

Transforming INIS: Uses of AI in Information Management

Brian Paul Bales, International Nuclear Information System, INIS-IAEA

Since the 1950s, and before, artificial intelligence has cycled through repeated waves of enthusiasm, hype, and inevitable disappointment. In 1958, Frank Rosenblatt, inventor of the Perceptron, said that “we are about to witness the birth of such a machine—a machine capable of perceiving, recognizing and identifying its surroundings without any human training or control.” (quoted in Lefkowitz, 2019) Like many other technologies, AI has followed a familiar path: initial excitement, followed by disillusionment, and finally, a productive phase of realistic expectations and practical integration.

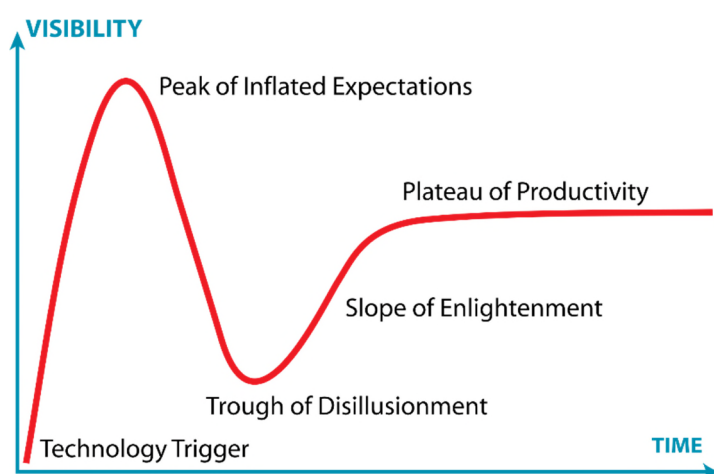


Figure 1: The Gartner Hype Cycle (Jeremykemp, CC BY-SA 3.0)

In recent months, AI use in Information Management appears to be entering a productive phase. While many applications remain unsuited for AI, with appropriate caution and oversight, AI tools can be successfully incorporated into IM workflows. This paper will present five current use cases illustrating successful integration and explore two prospective applications that hold promise for future development.

About INIS

The International Atomic Energy Agency’s International Nuclear Information System (INIS) was established in 1970 to collect, preserve, and disseminate knowledge products in all areas of nuclear science and technology. Its mission emphasizes relevance, currency, comprehensiveness, and accuracy. Today, INIS includes 134 Member States, who contribute journal articles, books, and reports, while the IAEA augments these submissions by harvesting bibliographic records and full texts from major publishers such as Elsevier, Springer, the Institute of Physics, and the American Physical Society.

INIS receives approximately 1.8 million unique users each year and has averaged more than 100,000 newly added records annually over the past two decades. However, estimates suggest that over 250,000 potentially relevant documents are produced each year within INIS’s scope. Many of these, particularly from conferences, smaller journals, and institutional reports, go

unrecorded, and their knowledge lost, meaning that the system falls short of its goal to be fully comprehensive and up to date.

Preparing for AI

For this reason and others, INIS embarked on a change in system. The previous system had been developed over 52 years, and the workflow had become unwieldy, with bottlenecks preventing the ability to process modern volumes of information. The system used a metadata schema called Tagged Text Format (TTF), which was based on MARC21. However, the schema had historical inconsistencies and anomalies in its usage.

In early 2025, INIS implemented Invenio Research Data Management repository software. Originally developed by the European Center for Nuclear Research (CERN), the system is open-source and a large group of organizations now contribute to its development, especially those in physics research. The system is well known and is based on the Data Cite Metadata Schema, which is developed and maintained by an even larger number of organizations.

Programming Assistance: AI as a Co-Worker

A key benefit of adopting widely used, standards-based systems is their visibility to large language models (LLMs), which are trained on a broad corpus of internet-based documentation. As a result, LLMs can readily generate useful code snippets, scripts, and integration tools for these platforms, requiring only minimal adaptation by a human with basic programming knowledge.

This has enabled INIS to use LLMs not merely as passive tools but as active, collaborative agents, similar to working with very knowledgeable remote co-workers who still need oversight guidance. For instance, INIS staff used prompts such as:

- Create a Python script that uses the American Physical Society's API to download bibliographic references, and put them in the InvenioRDM JSON format provided
- Create a Python script to enrich JSON files with affiliation data from Crossref
- Create a Python script to upload records and associated PDFs to Invenio RDM

While each script required refinement, the ability to create scripts quickly using LLM-generated code accelerated development. Additionally, AI systems were able to suggest resources for metadata enrichment and full-text, open-access downloads.

One of the key measures for the success of INIS is the number of records added each week. Historically, INIS has added between 2,000 and 2,500 items per week. Exceptional weeks in 2014 and 2015 saw highs of 6,752, 6,713, and 6,196 items added.

In 2025, with AI-assisted scripting and streamlined workflows, INIS achieved remarkable numbers: 17,733 records in week 12, and 10,987 in week 14. Notably, 30% of these records included open-access full-texts (with valid Creative Commons licensing), a substantial increase from the typical average of under 10%.

