CRIS2024
The power of generative AI for CRIS systems: a new paradigm for scientific information management

May 15, 2024

Anna Guillaumet
Anna.Guillaumet@sigmaaie.org
Aurelia de Andrés
aurelia.andres@fecyt.es

SIGMA
helping universities succeed

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

FECYT
INNOVACIÓN

euroCRIS
THE INTERNATIONAL ORGANISATION FOR RESEARCH INFORMATION

Contents

01

Introduction

1.1. The evolution of CRIS - Timeline

Summary of the evolution of CRIS systems since the term was coined in 1970.

1970s-1980s

- **Birth of CRIS:** At the end of the 70s, ***mainframe-based CRIS systems*** focused on basic functionality, focusing on the need to manage research information more efficiently. Early CRIS systems are developed primarily as ***institutional repositories*** to manage publications and research data.

1990-2000s

- **Standardization and expansion:** more widespread in Europe, establishment of ***standards*** for data and exchange interoperability. ***EuroCRIS*** **was created to help with the standardization** (2002). Integration with other institutional systems such as HR and Finance

2000s

- **Open Access:** The push for Open Access leads to **CRIS systems becoming *key* in managing and providing access to research outputs**.

2010s

- **Globalisation and metrics:** CRIS systems are **increasingly used for research *assessment* and performance metrics, influencing *funding* and *policy* decisions**. CRIS/IR ***survey*** by EUNIS and euroCRIS (2016)

2020s

- **Analytics and AI:** CRIS systems **start to incorporate AI and advanced analytics**, enhancing their capabilities for predictive modeling and ***complex data analysis*****.**

1.1. The evolution of CRIS systems - Today

ORCID
unpaywall
Scopus
Google Scholar
WEB OF SCIENCE
OpenAlex
PubMed
Dialnet
Altmetric

Bibliometrics

Legacy systems (HR, Finance)

CRIS
CERIF

RMS

Funding
Projects
Contracts
Research groups
Profiles
CV
Outputs

For years, efforts have been made to turn this vision into reality. CRIS systems aim to store data that is highly interoperable, free of redundancies, well-classified, and reusable. Having information organized in this manner enables us to undertake AI integration projects with significant confidence.

1.2. The evolution of AI - timeline

We could see similarities in the evolution of AI with respect to the evolution of CRIS systems as technological changes occurred over the last decades.

1950s

- **Birth of AI:** The term “Artificial Intelligence” was officially **coined at the 1956 Dartmouth conferenc**e. The first notable AI system was Theseus, a remote-controlled mouse that could find its way out of a labyrinth and remember its course.

1960s

- **AI Boom:** The creation and popularization of the **Perceptron artificial neural network** sparked a great deal of interest in the field.

1970s-1980s

- **AI Winter and Renaissance:** After an initial period of enthusiasm, AI research faced a period of **reduced funding and interest**, known as the “AI winter”. This was followed by a renaissance in the 1980s.

1990s

- **Speech and Video Processing:** The **1990s saw significant advancements** in speech and video processing.

2000s

- **Imitating humans:** AI systems **beating humans in tests** in domains like language understanding and image recognition..

All images created by DALL-E 3 (Bing image creator)

1.1. The evolution of AI – NEWS!!

On Monday, OpenAI launched a new model for ChatGPT that can process text, audio, and images. In a surprising turn, the company announced that this model, GPT-4o, would be available for free, no subscription required. fa 1 dia

A new model that reacts depending on your mood. Sad, happy, instant translations..

02

AI definition

2. AI definition

Principle of substitution by continuous progress:

“*any [natural] intelligence that performs an activity that can be improved by being performed by another [artificial] intelligence, will eventually be replaced by the latter*”.

By Sénen Barro (*Inteligencia Artificial: ¿quererla o Temerla? - YouTube*, n.d.)

2. Types of IA

Artificial Intelligence (AI) is generally classified into two types: weak AI and strong AI.

- **Weak AI:** narrow or specialized AI.

- Outperform humans with specific tasks for which it is designed
- operates under more restrictions than even the most basic human intelligence.
- It does not have consciousness or self-knowledge.

- Examples of weak AI include **Alexa, Siri, and Google Assistant**.

- **Strong AI:** Artificial General Intelligence (AGI) or deep AI.

- Solve problems and learn
- cognitive abilities similar to humans, which allows it to understand, learn, and solve problems autonomously.
- can understand and reason outside its designated function.

- Strong AI **does not yet exist,** but it seems to be on the horizon.

