UDC 004.89

Cybersecurity of Intellectual Information Aggregation Processes into Digital Archives

Yuriy Tsyrulnev[1]

[1] *DIGITAL DOCS® LLC, 5-7 Motornyi Lane, Kyiv, 03083, Ukraine*

Abstract

The article addresses the problem of cybersecurity in intellectual information aggregation (IIA) processes within digital archives, which arise during the automated collection, structuring, semantic enrichment, and analysis of heterogeneous data using artificial intelligence (AI), machine learning (ML), and large language models (LLMs). The study focuses on identifying vulnerabilities of IIA processes and their mathematical formalization across stages such as digitization, image processing, optical character recognition (OCR), classification, indexing, and archival system creation. Particular attention is given to formalizing cyber threats, including unauthorized access, integrity violations, metadata forgery, adversarial attacks on AI/ML models, data manipulation, prompt injection, data exfiltration, and digital signature forgery. For each threat category, mathematically grounded countermeasures are proposed, including encryption, multi-factor authentication, monitoring, anomaly detection, access control, metadata protection, and adversarial training. The paper emphasizes the emergent properties of combined defenses, highlighting the resilience of digital archives against cyber threats that arise from the interaction of individual safeguards. The proposed models can be applied to the assessment and strengthening of information system security in the context of state and societal digital transformation. Practical aspects of implementing digital archive creation processes have been validated through patented solutions for converting large collections of paper documents into digital information resources [15]. To support the functioning of intellectual information aggregation processes, specialized software packages are employed, the modules of Digital Docs® Technology, registered as a copyrighted work [16]. Practical deployment of the proposed solutions is carried out within the activities of DIGITAL DOCS®, registered as a trademark [17].

Keywords: Digital Information Resources; Digital Archives; Intellectual Information Aggregation; Cybersecurity; Cyber Threats; Artificial Intelligence; Machine Learning; Large Language Models.

Introduction

From the authors' perspective, whose joint experience is particularly related to the digitization – creation of digital information resources (DIR) and creation of digital archives (DA) of documents for Ukrainian government institutions and the development of comprehensive cybersecurity systems, digital archives have, under the conditions of state and societal digital transformation, become one of the key instruments for storing, processing, and analyzing large volumes of information.

They ensure accuracy, impartiality, completeness, and timeliness of analytics, thereby supporting evidence-based decision-making. Intellectual information aggregation enables the automation of processes of collection, structuring, analysis, and synthesis of information while creating digital archives. However, alongside its advantages, the use of intellectual information aggregation also introduces new cybersecurity challenges.

The primary aim of this paper is to analyze threats associated with the use of artificial intelligence (AI), machine learning (ML), and large language models (LLM) in the digitization and creation of digital archives of documents, and to emphasize the importance

of developing methods to counteract these threats. To achieve this aim, the study defines, systematizes, describes, and mathematically formalizes:

- The processes of intellectual information aggregation in digital archives.
- Cyber threats to these processes are associated with the application of AI, ML, and LLMs, including adversarial attacks, model bias, data manipulation, prompt injection, data exfiltration, disinformation, and misuse of results.
- Countermeasures to such threats in IIA processes.

A formalized set of countermeasures is proposed to mitigate these threats. Despite the relevance of this research, the authors have not identified scientific studies by Ukrainian scholars specifically devoted to the cybersecurity aspects of intellectual information aggregation processes in digital archives.

The scientific novelty of the study lies in the formulation and mathematical modeling of the relationships between intellectual information aggregation processes in digital archives, the associated cyber threats, and the countermeasures against these threats, as well as in identifying the necessity and directions for developing specialized cybersecurity measures for such processes. The results of the article emphasize the importance of a comprehensive approach to ensuring the security of digitization and digital archive creation processes under the conditions of Ukraine's digital transformation.

1. Intellectual Information Aggregation for Digital Archives

Within this study, the term Intellectual Information Aggregation (IIA) is introduced to denote a set of processes for automated collection, structuring, analysis, and semantic enrichment of data from heterogeneous sources using AI, ML, and LLM technologies. Unlike the classical technical concept of "aggregation," which in information systems typically refers to the mechanical merging or grouping of data (see ISO/IEC 2382:2015), intellectual aggregation encompasses the following characteristics:

- Semantic normalization of unstructured or partially structured information.
- Contextual classification and indexing with regard to the meaning and role of data.
- Identification of logical–semantic relations between fragments of information.
- Adaptive real-time processing with the involvement of heuristic or learning models.

The term "intellectual" in this context does not define the content of the information but rather indicates the method of its processing—through cognitive mechanisms. The added qualifier "intellectual" specifies the manner of aggregation, namely by means of AI/ML/LLM. The term is analytical rather than descriptive.

Thus, the notion of intellectual information aggregation reflects a new technological quality in the transformation of archival practice into a digital infrastructure with elements of automated semantic processing. Its introduction is supported by contemporary scientific literature describing similar approaches [11], [12]. The authors regard the use of this term as methodologically justified in the study of digital archive creation processes, taking into account the intellectualization of their digital population.

In this article, IIA is considered a modern approach to data management based on innovative technologies that enable efficient use of information resources, improve the accuracy of analytics, and ensure a high level of cybersecurity—making it indispensable in the current conditions of digital transformation. For example, study [1] explores AI in archival practice, particularly data processing automation, archive organization, and new forms of digital archives. It also emphasizes the necessity of integrating AI into archival systems while respecting archival accounting principles and applying a critical approach to the use of technology. Study [2] investigates the problem of accessibility of archives containing digitally born data. The authors consider the potential of AI and ML to facilitate access to digital archives, including automation of tasks such as privacy checking, while highlighting ethical principles such as transparency, fairness, and accountability in AI implementation.

Intellectual information aggregation in digital archives represents a set of processes of automated collection, structuring, analysis, and

synthesis of large data volumes from diverse sources, carried out using AI, ML, and LLM. These processes are aimed at transforming unstructured or partially structured data into formats suitable for effective search, classification, forecasting, and decision-making. Archival information aggregation is, in particular, examined in study [3], which presents results on applying AI, especially LLM and knowledge graphs, for organizing and analyzing oral history archival resources. Study [4] investigates the use of LLM to optimize archival work, particularly the processing and analysis of textual materials. The authors propose a novel model — Archival Generative Pre-trained Transformer (ArcGPT) — and evaluate its effectiveness in performing archival tasks, as well as developing a methodology for assessing user experience in working with LLM.

