkim strokovnjakom s področja umetne inteligence nove zahteve, vključno z opredelitvijo tveganj sistemov, katerih delovanje temelji na umetni inteligenci.

Čeprav je besedilo akta že na voljo in vsebuje vse pravne formulacije razen končnih preverjanj, še vedno ni povsem jasno, kako bo zakon vplival na digitalno informacijsko, komunikacijsko in arhivsko okolje v EU ter zunaj nje. Besedilo akta je v številnih vidikih nejasno, pogosto pa so navedene visoke stopnje tveganja za informacijske sisteme, pri čemer so podrobnosti opisane v povezanih smernicah, standardih ter zakonih in politikah držav članic EU.

Avtorica opisuje naslednje korake v zakonodajnem postopku, povezane z implementacijo orodij umetne inteligence v digitalnih arhivih, ter potrebne smernice in standarde, ki omogočajo pravilno izvajanje evropskega akta o umetni inteligenci.

O avtorici:

Lana Žaja je doktorica znanosti s področja informacijskih in komunikacijskih znanosti ter ustanoviteljica in direktorica podjetja Infinity Alpha Lab d. o. o., specializiranega podjetja za visoko analitično svetovanje na področju arhitekture informacijskih sistemov, dokumentnega sistema (DMS), digitalnega poslovanja, izobraževanja in programiranja, ki pomaga lastnikom podjetij, menedžerjem in vodjem skupin pri razvoju obstoječih podatkovnih baz na podlagi programskih rešitev. Že več kot 20 let dela Lana Žaja v dinamičnih poslovnih okoljih, kjer jo odlikuje strast do čim bolj sistematičnega in analitičnega pristopa k vsem delom procesa upravljanja, od iskanja rešitev za kompleksne tehnologije in sisteme (IKT) prek organizacijske sheme poslovnih procesov do razvoja in vzdrževanja kompleksnih specializiranih podatkovnih zbirk za prenos, shranjevanje, ustvarjanje, deljenje ali izmenjavo informacij.

Doktorirala je iz znanstvene analitike podatkov na Univerzi v Zagrebu z odliko in je prva avtorica knjige, ki obravnava razvoj umetne inteligence (UI) v Republiki Hrvaški z vidika družboslovja od začetkov do danes. Je tudi avtorica več znanstvenih člankov, poglavij v učbenikih in knjige o analitični scientometriji. Poleg tega je priznana kot recenzentka več znanstvenih člankov z multidisciplinarnih področij in urednica znanstvenih knjig.

About the author:

Lana Žaja holds a Ph. D. in Information and Communication Sciences and is the Founder and CEO of Infinity Alpha Lab d.o.o., specialized company for highly analytical consulting in information systems architecture, document magament system (dms), digital business, education and programming that helps business owners, managers and team leaders develop existing databases based on software solutions. For over 20 years, she has worked in dynamic Business Environments where she has distinguished herself with her passion for maximizing a systematic and Analytical approach to all parts of the management process, from finding solutions for complex technologies and systems (ICT) through the organizational scheme of business processes to the development and maintenance of complex specialized databases to transmit, store, create, share or exchange information.

Her Ph. D. in scientific data analytics at the University of Zagreb was awarded Magna Cum Laude and she is the first author of a book dealing with the development of Artificial Intelligence (AI) in the Republic of Croatia, from the perspective of Social Sciences from its beginnings to the present day. She is also the author of several scientific papers, chapters in textbooks and a book dealing with analytical scientometrics. Additionally, she is recognized as a reviewer of several scientific papers in multidisciplinary fields and an editor of scientific books.