Source: https://www.digital-adoption.com/ai-vs-generative-ai/

2. New Types of AI

Generative Artificial Intelligence (advanced form of weak AI).

- Creates something new
- It can replicate human creativity in various domains, (visual arts, literature, music, and programming, etc).
- Is a **form of deep learning** that can produce text, code, or images in response to prompts. Still limited to specific tasks and does not possess the consciousness or self-understanding associated with strong AI.

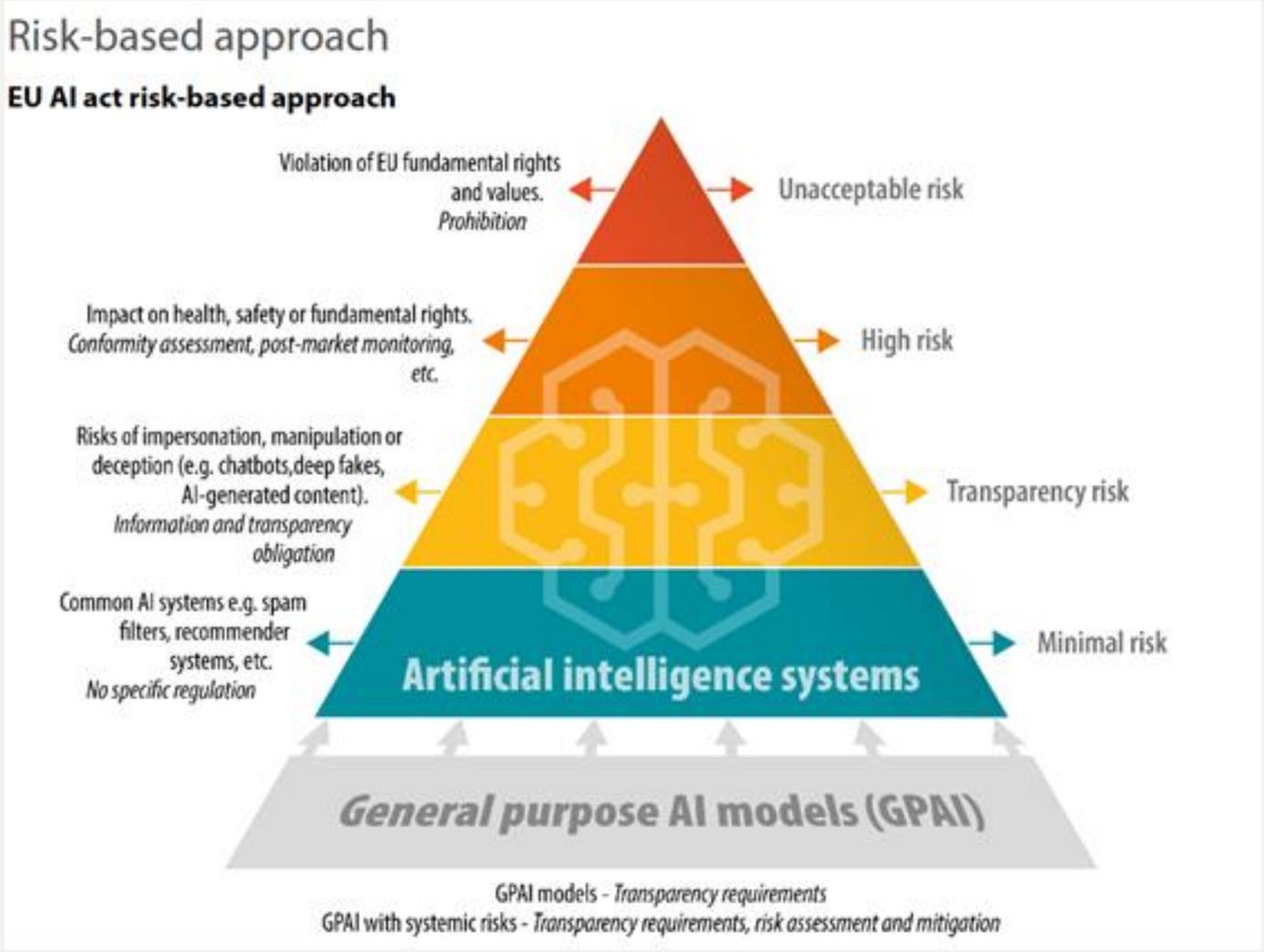
Artificial General Intelligence (AGI) (type of AI with capabilities that rival those of a human).