The main concepts associated with IIA in digital archives include AI [5], ML, LLM [6], and cybersecurity:

- AI enables automation of complex tasks such as speech recognition, natural language processing (NLP), and image analysis (IA). AI allows the system to autonomously make decisions related to IIA, for example, determining the importance of documents or detecting duplicates. Study [7] explores the use of AI to automate classification, appraisal, and disposal processes in digital archiving. Based on four case studies from Australian archival and governmental institutions, the authors analyze achievements, challenges, and prospects for AI use in records management.
- ML is employed for automatic pattern discovery in data, classification of documents, segmentation of information, and trend prediction. ML algorithms allow archives to adapt to new data types and improve processing efficiency based on accumulated experience. Monograph [5] is devoted to concept extraction, classification, and semantic networks using LLM to process large volumes of textual data. Study [6] investigates the interrelation between archives, data access, and AI, focusing on the use of digital and digitized archival collections.
- LLM are used for understanding, interpreting, and generating textual information. Architectures such as GPT, BERT, and others make it possible to analyze large volumes of text, extract key ideas, generate summaries, and respond to user queries. These models can also be integrated into digital archives to improve information retrieval and enable more intuitive interaction with data.

Data integration refers to the combination of data from various sources (texts, images, audio, video) into a unified system, enabling the acquisition of comprehensive information. The use of AI and ML allows for automatic identification of relationships among diverse data types, thereby enhancing their analytical value.

A critical aspect of information aggregation is data security and cybersecurity. AI can be used to detect anomalies, identify potential threats, and prevent unauthorized access to archives. LLMs, in particular, can be applied to log analysis and the detection of suspicious activities within the system.

Finally, the adaptability and scalability of IIA enable digital archives to operate effectively with growing volumes of data while adapting to new requirements and technologies.

2. IIA Functions for DA

The IIA system for DA is a software–hardware complex that provides automated collection, processing, structuring, analysis, and storage of large volumes of data from various sources using advanced technologies such as AI, ML, and LLM. This system is designed to transform unstructured or partially structured data into a user-friendly digital format, enabling effective solutions to tasks of search, classification, forecasting, analytics, and decision-making.

The functions of the IIA system in digital archives cover a wide range of tasks aimed at efficient data collection, management, processing, analysis, and protection. The IIA system implements its functions through a set of interrelated processes that ensure automation, accuracy, security, and reliability. It allows efficient management of large-scale data by transforming them into useful information for analytics, forecasting, and decision support. The main functions of the IIA system for digital archives include:

- Creation of digital information resources (DIR): using automatic and semi-automatic systems based on AI, ML, LLM,

image processing (IP), OCR, and neural networks (NN).

- Creation of DA: utilizing ECM/CSP.
- Population of digital archives with digital information resources: ensuring systematic accumulation, organization, and preservation of data for long-term use.

3. Notation and Symbols

In this study, the following notations are used to formally describe processes of intellectual information aggregation (IIA), associated cyber threats, and countermeasures:

D (*Documents/Data*) — set of documents (data) subject to digitization, processing, or storage.

M (Metadata/Models) — set of metadata associated with documents, or, where applicable, models (AI/ML) that operate on archival data.

S (Signatures) — set of digital signatures used for authentication and verification.

Q (Queries/Prompts) — set of queries or prompts submitted to AI/ML/LLM components.

A (Actions/Activities) — set of actions or activities (user or system) subject to monitoring and analysis.

U (Users) — set of users (subjects of access) interacting with the system.

X (Events/Anomalies) — set of events in which anomalies may be detected.

Each process P_i is defined as a transformation of an input set into an output set. Cyber threats PT_j are modeled as unwanted or adversarial transformations applied to these sets, while countermeasures TM_k are protective transformations aimed at neutralizing or mitigating threats.

4. IIA Processes (P)

4.1. Documents Scientific and Technical Arrangement (P10)

The scientific and technical arrangement of documents is a crucial process for ensuring effective document management, long-term preservation, and subsequent use. Its primary purpose is to establish an organized archival collection, prepare documents for further utilization, optimize documentation management, and guarantee the preservation of essential materials.

The results of this process include:

- A systematized archival collection;
- Reduction in the volume of documents subject to subsequent digitization;
- Improved accessibility of information;
- Preparation of documents for archival storage;
- Compliance with legal and regulatory requirements.

Formally, the arrangement process can be represented as the partitioning of the input document set D into subsets D_k corresponding to defined categories:

$$D = \bigcup_{k=1}^{\kappa} D_k,\ D_i \bigcap D_j = 0$$

for $i \neq \mathrm{j}$, where K - is the number of categories. (A detailed formalization of the ordering relation may be the subject of a separate study).

$$P_{10}(D) = \{arrange(d) \mid d \in D\},$$

where $arrange(d)$— function of scientific and technical arrangement: assigns a document d to a defined archival category according to classification rules.

4.2. Document Scanning Preparation (P20)

Document scanning preparation is an essential process to ensure high-quality digitization. The aim of this process is to bring documents subject to digitization into a condition that allows proper scanning. It may include unbinding archival files and documents into individual sheets, partial restoration of damaged pages, and sorting of documents by binding status, format, physical condition, paper density, and other defined characteristics. This ensures their suitability for scanning using different types of scanners, technologies, and methods.

Bound volumes are disassembled into individual documents, and documents into separate sheets. Artifacts are removed; sheets bound together are separated by carefully removing threads, staples, or other fasteners—except in cases where unbinding would compromise the legal status of the document.

Sheets are cleaned of dust and contaminants, and creases or folds that may obscure information or cause sheets to adhere to each other are flattened. Damaged originals are carefully reinforced, while residual adhesive materials at folds or perforations are removed.

This process ensures that sheets are not damaged during scanning and that complete and accurate images are obtained without information loss. The outcome of scanning preparation is a collection of original documents ready for digitization.

Formally, the set of prepared documents can be denoted as:

$$P_{20}(D) = \{prep(d) \mid d \in D\},$$

where $prep(d)$ — function of document preparation: transforms a document d into a state suitable for scanning (unbinding, cleaning, restoration, etc.).

4.3. Document Scanning (P30)

Document scanning consists in creating electronic copies of documents in the form of raster images using scanners and specialized software that perform the functions of analog-to-digital converters. The purpose of this process is to obtain high-quality electronic graphical copies of pages in raster format, suitable for subsequent image processing.

The process may include scanner calibration and software configuration, selection of optimal scanning modes, installation and adjustment of drivers to ensure high-quality digitization of documents of various types, as well as parallel real-time monitoring and correction of scanning results. The scanning of sheets prepared for digitization is carried out directly.

The result of the process is a dataset of electronic graphical copies in raster format of all document pages, which can be used for further image processing: quality enhancement, OCR, classification, indexing, including with the AI tools.

$$P_{30}(D^*) = \{scan(d) \mid d \in D^*\},$$

where $scan(d)$ — scanning function: converts a prepared document d into a raster image.