Categorization and Tagging With NADIA

One of the most resource-intensive stages in the INIS workflow has historically been subject classification and descriptor assignment. Each incoming record, whether a journal article, report, or conference paper, had to be reviewed by an expert in an area of nuclear science and technology who manually selected appropriate categories and index terms from the INIS

Thesaurus, a 31,000+ term multilingual controlled vocabulary specific to nuclear science and technology. This activity created a significant bottleneck. It was labor-intensive and marked by inconsistencies due to variations in interpretation among indexers.

To address these challenges, INIS developed NADIA (Nuclear Artificial Intelligence for Document Indexing and Analysis), a specialized, pre-LLM AI tool designed to automate subject indexing with greater efficiency and reliability. Index terms, titles and abstracts are vectorized with a customized version of BERT, fine-tuned specifically on nuclear science literature to capture domain-specific language and terminology. NADIA then uses 234 purpose-trained neural networks, which classify 11,500 of the most commonly used descriptors. For terms not used as often, a grid search algorithm is employed. This means that both commonly used, and lesser used terms will be selected for tagging.

NADIA processes new records nightly, analyzing titles and abstracts to assign subject categories and controlled vocabulary terms ("descriptors") in alignment with the INIS Thesaurus. The system has improved throughput and consistency. It applies rules uniformly, avoiding subjective drift. NADIA provides a "confidence score" and is not confident in the indexing of 3-4% of the records provided. This means that 96-97% are likely well indexed, which is much better than the initial target of 90%. It works with batches of 100 and is able to do a batch in about 40 minutes, a maximum rate of 3,600 per day. This is 7 times larger than the typical daily rate of 500 as achieved previously by a team of experts.

By removing this long-standing bottleneck, NADIA has allowed expert staff to focus on higher-value tasks such as quality assurance, outreach, innovation, and finding lesser known, but valuable sources of information.

PDF Analysis and Metadata Extraction

In 2022, INIS received a substantial archive of approximately 20,000 digitized reports and technical papers from Forschungszentrum Jülich in Germany. These documents captured decades of high-temperature gas-cooled reactor (HTGR) research, a programme discontinued following Germany's phase-out of nuclear energy. Despite their significance, many of the documents remained effectively inaccessible to the broader research community. Over two years, a bibliographic specialist had manually processed around 4,000 items. At that pace, ingest of the full collection would take 8 more years.

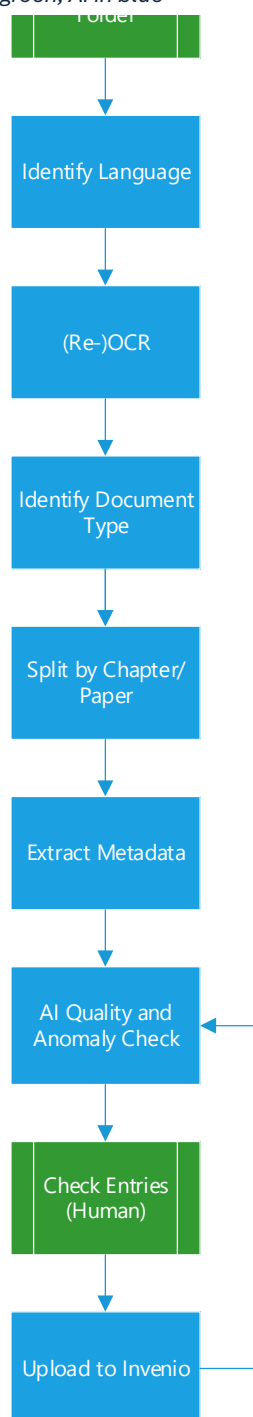
In 2025, however, INIS implemented a new AI-assisted metadata extraction workflow using GPT-4o-mini via Microsoft Azure OpenAI. By working through the API with a carefully constructed prompt (see Appendix) and a low temperature setting (0.2), the model was able to generate fully structured JSON files suitable for direct upload into the INIS repository. The results were both rapid and remarkably accurate. This work was anticipated and inspired by Savić (2024), who proposed a system of web scraping and metadata creation for INIS and other grey literature repositories.

Despite the disordered and often OCR-corrupted nature of the PDF text, full of formatting artifacts and layout noise, the system reliably identified key metadata elements: titles (in German and English), author names and affiliations, report numbers, publication dates, abstracts in both languages, and added static metadata elements (such as country of publication). A system prompt, re-sent with every API call, enforced a strict output format aligned with INIS standards. Even difficult translation challenges were handled well: for example, the term *Kugelhaufenreaktor*

(literally “ball heap reactor”) was correctly translated as “Pebble Bed Reactor,” rather than the misleading “Globular Cluster” rendered by online translation tools.

In early stages, the model made some interesting mistakes, such as translating “this paper” into the awkward *dieses Papier*, but these diminished after adjusting the prompt to require summaries in the original language followed by English translations. The model’s ability to infer structure from cluttered input was especially striking. Human readers might struggle to make sense of the raw extracted text, but the AI consistently found authors, affiliations, and summaries with apparent ease.