- Purely theoretical at this point, someday AGI could replicate human-like cognitive abilities including reasoning, problem-solving, perception, learning, and language comprehension, to pass, without problems, the Turing test.

03

AI law

4. IA Law Europe: EU AI Act, first regulation on artificial intelligence



The European AI Act was approved by the European Parliament with a large majority. The Regulation, agreed in negotiations with Member States in **December 2023**, received 523 votes in favor, 46 against, and 49 abstentions.

Objective:
To protect fundamental rights, democracy, the rule of law, and environmental sustainability from high-risk AI, while promoting innovation and positioning Europe as a leader in the sector.
The law establishes a classification of AI risks to ensure safe, ethical, and responsible use of artificial intelligence

Classification of AI risks:
- Unacceptable risk → prohibited
- High risk
- Limited risk
- Minimal/no risk.

Each member state should define their own law based on the Act.

Source: https://medium.com/generative-ai-revolution-ai-native-transformation/eu-passed-the-worlds-first-ai-law-what-it-means-for-you-and-the-future-eaf7ea7c31f2

Spanish strategy on Artificial intelligence (ENIA)

The Spanish Government released its **National AI strategy in December 2020**, **aligned with the Digital Spain strategy 2026,** to develop a policy framework defining the various actions that the governmental administrations will undertake to facilitate the development and deployment of AI in the economy and society.

ENIA objectives in relation to Artificial Intelligence and the Data Economy:

- To turn Spain into a benchmark in the transformation towards a Data Economy.
- Promote Artificial Intelligence as a driver of innovation and social, inclusive and sustainable economic growth.
- Establish regulatory frameworks that delimit and guide the design of AI, so that the resulting applications respect the rights of citizens.
- Foster data and technological infrastructures that support AI.
- Strengthen competitiveness through R&D activities in the field of Digital Enabling Technologies (DETs).

MINISTERIO DE ECONOMÍA, COMERCIO Y EMPRESA

MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA

ENIA actions and measures in the field of IA and data economy

- Creation of the Data Office and the figure of the Chief Data Officer
- Launch of the Artificial Intelligence Advisory Board
- Publication of the Charter of Digital Rights
- Approval of the National Plan for Language Technologies.
- PERTE: New Language Economy
- Quantum Spain initiative, a strategic supercomputing project
- GAIA-X, Spanish hub for the national development of sectorial data space and data management
- National Program of Green Algorithms (PNAV)
- Regulatory sandbox for the implementation of the future European AI Regulation.
- Creation of the National Centre for Neurotechnologies
- Creation of the National Agency for Artificial Intelligence Oversight

Spanish strategy on Artificial intelligence (ENIA)

MINISTERIO DE ECONOMÍA, COMERCIO Y EMPRESA

MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA

ENIA impact for universities and research centers:

1. **Funding**: ENIA allocated funds to finance over 250 innovative projects, to drive technological innovation for the benefit of citizens and society.

2. **Talent Development Programs:** initiatives like the **Spain Talent Hub** to attract and retain AI talent at leading Spanish centers and develop new high-potential research lines, to enhance the skills and knowledge of researchers and students in the field of AI.

3. **Collaboration with Industry:** through the creation of **ENIA Chairs**, such as the "Cátedra ENIA UAM-VERIDAS" in Responsible AI.

4. **International Cooperation:** offering opportunities for Spanish researchers to engage with **global AI initiatives**, sharing knowledge, resources, and best practices, fostering a collaborative environment for AI development.

5. **Advanced Training Opportunities:** The creation of specialized positions for AI research, including **PostDoc and PhD positions** with competitive salaries and funding for research activities.

6. **Regulatory:** A first **regulatory sandbox** is being established in Spain in collaboration with the European Commission. It serves as a **testing space** where EU regulatory requirements for high-risk AI systems are experimented with, ensuring their suitability and efficacy before widespread application. And **AI quality seal**, to guarantee robust ethics for non-high risk IAs.

7. **Infrastructure Development:** **Spanish Neurotech** (Center specialized in Neurotechnology), **Quantum Spain** (First quantic computer) and the developing of **neuronal models in Spanish**.