4.4. Digital Image Processing (P40)

Digital image processing refers to the transformation of electronic graphical copies of document pages into a state necessary and sufficient for readability, printing, and subsequent processing using OCR, classification, and indexing tools supported by AI, ML, LLM, and neural networks (NN).

This process may include: cropping of page images along the edges; background padding of page images; normalization of page images to standard formats; adjustment of contrast, color, and brightness; correction of faded text; alignment of page images horizontally, vertically, or relative to lines and text; and removal of undesirable background from page images (including images of drawings on tracing paper or blueprint photocopies), while preserving informative elements such as seals, stamps, signatures, and annotations.

Digital image processing can be performed in automatic, semi-automatic, or manual modes using scanner drivers and/or specialized software.

As a result of digital image processing, a dataset of electronic graphical copies of all document pages is created, prepared for further processing with AI, ML, LLM, NN, and OCR tools.

The processed electronic copies are obtained by means of the processing function:

$$P_{40}(I) = \{process(i) \mid i \in I\},$$

where $process(i)$ — image processing function: transforms a raster image i by normalization, enhancement, and artifact removal.

4.5. Optical Character Recognition, OCR (P50)

The recognition of symbols (letters, digits, and characters) contained in digital graphical copies of document pages is performed through OCR and/or intelligent recognition methods using AI, ML, LLM, and NN. Its objective is the transformation of handwritten, typewritten, or printed text images into sequences of codes represented in ANSI/ASCII symbols, which are interpretable by electronic processing systems.

The conversion of symbols into ANSI/ASCII characters enables word and number search within the text, copying of text into word processors, text editing, display or printing without loss of quality, information analysis, as well as the application of electronic translation, formatting, or text-to-speech transformation.

As a result of character recognition from digital raster copies of documents, a dataset of electronic document page copies is produced, each containing two layers: the digital raster graphical copy of the document on one layer and the recognized text in ANSI/ASCII symbols on the other.

The process of text recognition from processed images can be formalized as:

$$P_{50}(I) = \{OCR(i) \mid i \in I\},$$

where $OCR(i)$ — optical character recognition function: converts a raster image i into a machine-readable text representation.

4.6. Document Classification (P60)

The classification of electronic document copies consists in assigning scanned and processed copies to categories based on their content, format, and metadata. This process employs AI, ML, LLM, NN, and OCR technologies.

The primary objective of classification is to ensure the systematic organization of electronic document copies, thereby simplifying subsequent search and use of documents within the DA.

At the classification stage, each document is analyzed, and on the basis of defined characteristics it is assigned to a particular category.

Metadata describing the class affiliation of the document are also added to specialized registers for further processing.

Formally, documents are classified into categories using AI, ML, LLM, and OCR as follows:

$$P_{60}(T) = \{classify(t) \mid t \in T\},$$

where $classify(t)$ — classification function: assigns a text t to a predefined class based on content and metadata.

4.7. Indexing (P70)

The indexing process involves the creation of unique search keys (tags, metadata, etc.) for each document.

This significantly reduces the time required to retrieve information in DA. Indexing is performed through analysis of document content, keywords, and metadata.

The results of indexing are stored as structured data that are associated with the corresponding documents.

This ensures a high level of automation during information retrieval in the archive and provides technical simplification of search by creating a structured dataset in addition to the unstructured mass of information.

Formally, document indices are generated as a set of keywords (tags/metadata):

$$P_{70}(T) = \{index(t) \mid t \in T\},$$

where $index(t)$ — indexing function: generates metadata and search keys from a text t.

4.8. Merging of Pages (P80)

The merging of pages into documents is a necessary stage for restoring the integrity of multi-page documents.

During this process, individual pages belonging to the same document are combined into a single electronic structure. Merging is performed on the basis of unique identifiers or metadata that specify the correct sequence of pages.

This stage is critically important for documents with a strictly defined page order.

The merging of pages into a single document can be represented as:

$$P_{80}(P) = \{merge(p) \mid p \in P\},$$

where $merge(p)$ — merging function: restores the sequence of pages p into a complete document.

In practical implementation, this process employs document binding devices patented in [18].

4.9. Binding of Document Files (P85)

The binding of documents into folders is performed to group multiple documents into a single logical unit, such as an archival volume or dossier.

This process is applied to documents that share a thematic or functional relationship.

During binding, the common characteristics of the documents are identified, after which they are combined into a folder with the creation of a unified index and metadata.

The aggregation of documents into folders can be formalized as:

$$P_{85}(F) = \{bind(f) \mid f \in F\},$$

where $bind(f)$— binding function: groups related files f into a dossier or volume.

4.10. Digital Archive System Creation (P90)

The creation of a digital archive system is a comprehensive process that includes the configuration of ECM/CSP systems and the integration of all archive components.

The system incorporates modules for managing storage, search, protection, and document updating.

At this stage, the primary objective is to ensure high data availability and security, as well as to configure automation mechanisms for processing.

The result is a DA that provides reliable access to information while taking user rights into account.

The digital archive system consists of a database that includes:

$$P_{90}(S) = \{build(s) \mid s \in S\},$$

where $build(s)$ — archive creation function: integrates subsystem s into a digital archive (ECM/CSP platform).

5. Cyber Threats to the processes of intellectual information aggregation in digital archives (PT)

As shown in Table 1 and Figure 1, each of the described above IIA Processes P_i has specific threats $PT = \{PT_1, PT_2, \dots PT_n\}$ that require detailed analysis and protective measures. For example, in the case of data collection, ensuring the security of transmission is of critical importance, whereas for the process of analysis, protection against model bias is key. In [8], for instance, the authors define security through the lens of integrity, confidentiality, availability, intruder detection, and protection against physical damage and viruses.

Below are examples of possible threats and vulnerabilities to IIA processes (it should be noted that these examples are valid but not exhaustive—the list may be extended).

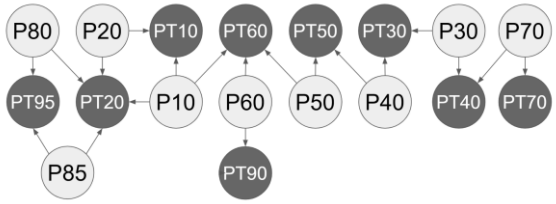
Table 1

Matrix of IIA Processes Threats (P–PT)

	Process \ Threat	PT10 Unauthorized Access	PT20 Data Interception in Transit	PT30 Data Integrity Violation	PT40 Metadata Forgery	PT50 Attacks on AI/ML Models	PT60 Data Manipulation	PT70 Prompt Injection	PT80 Data Exfiltration	PT90 Disinformation	PT95 Digital Signature Forgery
P10	Documents Scientific and Technical Arrangement	1	1	0	0	0	1	0	0	0	0
P20	Document Scanning Preparation	1	1	0	0	0	0	0	0	0	0
P30	Document Scanning	0	0	1	1	0	0	0	0	0	0
P40	Digital Image Processing	0	0	1	0	1	0	0	0	0	0
P50	Optical Character Recognition	0	0	0	0	1	1	0	0	0	0
P60	Document Classification	0	0	0	0	0	1	0	0	1	0
P70	Indexing	0	0	0	1	0	0	1	0	0	0
P80	Merging of Pages	0	1	0	0	0	0	0	0	0	1
P85	Binding of Document Files	0	1	0	0	0	0	0	0	0	1

