

INTELIGENȚA ARTIFICIALĂ

Tehnologie emergentă în soluțiile de arhivare electronică

Aplicații, beneficii și provocări pentru infrastructurile digitale de cunoaștere

Cuprins

- 1 Contextul și provocarea arhivării moderne
- 2 AI aplicată în arhivistică — concepte cheie
- 3 Domenii de aplicare
- 4 Beneficii și rezultate
- 5 Provocări și riscuri
- 6 Exemple și inițiative
- 7 Concluzii și direcții viitoare

Provocarea: volumul de date crește mai rapid decât capacitatea umană de procesare

Instituțiile de arhivă se confruntă cu presiuni tot mai mari, sintetizate clasic prin cei „trei V ai datelor mari”, aplicabili direct fondurilor arhivistice digitale.



VOLUM

Fonduri fizice și digitale în continuă extindere, adesea depășind capacitatea echipelor de a le indexa manual.



VARIETATE

Documente text, manuscrise, imagini, hărți, înregistrări audio-video — formate eterogene greu de tratat unitar.



VITEZĂ

Ritmul de creare și digitizare a conținutului depășește ritmul clasic de catalogare și descriere arhivistică.

AI în arhivistică: trei tehnologii cheie

Termenul „inteligență artificială” acoperă, în context arhivistic, câteva familii distincte de tehnici, folosite adesea complementar.



ML

Machine Learning

Algoritmi care învață din exemple adnotate pentru a clasifica documente, a prezice priorități de digitizare sau a detecta anomalii în fonduri.



NLP

Procesarea limbajului natural

Extragere de entități (nume, locuri, date), rezumare automată, traducere și analiză a conținutului textual al documentelor.



CV

Computer Vision

Recunoaștere optică a caracterelor (OCR/HTR), segmentarea paginilor, identificarea deteriorărilor fizice ale suporturilor.

Recunoaștere optică avansată a textului (OCR/HTR)

Sistemele de tip Handwritten Text Recognition, antrenate pe corpusuri de manuscrise, permit transformarea documentelor scrise de mână — inclusiv cele istorice, cu grafii vechi — în text digital căutabil.

Aceasta reduce drastic timpul de transcriere manuală și deschide accesul cercetătorilor la fonduri altfel greu consultabile.

Manuscris



Text digital căutabil

Beneficiu direct: acces la conținut, nu doar la imaginea paginii.

Clasificare automată și generare de metadata



Etichetare automată

Modelele de clasificare atribuie categorii tematice, tipuri de document și niveluri de confidențialitate pe baza conținutului.



Extragere de entități

Identificarea automată a numelor de persoane, instituții, locuri și date calendaristice pentru îmbogățirea descrierii arhivistice.



Aliniere la standarde

Generarea de metadata compatibile cu scheme consacrate (Dublin Core, EAD), facilitând interoperabilitatea între instituții.

Căutare semantică și regăsirea informației

Căutare clasică pe cuvinte-cheie

Returnează doar documentele care conțin exact termenul căutat, ratând sinonime, variații ortografice sau formulări vechi.

Căutare semantică bazată pe AI

Reprezentări vectoriale ale sensului textului permit regăsirea documentelor relevante chiar și fără potrivire exactă de cuvinte, inclusiv între limbi diferite.

Impact practic: cercetătorii pot regăsi documente relevante folosind întrebări formulate în limbaj natural, nu doar termeni exacti din index.

Conservare predictivă și prioritizarea digitizării

1

Analiza stării fizice

Modele de viziune artificială detectează semne incipiente de degradare a suporturilor (hârtie, peliculă, benzi magnetice).

2

Estimarea riscului

Algoritmii estimează probabilitatea și viteza de deteriorare pentru fiecare unitate arhivistică.

3

Prioritizarea acțiunilor

Fondurile cu risc ridicat sunt propuse prioritar pentru digitizare sau intervenții de conservare.

Beneficii cheie

V Viteză

Procesare a mii de pagini în timp ce metodele manuale ar necesita luni de muncă.

S Scalabilitate

Aceleași modele pot fi aplicate unor fonduri de dimensiuni foarte diferite, fără efort proporțional.

A Accesibilitate

Conținutul devine căutabil și disponibil unui public mult mai larg de cercetători și cetățeni.

E Eficiență pe termen lung

Automatizarea sarcinilor repetitive eliberează timp pentru expertiza arhivistică de valoare adăugată.

Provocări și riscuri

Adoptarea AI în arhivistică nu este lipsită de riscuri, care trebuie gestionate prin guvernanță și expertiză umană.



Bias algoritmic

Modelele pot reproduce sau amplifica părtiniri prezente în datele de antrenament, afectând echitatea clasificării.



Autenticitate și proveniență

Este necesară păstrarea trasabilității complete între documentul original și rezultatele generate automat.



Costuri inițiale

Implementarea infrastructurii AI presupune investiții în tehnologie, date de antrenament și formare a personalului.



Guvernanță și expertiză umană

Validarea arhivistică rămâne indispensabilă — AI asistă, nu înlocuiește judecata profesională.

Exemple și inițiative relevante

Transkribus

Platformă specializată în recunoașterea automată a textului scris de mână (HTR), folosită pe scară largă pentru manuscrise istorice.

Europeana

Platformă europeană care agregă și facilitează accesul digital la patrimoniul cultural și arhivistic al instituțiilor membre.

Inițiative naționale

Biblioteci și arhive naționale integrează tot mai frecvent module de AI în fluxurile proprii de digitizare și descriere.

Concluzii și direcții viitoare

- AI nu înlocuiește arhivistul, ci îi extinde capacitatea de a gestiona volume mari de conținut eterogen.
- Beneficiile în viteză, accesibilitate și scalabilitate sunt condiționate de o guvernare atentă a datelor și modelelor.
- Colaborarea între instituțiile de arhivă, cercetători și dezvoltatori de tehnologie rămâne esențială pentru rezultate de încredere.
- Următorul pas firesc este integrarea treptată a AI în fluxurile existente, cu evaluare continuă a calității rezultatelor.

Mulțumesc!

Întrebări și discuții

Mădălina Giurgescu