Quantum SPAIN

Sandbox

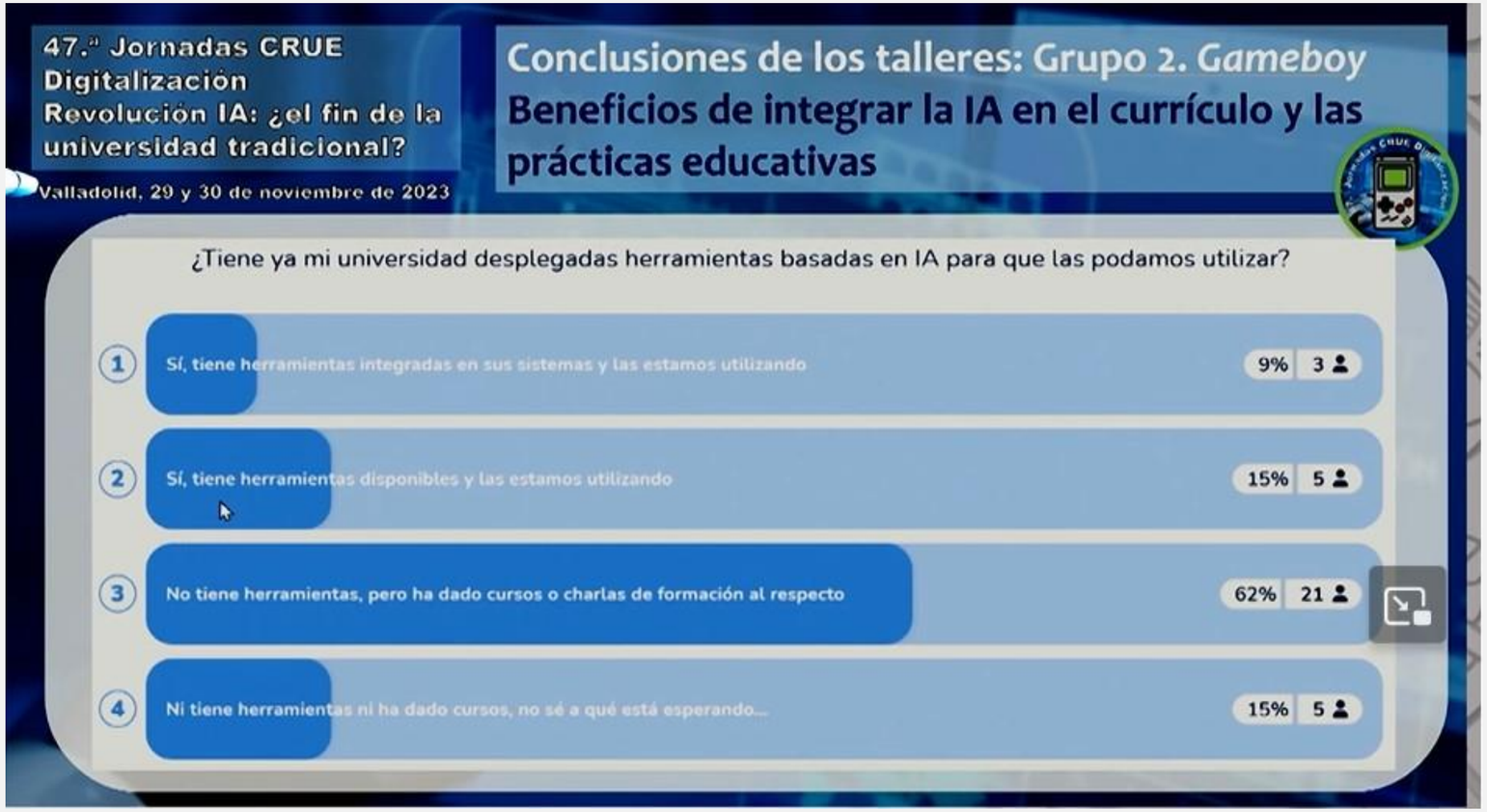
Spain Talent Hub

Access to Sandbox: https://portal.mineco.gob.es/es-es/digitalizacionIA/sandbox-IA/Paginas/sandbox-IA.aspx

Spanish strategy on Artificial intelligence (ENIA)

But …

An examination of the implementation of AI in CRIS systems within Spanish universities reveals it to be in its early stages, with the application of AI being rather limited. In fact, more than 62% of institutions surveyed do not yet utilize AI-related tools. Thus, there is still significant progress to be made in this area.