Figure 1

Threats to IIA Processes (P-PT)



5.1. Unauthorized Access (PT10)

This threat consists in the unauthorized acquisition of access to digital archive data by adversaries or improperly authorized users. It arises from vulnerabilities in authentication mechanisms, insufficient access control, or credential leakage. Potential sources include hacker attacks, internal insiders, or misconfigurations of security systems. Unauthorized access to data implies obtaining archival information without the appropriate permission, which may lead to a compromise of confidentiality. This threat can be mathematically formalized as follows:

$$PT_{10}(D) = \{breach(d) \mid d \in D\},$$

where $breach(d)$ — unauthorized access to document d.

5.2. Data Interception in Transit (PT20)

The threat of data interception arises during transmission between a user and a server or between servers of the digital archive. It occurs due to the use of unsecured communication channels (e.g., HTTP instead of HTTPS), the absence of encryption, or man-in-the-middle (MITM) attacks. The sources of such threats include adversaries with access to network infrastructure or those employing specialized spyware.

Data interception may occur when information is transmitted between systems without proper encryption or protective mechanisms.

This threat can be mathematically formalized as follows:

$$PT_{20}(D) = \{intercept(d) \mid d \in D\},$$

where $intercept(d)$ — interception of data d during transmission.

5.3. Data Integrity Violation (PT30)

This threat involves intentional or accidental modifications of archival data, resulting in the loss of their authenticity.

It arises from attacks on the file system, errors during data processing, or malicious software.

The sources may include internal users with editing rights as well as external adversaries who have gained access to the archive.

The threat encompasses deliberate or accidental alteration of archival data and can be mathematically formalized as follows:

$$PT_{30}(D) = \{corrupt(d) \mid d \in D\},$$

where $corrupt(d)$ — violation of integrity, deliberate or accidental modification of document d.

5.4. Metadata Forgery (PT40)

This threat consists in the deliberate modification of document metadata, which may result in the loss of correct information about provenance, authorship, or creation date. It arises from insufficient control of metadata authenticity, the possibility of editing without verification, or attacks on the database. The sources include internal employees manipulating metadata or adversaries altering it to bypass control systems.

The modification of document metadata may lead to incorrect identification or loss of authenticity. Such actions can be mathematically formalized as follows:

$$PT_{40}(M) = \{forge(m) \mid m \in M\},$$

where $forge(m)$ — forgery or alteration of metadata item m.

5.5. Attacks on AI/ML Models (PT50)

This threat is associated with the deliberate manipulation of input data in order to deceive artificial intelligence and machine learning algorithms used in digital archives. It arises from the insufficient robustness of models to malicious modifications in input data (e.g., poisoning of the training dataset). The sources may include adversaries injecting manipulated data or creating specially crafted adversarial examples.

Attacks aimed at altering the system's AI-driven outputs through manipulations of input data can be mathematically formalized as follows:

$$PT_{50}(M) = \{attack(model) \mid model \in M\},$$

where $attack(model)$ — adversarial attack on AI/ML model (e.g., poisoning, adversarial input).

5.6. Data Manipulation (PT60)

Data manipulation involves the deliberate alteration of information with the purpose of disinformation, misleading users, or preparing fraudulent activities. It arises from vulnerabilities in document editing systems or insufficient control of modifications. The sources include internal adversaries or hackers who alter data to achieve their objectives.

Unauthorized modification of information stored in the archive can be mathematically formalized as follows:

$$PT_{60}(D) = \{manipulate(d) \mid d \in D\},$$

where $manipulate(d)$ — manipulation of data d for misleading or fraudulent purposes.

5.7. Prompt Injection (PT70)

This threat consists in the injection of malicious commands or prompts into automated systems that employ AI/ML, with the aim of obtaining unauthorized results or altering model behavior.

It arises from the absence of input filtering and insufficient protection of response generation mechanisms.

The sources include adversaries manipulating textual queries to gain access to confidential information or to bypass imposed restrictions.

The injection of malicious commands into the system to alter its behavior can be mathematically formalized as follows:

$$PT_{70}(Q) = \{ inject(q) \mid q \in Q \},$$

where $inject(q)$ — prompt injection, maliciously altering the behavior of query q.

5.8. Data Exfiltration (PT80)

Data exfiltration refers to the unauthorized copying and transfer of information to third parties.

It arises from insufficient access control, the absence of user anomaly monitoring, or the use of malicious software.

The sources include both internal insiders and external adversaries exploiting malicious scripts or attacks on archive servers.

Unauthorized copying or transmission of data can be mathematically formalized as follows:

$$PT_{80}(D) = \{exfiltrate(d) \mid d \in D\},$$

where $exfiltrate(d)$ — unauthorized copying or transfer of data d.

5.9. Disinformation (PT90)

Disinformation refers to the deliberate introduction of false data into the digital archive with the purpose of distorting genuine information.

It arises from insufficient control over data authenticity, the absence of verification mechanisms, or malicious manipulations.

The sources may include hackers, state-sponsored actors, or internal employees disseminating falsified information.

The insertion of false or manipulative information into the system can be mathematically formalized as follows:

$$PT_{90}(D) = \{disinform(d) \mid d \in D\},$$

where $disinform(d)$ — injection of false or misleading content into data d.

5.10. Digital Signature Forgery (PT95)

This threat consists in the creation or forgery of digital signatures for documents, intended to mislead users about the authenticity of files.

It arises from vulnerabilities in cryptographic algorithms, leakage of private keys, or insufficient protection of signing mechanisms.

Sources of such threats include adversaries who have obtained signing keys or who mount attacks against certification authorities.

The falsification of a document's digital signature can be formally expressed as follows:

$$PT_{95}(S) = \{forge(s) \mid s \in S\},$$

where $forge(s)$ — digital signature forgery for signature s.

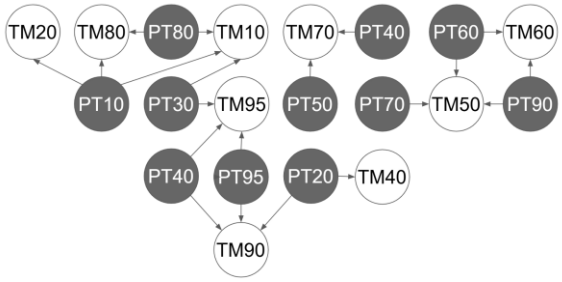
6. Threat Mitigation (TM)

As shown in Table 2 and Figure 2, each of the described above Threats PT_i must be applied by Mitigation TM_j.