Source: https://www.youtube.com/watch?v=fwWxqqejR48&list=PLSbo9kXA_LcykgmcHLm6j3AuY3UG7mmgz&index=5

AI in knowledge production

3. AI in knowledge production

According to scientific literature, Artificial Intelligence (AI) will revolutionize various aspects of research. The key domains that will be impacted are:

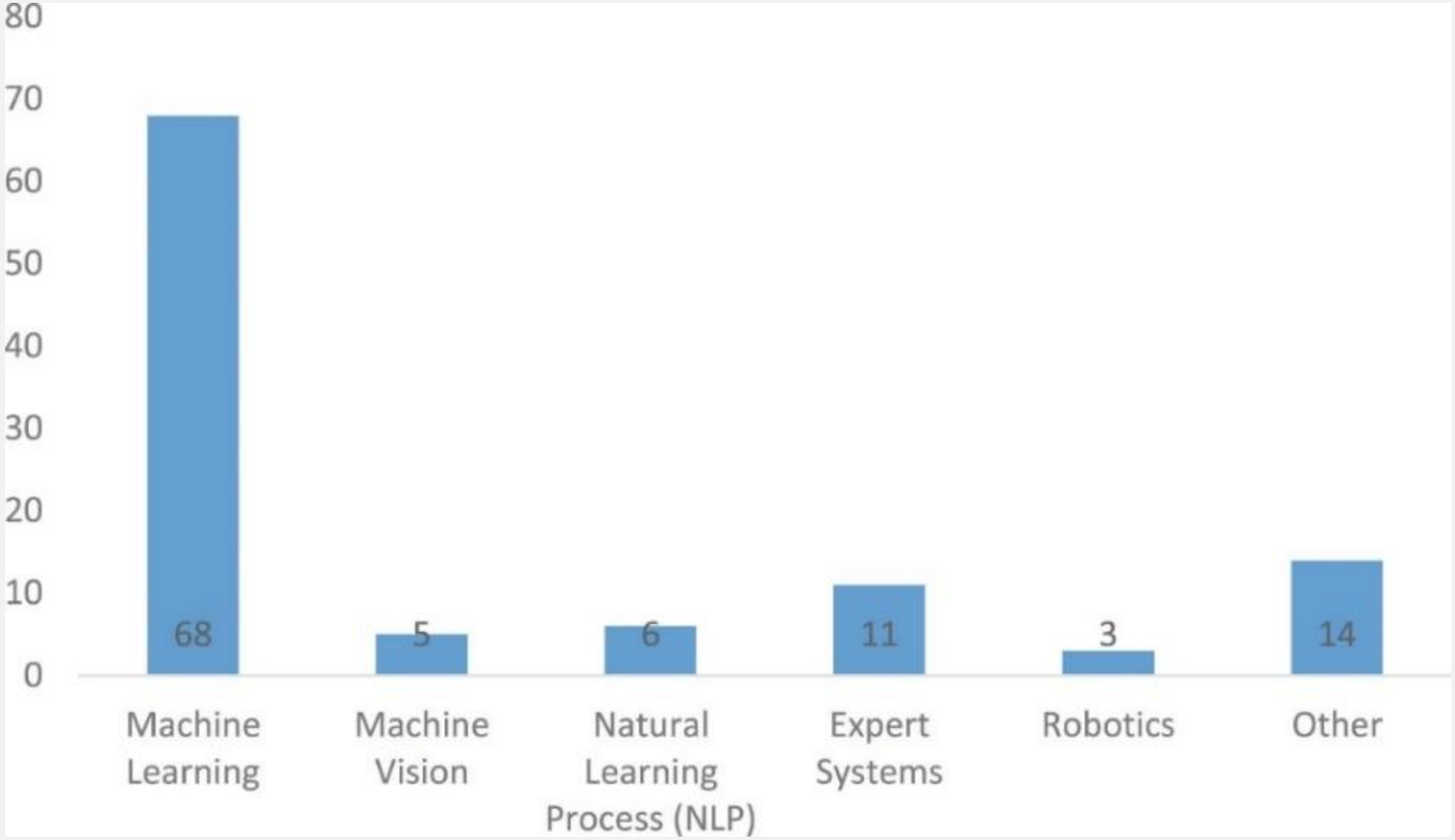
Funding	• Identifying **promising research areas**, optimizing fund allocation by predicting which research endeavours are more likely to succeed and **enhancing public-private collaboration**, driving demand for applied research.
Outputs	• Accelerating research pace by helping researchers finding **patterns, keywords, trends, assisting writing, foster collaborations,** applying different AI algorithms.
Standards	• **New standards, guidelines, and protocols** to ensure ethical and reliable practices. European Union new AI standards. **Data reusability, accessibility and CERIF** should be improved with AI.
Infrastructures	• Investment in massive Computing Power with a robust cloud infrastructure. Establishing a European institute for AI in science would provide **high-performing computational resources**. (data storage, model training, and deployment). Investment in high data quality and accesibility, AI training and ethical & responsable adoption of AI
Legal aspects	• Adjustments in **intellectual property rights, authorship & contributions**, **data privacy & protection, ethical standards, liability & accountability, quality contro**l, etc.
Collaboration	• **Removing language barriers** with real-time translations, facilitating **resource sharing**, potential collaboration and complex project management.

By considering these six aspects, researchers can embrace the potential of AI and foster a collaborative, innovative, and ethical research ecosystem.

3. AI in knowledge production

As CRIS systems have several purposes for researchers, research managers and decision makers, different AI algorithms can be incorporated that support these profiles and facilitate their work.

Some of the most common types of AIs used in research are, according to ((Collins et al., 2021) are:



Generative AI, with its capacity for natural language processing (NLP) and machine learning (ML), offers unprecedented opportunities to support the multifaceted needs of researchers, research managers, and decision-makers.

A **combination of AI techniques** can be used to help the research environment.

05

Use cases

CRIS USE CASES

DATA management and Classification

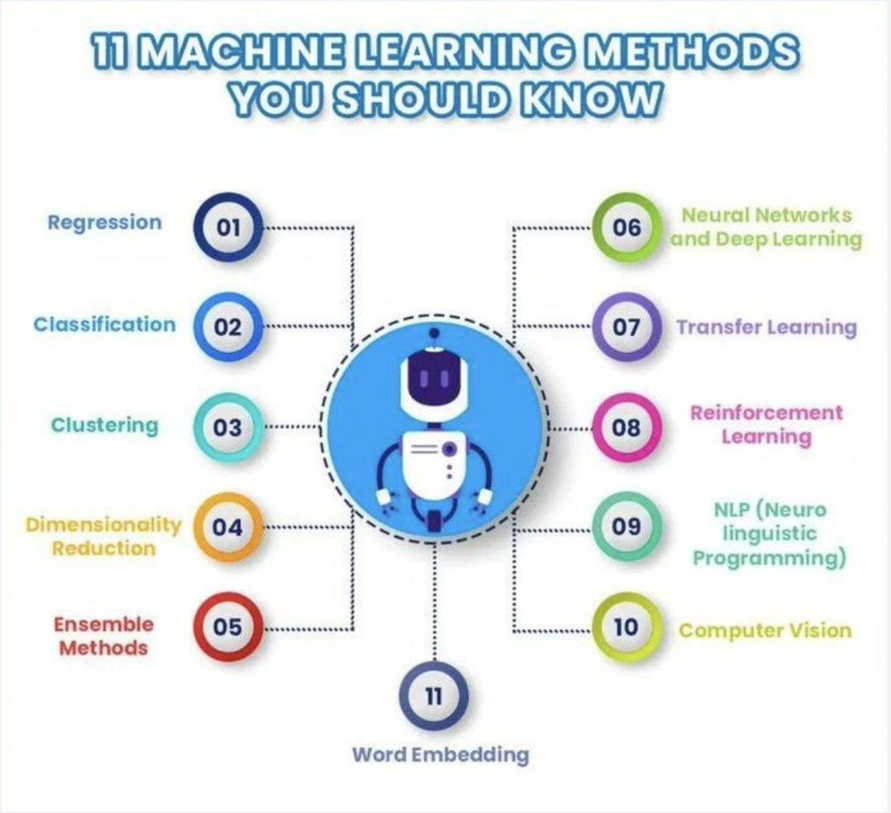
CRIS systems can benefit from the implementation of AI tools and models. Some examples:

1. Researchers profile automatization:

- **Areas of expertise** → *Text classification* algorithms (like SVM) or neural networks, classify documents into predefined categories.
- **Biography, personal and academic data** → *Named entity recognition* algorithms (NER) identifies and classifies key elements in the text.
- **Abstracts** → LLM, automatic summarization, *deep learning* and recognition systems algorithms, can generate summaries for one or more documents.
- **Impact** → *Sentiment analysis* algorithms access opinions and emotions expressed in publications or comments about researchers.

2. Research Groups and projects classification:

- **Similar interests' classification** → *Clustering* algorithms such as K-Means or hierarchical clustering to automatically organize researchers' profiles into groups based on similarities in their research interest and publications.



Note: All these automatisms are an aid and will require the supervision of the researchers.

CRIS USE CASES

DATA management and Classification

3. Areas of research

- **Emerging research areas** → Using *predictive analysis* with *topic model* algorithms like LDS to discover hidden topics in documents collections which can reveal emerging areas of research.