Table 2
Matrix of Threat's Mitigations (PT–TM)

Threat \ Mitigation	TM10 Data Encryption	TM20 Multi-Factor Authentication	TM30 Activity Monitoring	TM40 Data Updating and Backup	TM50 Request Verification and Filtering	TM60 Anomaly Detection	TM70 Protection Against AI/ML Attacks	TM80 Access Control and Logging	TM90 Digital Signature Verification	TM95 Metadata Protection
PT10 Unauthorized Access	1	1	0	0	0	0	0	1	0	0
PT20 Data Interception in Transit	0	0	0	1	0	0	0	0	1	0
PT30 Data Integrity Violation	1	0	0	0	0	0	0	0	0	1
PT40 Metadata Forgery	0	0	0	0	0	0	0	0	1	1
PT50 Attacks on AI/ML Models	0	0	0	0	0	0	1	0	0	0
PT60 Data Manipulation	0	0	0	0	1	1	0	0	0	0
PT70 Prompt Injection	0	0	0	0	1	0	0	0	0	0
PT80 Data Exfiltration	1	0	0	0	0	0	0	1	0	0
PT90 Disinformation	0	0	0	0	1	1	0	0	0	0

Figure 2
Mitigations to Threats (TM-PT)



6.1. Data Encryption (TM10)

Data encryption is a fundamental method for ensuring the confidentiality and integrity of information.

It prevents unauthorized access by transforming data into a format unreadable without the corresponding decryption key.

Two primary approaches are employed: symmetric encryption (e.g., AES, DES), where the same key is used for both encryption and decryption, and asymmetric encryption (e.g., RSA, ECC), which relies on a pair of keys — a public key for encryption and a private key for decryption.

Within the cybersecurity of digital archives, encryption is applied both to data at rest (e.g., encrypted file systems, databases) and to data in transit (e.g., TLS, SSL, VPN).

Properly implemented encryption safeguards archival information against interception, unauthorized access, and data leakage during storage, processing, and transmission. Formally:

$$TM_{10}(D) = \{encrypt(d) \mid d \in D\},$$

where $encrypt(d)$ — encryption of data d (at rest or in transit).

6.2. Multi-Factor Authentication (TM20)

Multi-factor authentication (MFA) provides an enhanced level of access security for digital archives by requiring users to verify their identity through multiple independent factors. The primary categories of authentication factors include:

Personal Identification Numbers (PINs) or answers to security questions;

One-Time Passwords (OTPs) delivered via Short Message Service (SMS), dedicated applications, hardware tokens, or smart cards;

Biometric data, such as fingerprints, facial recognition, or iris scans.

The combination of at least two of these factors substantially complicates account compromise. Formally:

$$TM_{20}(U) = \{authenticate(u) \mid u \in U\},$$

where $authenticate(u)$ — multi-factor authentication of user u.
or:

$$TM_{20}(u) = \begin{cases} 1, if\ \exists\ f_i, f_j \in F, i \neq j: verify(u, f_i) \wedge verify(u, f_j) \\ 0, otherwise \end{cases}$$

where $verify(u, f)$ denotes the successful verification of user $u \in U$ by factor $f \in F$.

6.3. Activity Monitoring (TM30)

Activity Monitoring is the process of collecting, analyzing, and correlating events associated with the use of a digital archive. It enables the detection of anomalous activities, including attempts of unauthorized access, suspicious modifications of document contents, or large-scale file exfiltration. The main methods of monitoring include:

- Logging — recording system events, including user activities, into audit logs;
- Behavioral Analysis — defining baseline user behavior and detecting deviations from the norm;
- Anomaly Detection Systems — machine learning algorithms that analyze traffic patterns and behavioral profiles to identify potential threats.

$$TM_{30}(A) = \{monitor(a) \mid a \in A\},$$

where $monitor(a)$ — monitoring and logging of action a.

Let $A(t)$ — denote the user's activity at time t, while μ and δ represent the mean value and the standard deviation of activity under the assumption of a normal distribution. The deviation is quantified using the *Z*-score:

$$Z = \frac{A(t)-\mu}{\delta}, \text{ if } |Z| > \tau,$$

where τ — is the anomaly threshold, the system flags the activity as suspicious.

6.4. Data Updating and Backup (TM40)

Backup is a fundamental mechanism for ensuring the continuity of operations of a digital archive. Several backup strategies exist:
- Full Backup — creation of a copy of all data.
- Differential Backup — saving only the data changed since the last full backup.
- Incremental Backup — saving the data changed since the last backup of any type.

The system must periodically update backup copies and verify their correctness. Let $D(t)$ denote the state of data at time ttt, and $B(t)$ the backup copy. The creation of a backup is described by the equation: $B(t) = D(t)$. Data recovery is performed in the event of failure or loss according to: $D(t') = B(t'), t' > t$.

$$TM_{40}(D) = \{backup(d) \mid d \in D\},$$

where $backup(d)$ — creation of backup copy of document d.

6.5. Request Verification and Filtering (TM50)

The process of request verification and filtering is aimed at detecting and blocking malicious requests to the digital archive system. The primary threats include SQL injections, XSS attacks, and manipulations with API requests. The filtering mechanism includes:
- Syntactic analysis of the request — checking the correctness of input;
- Blacklists and whitelists of requests — blocking dangerous commands;
- Input data validation — verifying that the request conforms to the expected format.

Let Q denote the input request, and $verify(Q)$ the verification function. Then:

$$verify(Q) = \begin{cases} 1, if\ the\ request\ is\ valid\ and\ safe; \\ 0, if\ the\ request\ is\ malicious\ or\ invalid. \end{cases}$$

$$TM_{50}(Q) = \{verify(q) \mid q \in Q\},$$

where $verify(q)$ — verification and filtering of query q.

6.6. Anomaly Detection (TM60)

Anomaly detection mechanisms are employed to automatically identify suspicious or malicious activities in the digital archive system.

They are based on comparing current events with predefined behavioral norms. The main approaches to anomaly detection include:
- Statistical methods — identifying deviations from mean values and distributions;
- Machine learning methods — analyzing historical data to detect anomalous behavior patterns;
- Signature-based detection methods — verifying activity against known attack types.

These approaches allow for the detection of both known threats (e.g., repeated failed login attempts) and novel attacks that have not previously been observed in the system.