- **SDG classification** → Using *ML* algorithms (for pattern recognition) and *NL* processing (for text classification) to analyze large datasets and identify patterns and trends that align with specific SDGs and order it by relevance. Can assign SDG labels to research abstracts and publications

Note: All these automatisms are an aid and will require the supervision of the researchers.

CRIS USE CASES

CERIF standard

CERIF as a standard for research information systems such as CRIS systems can benefit from the implementation of AI tools and models. For example:

1. Semantic enrichment

Natural language processing (NLP) techniques can extract meaningful information from research documents and link it to CERIF entities. Researchers could use AI tools to automatically annotate research outputs with relevant CERIF concepts, improving data quality and interoperability

2. Automated data integration

CERIF could incorporate AI-based *data mapping and alignment techniques* to harmonize research data from various CRISs, repositories, and databases.

3. Enhanced collaboration networks

AI can analyze *collaboration patterns* and recommend potential research partners. CERIF could evolve to include AI-generated collaboration networks, fostering interdisciplinary research.

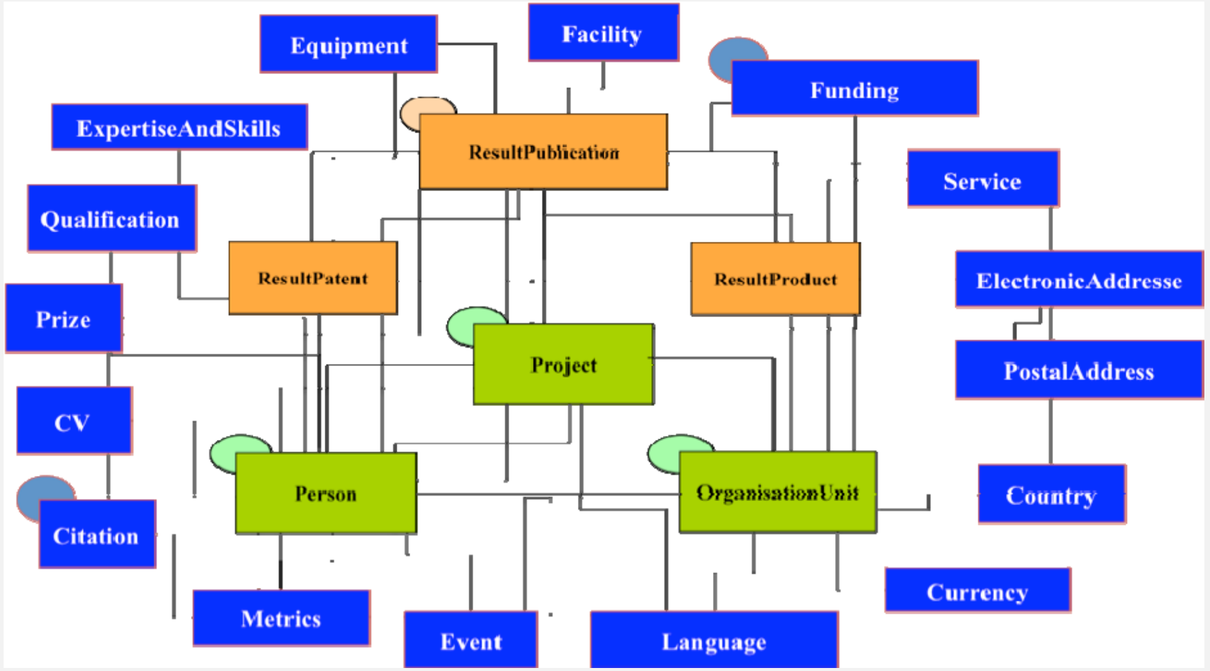
Note: All these automatisms are an aid and will require the supervision of the researchers.

CRIS USE CASES

CERIF standard

4. Quality assurance and peer-review

CERIF might integrate AI-driven metrics for assessing research impact, credibility, and novelty.



However, CERIF should address ethical aspects. Guidelines for responsible AI use within CERIF should be established to prevent biases and ensure transparency.

06

IntelComp Project

An example that analyzes large volumes of heterogeneous data from the research ecosystem to obtain key information, which is one of the great benefits that AI can bring, as we have seen, to the field of research.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004870. H2020-SC6-GOVERNANCE-2018-2019-2020 / H2020-SC6-GOVERNANCE-2020

IntelComp project

OpenSource project coordinated by **FECYT (Spanish Foundation for Science and Technology)**

Project goal: artificial intelligence-based platform for analyzing large volumes of textual data. This platform automates tasks within Public Administrations related to science, technology, and innovation policy, involving stakeholders such as academia, businesses, and citizens.