Let X be the set of user activity features, and $f(X)$ the model of predicted normal behavior. An anomaly is detected under the condition:

$$anomaly(X) = \begin{cases} 1, if\ A(X) \neq f(X); \\ 0, otherwise \end{cases}$$

or, using a statistical approach:

$$anomaly(X) = \begin{cases} 1, if\ \dfrac{A(X)-\mu}{\sigma} > k; \\ 0, otherwise \end{cases}$$

where $\mu$is the mean activity parameter, $\sigma$is the standard deviation, and k is the threshold value.

$$TM_{60}(X) = \{detect(x) \mid x \in X\},$$

where $detect(x)$ — detection of anomaly in event x.

6.7. Protection Against AI/ML Attacks (TM70)

Attacks on AI/ML models may include training data poisoning, manipulations of input queries, or the creation of adversarial examples specifically designed to alter model behavior. To mitigate these threats, the following protective measures are commonly employed:

- Filtering and validation of training data — preventing data poisoning by detecting and eliminating corrupted samples;
- Defense against input manipulation — developing robust models that integrate techniques resistant to adversarial modifications;
- Adversarial training — training models to enhance their resilience against attacks through exposure to adversarially crafted examples.

Let $f_{\boldsymbol{\theta}}(X)$ — prediction function of the model with parameters θ . A model's vulnerability implies that there exists a minor perturbation of the input data δ that:

$f_\theta(X) \neq f_\theta(X+\delta)$ where $||\delta||$ — a minor perturbation causing a major shift in model predictions.

$TM_{70}(M) = \{defend(model) \mid model \in M\}$, where $defend(model)$ — protection of AI/ML model through adversarial training or data sanitization.

6.8. Access Control and Logging (TM80)

Access control enables the restriction of user rights within the system according to their privileges.

This reduces the risks of data compromise resulting from internal or external threats.

The primary access control mechanisms include:

- Role-Based Access Control (RBAC) — access is determined based on user roles;
- Attribute-Based Access Control (ABAC) — restrictions are enforced according to user attributes and the context of the request;
- Two-step verification — combines authorization with additional identification methods.

Simultaneously with access control, all user actions are logged for subsequent auditing and analysis. Let $P(u,r)$ denote the access rights of user u to resource r:

$$P(u,r) = \begin{cases} 1, if\ user\ u\ is\ autorizes\ to\ access\ resource\ r; \\ 0, otherwise \end{cases}$$

where t — the time point of operation execution.

$$TM_{80}(U) = \{authorize(u) \mid u \in U\},$$

where $authorize(u)$ — enforcement of access control rules for user u.

6.9. Digital Signature Verification (TM90)

Digital Signature ensures both the authenticity and the integrity of a document. It is based on public-key cryptographic algorithms, in particular the Rivest–Shamir–Adleman (RSA) algorithm and the Elliptic Curve Digital Signature Algorithm (ECDSA), and allows verification of whether a document has been altered after being signed. The main stages of digital signature verification are as follows:

- Hash function computation — a mathematical algorithm that transforms arbitrary input data (text, file, message, etc.) into a fixed-size hash string/hash value (digest) using a cryptographic algorithm (e.g., Secure Hash Algorithm 256, SHA-256), which generates a fixed 256-bit (32-byte) number from input data of any size;
- Signature decryption with the public key;
- Comparison of the computed hash value with the hash of the document.

Let $S(D)$ denote the digital signature of a document d , and $H(d)$ — its hash. The signature is considered valid if:

$$verify(S,d) = \begin{cases} 1, if\ H(d) = D_{K_{pub}}(S(d)); \\ 0, otherwise. \end{cases}$$

where $D_{K_{pub}}$ — the function of signature decryption with a key.

$$TM_{90}(S) = \{verify(s) \mid s \in S\},$$

where $verify(s)$ — verification of digital signature s.

6.10. Metadata Protection (TM95)

Metadata contain supplementary information about documents, such as authorship, creation date, format, and location.

Adversaries may tamper with metadata to conceal traces of their activity, alter access rights, or manipulate the document's provenance. Metadata protection involves the following measures:

- Encryption and hashing of metadata — to prevent unauthorized modifications;
- Access control for metadata — allowing only authorized users to view or modify metadata;
- Change logging — recording all operations with metadata for subsequent audit.

Let $M(D)$ denote the set of metadata of document d. Metadata protection requires that for all d' potentially subject to alteration:

$$H(MD(d)) = H(MD(d'))$$

where H is a hash function ensuring integrity verification. If $H(MD(d)) \neq H(MD(d'))$, this indicates an attempted metadata forgery.

$$TM_{95}(M) = \{protect(m) \mid m \in M\},$$

where $protect(m)$ — integrity protection of metadata m.

6.11. Use Mitigations in IIA Processes

As shown in Table 3 and Graph 3, Countermeasures must be applied within the Processes.

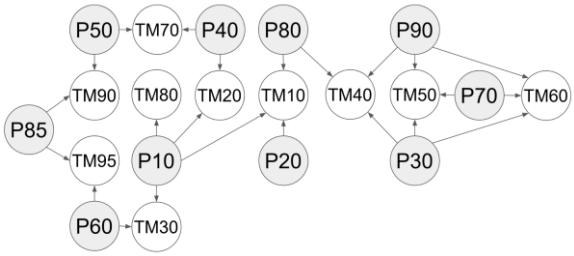
Table 3

Matrix of Processes and Threat Mitigations (P–TM)

Process \ Mitigation		TM10 Data Encryption	TM20 Multi-Factor Authentication	TM30 Activity Monitoring	TM40 Data Updating and Backup	TM50 Request Verification and Filtering	TM60 Anomaly Detection	TM70 Protection Against AI/ML Attacks	TM80 Access Control and Logging	TM90 Digital Signature Verification	TM95 Metadata Protection
P10	Documents Scientific and Technical Arrangement	1	1	1	0	0	0	0	1	0	0
P20	Document Scanning Preparation	1	0	0	0	0	1	0	0	0	0
P30	Document Scanning	0	0	0	1	1	1	0	0	0	0
P40	Digital Image Processing	0	1	0	0	0	0	1	0	0	0
P50	Optical Character Recognition	0	0	0	0	0	0	1	0	1	0
P60	Document Classification	0	0	1	0	0	0	0	0	0	1
P70	Indexing	0	0	0	0	1	1	0	0	0	0
P80	Merging of Pages	1	0	0	1	0	0	0	0	0	0
P85	Binding of Document Files	0	0	0	0	0	0	0	0	1	1
P90	Digital Archive System Creation	0	0	0	1	1	1	0	0	0	0

Figure 3

Threat Mitigations to IIA Processes (P–TM)



7. Emergence of Countermeasure Systems Against Cyber Threats in IIA Processes into DA

Emergence refers to the effect whereby a system acquires properties or behaviors that cannot be predicted solely from the properties of its individual components.