Definition:

Textual Information Exploitation

The integrated open-source platform includes:

- Tools for Public Administrations to diagnose the context in STI policy, evaluate proposals for funding and train, edit and curate AI models.
- AI services (topic modelling, automatic classifiers –FOS, SDGs, subcorpus generation and graph analysis) **useful for CRIS systems.**

The tools have been tested in **three specific domains through living labs**: R&D in artificial intelligence, Science and innovation activities for the sustainability of the energy and agri-food sectors and Cancer research.

Intelcomp Platform: https://intelcomp.eu/services; https://github.com/IntelCompH2020

Data sources:

Science, Technology, and Innovation (STI) Space Data

Large datasets related to science, technology, and innovation. These dynamic, multilingual, and heterogeneous data include information such as funded projects, scientific publications, patents, and more.

Processed within a High Performance Computing (HPC) environment as part of the European Open Science Cloud (EOSC), using the openAIRE graph (Barcelona Supercomputing Center).

IntelComp Data Catalogue: https://catalogue.intelcomp.bsc.es/search

STI – Science, Technology and Innovation.

IntelComp project

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004870. H2020-SC6-GOVERNANCE-2018-2019-2020 / H2020-SC6-GOVERNANCE-2020

intelcomp

EUROPEAN OPEN SCIENCE CLOUD

AI Tools and Services

- **Topic Modelling**: An unsupervised analysis of any set of documents, extracting from the text the main themes the collection of documents deals with, and the relation of each document with the identified topics.

- **Automatic Classifiers:** Service for document taxonomical classification (FOS, SDG,…)

- **Service for Subcorpus Generation:** Classification of all documents from a corpus according to a category that is specified by the user (an expert in the domain).

- **Service for Graph Analysis**: Software for the computation, processing and analysis of graph collections.

- **Interactive Model Trainer**: Experts in a scientific or technological domain can train, edit, and curate topic models, as well as to generate sub-corpora and evaluate models in a user-friendly manner.

Potential Applications:

- **Scientific Research:** can assist researchers in discovering **new connections** between studies, identifying **potential collaborators**, and **exploring emerging areas**.

- **Information Management:** Academic and research institutions can use IntelComp to manage and visualize research-related data, such as publications, projects, and collaborators.

- **Agenda setting and policy formulation:** Policymakers can understand the current STI landscape and its evolution with advanced visualizations, analyzing structured and unstructured data, to help in observing emerging topics and detecting key players.

- **Evaluation of proposals for funding in programs supporting STI activities:** Officials at funding agencies can automatize some tasks of the evaluation process.

IntelComp Project - Access

More information: https://intelcomp.eu/

07

Conclusions

Conclusions

We have seen the pivotal role of AI in revolutionizing research processes, the importance of ethical considerations in AI implementation, and the potential for AI to drive innovation, collaboration, and knowledge discovery in the field of science, technology, and innovation.

- The evolution of CRIS systems and AI, concluding that the integration of AI and advanced analytics in CRIS systems in the 2020s has enhanced predictive modeling and data analysis capabilities.

- The transformative impact of AI on research funding allocation, scientific outputs, data standards, infrastructures, legal frameworks, and international collaborations, has been highlighted.

- The potential of AI to optimize research processes, drive innovation, enhance accuracy, and necessitate new standards and ethical practices in research is demonstrated.

- The rapid evolution of these technologies requires the creation of new regulations and laws for their regulation. To protect fundamental rights, promote innovation, and classify AI risks for safe and responsible.

- Projects like IntelComp Project, focused on extracting knowledge from unstructured text data related to science, technology, and innovation using AI-driven tools and natural language processing, to assist scientific research, managing research-related data, competitive intelligence analysis, and fostering collaboration in STI ecosystems.

Conclusions

So, no worries about chatbots becoming outdated anytime soon!

THANK YOU FOR YOU ATTENTION!

Questions, concerns, opinions?

Anna Guillaumet
Anna.Guillaumet@sigmaaie.org
Aurelia de Andrés
aurelia.andres@fecyt.es

SIGMA
helping universities succeed

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
FECYT INNOVACIÓN

euroCRIS
THE INTERNATIONAL ORGANISATION FOR RESEARCH INFORMATION