In other words, it is the situation in which the whole becomes greater than the sum of its parts. The main characteristics of emergence include:

- Novel quality — properties of the integrated system that are absent in its individual components;
- Interaction of components — emergent properties arise through interactions among the system's elements;
- Unpredictability — system-level properties that cannot be foreseen on the basis of analyzing separate components alone.

The mathematical formalization of emergence presupposes the description of how new properties or behaviors of the system arise as a result of its structure and the interactions of its components.

General approaches to such formalization can be expressed as follows:

$$E(DA) = f(\{C_1\}, I(C_i, C_j)),$$

where $E(S)$ demotes the emergent properties of the DA, $\{C_i\}$ is the set of its components, and $I(C_i, C_j)$ represents the interactions among components.

7.1. Basic Components of the Countermeasure Systems Against Cyber Threats in IIA Processes into DA System

Let S be a system consisting of nnn components:

$$S = \{C_1, C_2, \ldots, C_n\}.$$

where C_i is an individual component of the system.

7.2. Properties of Components

Each component C_i has a set of properties or states $P(C_i)$:

$$P(C_i) = \{p_1(C_i), p_2(C_i), \ldots, p_k(C_i)\}.$$

7.3. Interaction of Components

The interaction between components is defined by a function I, which describes the relationship between C_i and C_j:

$$I(C_i, C_j): P(C_i) \times P(C_j) \to \mathbb{R}$$

This function may be nonlinear, stochastic, or dependent on external conditions.

7.4. Global Property of the System

The property of the system $G(S)$ is determined as an aggregated function:

$$G(S) = F(I(C_1, C_2), I(C_2, C_3), \ldots),$$

where F is an aggregation function (e.g., sum, mean, integral, etc.).

7.5. Condition of Emergence

Emergence arises if the global property $G(S)$ cannot be predicted solely from the local properties $P(C_i)$ of the components.

Formally:

$$\exists G(S): G(S) \neq \sum_{i=1}^{n} g(P(C_i)),$$

where $g(P(C_i))$ is a function of the local properties of the components.

7.6. Metric Evaluation of Emergence

For quantitative evaluation of emergence, one may use the distance between the actual system behavior and its predicted behavior:

$$E(S) = ||G(S) - \hat{G}(S)||,$$

where $\hat{G}(S)$ is the model of system behavior constructed on the basis of the properties of individual components without considering their interactions.

7.7. Use in Data Cybersecurity Systems

System components C_i: servers, routers, software.

Component properties $P(C_i)$: resilience to attacks, throughput capacity, response speed.

Global property $\hat{G}(S)$: overall resilience of the system to cyberattacks.

Emergence: within the system, an unexpected vulnerability may arise that cannot be foreseen solely by analyzing the properties of individual components.

This formalization allows the modeling of emergence in complex systems and enables the study of how new properties arise from the interaction of components.

In the context of your dissertation, it may be useful for describing the emergence of properties in digital archives when implementing intelligent algorithms.

8. Requirements for Countermeasure Systems Against Threats to Intellectual Information Aggregation Processes in Digital Archives

A system designed to counter threats associated with the use of large language models (LLMs) must be comprehensive, integrating technical, organizational, and ethical measures. Such a system should operate across all stages of data handling, from collection to the generation of results, thereby reducing risks and ensuring both secure and responsible use of LLM technologies.

From a general perspective, several foundational requirements can be identified. First, the system must be scalable, capable of processing large volumes of data and adapting flexibly to newly emerging threats. Second, it should support seamless integration with existing archival infrastructures and software tools. Continuous monitoring and analytic functionality represent another essential element, enabling the detection and reporting of threats in real time. Finally, the system must be accompanied by thorough documentation describing its functions, protective procedures, and operational mechanisms.

When considering protection against adversarial attacks (AA), the system must address attempts to deceive ML or DL models through deliberately manipulated input data. These attacks typically involve subtle perturbations that remain invisible to human observers yet result in misclassification or erroneous decision-making. To counter such risks, models should be designed or adapted to exhibit robustness against adversarial perturbations. Regular testing for vulnerabilities, the deployment of filtering mechanisms for suspicious inputs, and the establishment of continuous monitoring procedures are critical. Equally important is the training of personnel in recognizing and responding to adversarial scenarios.

Another requirement relates to defense against malware. Since malicious software represents a persistent risk, the system should ensure regular updating of libraries and dependencies, use only verified software sources, and incorporate vulnerability scanning. Security policies concerning third-party libraries must be enforced, while systematic audits of both software and AI models should be conducted to detect potential biases. Addressing such biases through retraining, alongside ethical training for developers, forms part of the organizational response to this threat.

Protection against data manipulation requires the verification of data sources and the implementation of algorithms capable of detecting falsified or corrupted information. Regular archival audits and verification procedures serve as preventive measures, while staff must be adequately trained to recognize manipulation attempts.

Similarly, prompt injection constitutes a distinct threat category. Countermeasures here include limiting the complexity of user queries, filtering inputs to block potentially harmful prompts, and monitoring query logs. Security policies governing the formulation of prompts are necessary, as is user training in secure and responsible interaction with AI systems.

With respect to data exfiltration, the system must ensure encryption of data both in transit and at rest, implement rigorous access control policies, and conduct regular vulnerability audits. Training users in cybersecurity awareness complements these technical safeguards.

A further concern involves disinformation. To prevent the introduction of false or manipulative content into digital archives, the system must validate outputs, rely on verified training data, and conduct systematic audits of models. Verification of results and training personnel in disinformation detection provide additional resilience.

Finally, measures are required to counter the malicious use of generated outputs. This includes filtering harmful queries, monitoring the use of LLM outputs, and applying mechanisms to restrict access where necessary. Establishing policies for the ethical use of results, supported by user training in responsible AI practices, is crucial for minimizing risks.

Taken together, these requirements underscore the necessity of a holistic approach to cybersecurity in digital archives. By integrating technical safeguards with organizational procedures and ethical principles, such systems can provide robust protection against a wide spectrum of threats, ensuring both the resilience and trustworthiness of intellectual information aggregation processes.

Conclusions

This article presents a comprehensive approach to the cybersecurity of Intellectual Information Aggregation (IIA) processes in digital archives. The proposed framework encompasses the formalization of digital lifecycle stages (digitization, image processing, OCR, classification, indexing, and archive system creation) together with corresponding threats and countermeasures,

with a focus on mathematically justified safeguards (encryption, multi-factor authentication, monitoring, anomaly detection, metadata protection, adversarial training). The models address risks arising in the automated collection, structuring, semantic enrichment, and analysis of data using AI/ML/LLM.

Key classes of threats to IIA are systematically analyzed and mathematically formalized, including data exfiltration, disinformation, digital signature forgery, as well as scenarios inherent to digital archives interacting with AI/ML/LLM (prompt injection, data manipulation, etc.). For each class, formal definitions and predicates are introduced, enabling the verification of security properties and the construction of verifiable countermeasures.

The scientific novelty of this work lies in:

- Introducing and methodologically substantiating the concept of intellectual information aggregation as a set of processes ensuring semantic normalization, contextual classification, and adaptive data processing through AI/ML/LLM;
- Formalizing the "process–threat–countermeasure" relations for the IIA stages in digital archives;
- Emphasizing the emergent properties of combined protection, where system resilience increases due to the interaction of individual safeguards.

It is demonstrated that protection integration must cover all stages of data processing — from acquisition to result generation — and adhere to the requirements of countermeasure systems: scalability, compatibility with existing infrastructure, continuous monitoring and analytics, and documented operational procedures. This enables real-time detection and localization of adversarial influences on models, reducing the likelihood of erroneous decisions.

The classification and description of IIA processes in digital archives are provided as a consistent sequence of production stages (P10–P90), facilitating the mapping of specific threats and countermeasures to concrete operations within the archive lifecycle. Such decomposition establishes a basis for constructing metric risk profiles and conducting compliance audits.

A significant contribution of the study is the formalization of the emergent properties of countermeasure systems: expressed through computable metrics of the distance between the actual behavior of integrated defense and the predicted "sum of local properties." This enables a quantitative assessment of the integral resilience of the archive as a cyber-socio-technical system.

Practical relevance is confirmed by implementations protected by intellectual property rights: technological solutions for converting large sets of paper documents into digital resources, the DIGITAL DOCS® software modules, and the associated trademark. These instruments have been applied in production environments for the implementation of IIA processes.

The results obtained can be used to evaluate and strengthen the cyber-resilience of information systems in the context of national and societal digital transformation — particularly where archival data support analytics and evidence-based decision-making.

Future research directions include expanding the formal apparatus of countermeasures against new categories of threats accompanying the evolution of AI/ML/LLM, advancing models of emergent resilience with consideration of socio-technical factors, and scaling integrated monitoring and security auditing methods for multi-component archival ecosystems (digital archives).

References

[1] Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2021). Archives and AI: An overview of current debates and future perspectives. ACM Journal on Computing and Cultural Heritage (JOCCH), 15(1), 1–15. https://doi.org/10.1145/3479010

[2] Jaillant, L., & Caputo, A. (2022). Unlocking digital archives: Cross-disciplinary perspectives on AI and born-digital data. AI & Society, 37(3), 823–835. https://doi.org/10.1007/s00146-021-01367-x

[3] Sun, Y., Yang, W., & Liu, Y. (2024, June). The application of constructing knowledge graph of oral historical archives resources based on LLM-RAG. In Proceedings of the 2024 8th International Conference on Information

System and Data Mining (pp. 142–149). https://doi.org/10.1145/3686397.3686420

[4] Zhang, S., Peng, S., Wang, P., & Hou, J. (2024). Archives meet GPT: A pilot study on enhancing archival workflows with large language models. In iConference 2024 Proceedings. University of Illinois. https://hdl.handle.net/2142/122806

[5] Lande, D., & Strashniy, L. (2023). GPT Semantic Networking: A Dream of the Semantic Web – The Time is Now. Kyiv: Engineering. ISBN 978-966-2344-94-3. https://doi.org/10.5281/zenodo.14278893

[6] Jaillant, L. (2022). Archives, access and artificial intelligence. Bielefeld: Bielefeld University Press. ISBN 978-3-8376-5584-1.

[7] Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoupova, T., & Stuart, K. (2019). More human than human? Artificial intelligence in the archive. Archives and Manuscripts, 47(2), 179–203. https://doi.org/10.1080/01576895.2018.1502088

[8] Donaldson, D. R., & Bell, L. (2019). Security, archivists, and digital collections. Journal of Archival Organization, 15(1–2), 1–19. https://doi.org/10.1080/15332748.2019.1609311

[9] Gupta, P., Sehgal, N. K., & Acken, J. M. (2024). Evolution and risks of LLMs. In Introduction to machine learning with security: Theory and practice using Python in the cloud (pp. 371–379). Cham: Springer. https://doi.org/10.1007/978-3-031-59170-9_12

[10] Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., & Zhang, Y. (2024). A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly. High-Confidence Computing, 100211. https://doi.org/10.1016/j.hcc.2024.100211

[11] Moorthy, E., Shanthakumar, M., & Janarthanam, S. (2022). Intellectual data aggregation using independent cluster-based Medicaid method for network functionality fabrication. Indian Journal of Science and Technology, 15(44), 2432–2440. https://doi.org/10.17485/IJST/v15i44.1555

[12] Abbasi, A., Zubair, M., Ali, R., Khan, S., & Hameed, I. A. (2023). Federated learning-assisted data aggregation scheme for smart grids. Applied Sciences, 13(17), Article 9813. https://doi.org/10.3390/app13179813

[13] AIIM. (2000). Enterprise content management (ECM) definition. Retrieved from https://info.aiim.org/what-is-ecm

[14] Gartner. (2017). Definition of content services platform (CSP). Retrieved from https://www.gartner.com/en/information-technology/glossary/content-services-platform-csp

[15] Tsyrulnev, Y. B. (2024). DIGITAL DOCS: Certificate of trademark registration No. 342775, 17.01.2024. State Organization "Ukrainian National Office of Intellectual Property and Innovations." Retrieved from https://sis.nipo.gov.ua/uk/search/detail/1684173/

[16] Tsyrulnev, Y. B., Tsyrulnev, A. Y., Yevdokymov, A. O., Mykhailovskyi, N. Y., & Markelova, K. Y. (2021). Method for converting an array of paper documents into a digital archive of electronic information resources and documents: Utility model patent No. 147523, filed 01.03.2021, published 12.05.2021, Bulletin No. 19. State Enterprise "Ukrainian Institute of Intellectual Property." Retrieved from https://sis.nipo.gov.ua/uk/search/detail/1593282/

[17] Tsyrulnev, Y. B., & Yevdokymov, A. O. (2021). Software modules package for electronic databases and digital archives of digital information resources and documents (DIGITAL DOCS TECHNOLOGY, DDT): Copyright registration certificate No. 102327, 04.02.2021. State Enterprise "Ukrainian Institute of Intellectual Property." Retrieved from https://sis.nipo.gov.ua/uk/search/detail/1578815/

[18] Tsyrulnev, Y. B., & Mykhailovskyi, N. Y. (2021). Device for document binding: Utility model patent No. 146181, filed 20.10.2020, published 21.01.2021, Bulletin No. 3. State Enterprise "Ukrainian Institute of Intellectual Property." Retrieved from https://sis.nipo.gov.ua/uk/search/detail/1473480/