Figure 2: Proposed AIDAN Workflow, human actions in green, AI in blue



There are several important caveats to consider. This workflow is only reliable when used via the API. In interactive chat mode, while it may work the first time, the model’s performance deteriorates after a few rounds as the context becomes longer and the model becomes increasingly “creative.” Using the API, each call begins with a clean context - instructions are resent every time, and with the previously mentioned low temperature setting, which reduces randomness. Additionally, only the first 20,000 characters of each PDF are used for analysis, which generally captures the essential metadata. GPT-4o-mini, while not the most powerful model available, offers good performance at a fraction of the cost of GPT-4o (approximately 20 times cheaper).

Thanks to this workflow, the remaining 16,000 documents are now expected to be fully processed in just a matter of weeks.

Toward Full Automation: The AIDAN Workflow

Building on the success of AI-assisted metadata extraction, INIS is now exploring the development of a fully automated, multi-agent workflow system tentatively named AIDAN. Whereas NADIA focuses on classification and indexing, AIDAN (Automated Information Derivation and Analysis for Nuclear research) envisions a start-to-finish pipeline capable of transforming raw documents into structured, quality-assured knowledge products with minimal human intervention.

The conceptual model for AIDAN is agent-based, with each agent responsible for a specific task in the document processing lifecycle. The workflow would begin with a drop folder. PDFs deposited here would trigger the sequence. A first agent would identify the language of each document, followed by an OCR module to ensure clean, consistent text input. Even documents that have already undergone OCR might benefit from reprocessing, as newer tools offer better recognition and layout handling.

Next, a document-type agent would classify the item as a journal article, book, report, or conference proceeding. In the case of books or proceedings, a segmentation agent would split the volume into individual chapters or papers. Each sub-document would then be passed to a metadata extraction agent equipped with type- and

language-specific instructions, allowing it to extract titles, authors, affiliations, abstracts, descriptors, and report numbers with high accuracy.

A subsequent quality control agent would perform anomaly detection—flagging records with implausible values (e.g., publication dates in the future), missing critical metadata, or possible duplicates. An additional content-checking agent could provide further validation by identifying outdated, factually incorrect, or anomalous content that merits human review. Such content might be tagged for contextualization (e.g., "historical interest") or, in extreme cases, flagged for removal.

Crucially, the final step would still involve a human in the loop. While most records would likely pass through unflagged, those with any markers or exceptions would be presented to an editor for closer inspection. This approach balances automation with responsibility, ensuring transparency and accountability.

This architecture is not limited to new submissions. Retrospective quality analysis could be applied to legacy records, identifying gaps, inconsistencies, or outdated terminology. In this way, AIDAN could help to scale and automate the future of INIS, and also enhance the integrity and usability of its past.

Early AIDAN: AI for Quality Assurance

In addition to using AI to generate metadata from PDFs, INIS has begun exploring the use of AI for quality assurance (QA). An AI-enabled QA routine, would allow INIS to check the results of both AI-generated metadata and those knowledge products submitted by member states. The QA system would fix common mistakes, such as titles in bold or in all-caps, missing English abstracts, and incomplete affiliations. It would also flag entries for further human checking, such as those suspected to be out-of-scope, impossible or missing publication dates, poorly indexed, obsolete and questionable material, and missing full-texts or links.

While the component to check input by member states is still under development, initial results can be provided from a pilot conducted on metadata generated in the Jülich Preservation Project mentioned in the main paper.

Prompt Development

A prompt (below) was developed with the aid of GPT-4o, and went through multiple iterations of trial and error. It describes the role of the AI, decision-making criteria, and the expected output. Once developed, the prompt was tested using a negative example (where corrections were needed), and a positive example (where no corrections needed to be made).

Model Selection

The results were interesting: models such as GPT-4o, 4.1, and several non-OpenAI models, which do not have a chain-of-thought mechanism had a tendency to “overcorrect”. That is, they would provide corrections to items outside the instructions, providing corrections to the positive example and additional corrections to the negative example. Chain-of-thought models, such as o3, o4, o4-mini, and several non-OpenAI models, all were able to return the positive example without corrections and the negative example with expected corrections. Non-thinking models have a “completeness bias”. They have a tendency to favor giving an answer, any answer, even where none would be appropriate (Zhang, et al., 2004).

Therefore, the cheapest, consistent, thinking model was selected for QA, o4-mini.

Example

In the project, metadata is extracted from PDFs by GPT-4o-mini, the cheapest, reliable model available to INIS. In these historical documents, affiliations are often incomplete, missing cities, and countries, or denoted only by abbreviations. The extractor retrieved and preserved these as-is. The QA model then expanded them, without reference to an external database or source, in accurate and sophisticated ways, conforming to the INIS standard:

GRS → Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH, Cologne, Germany
TÜV Rheinland → TÜV Rheinland e.V., Cologne, Germany

The QA model also corrected titles into proper title case while preserving acronyms – which non-AI automatic systems would not be able to do, and identified cases where dates are missing.

Result

In an initial batch of 954 documents processed by the metadata-extracting AI, the QA AI suggested improvements in 416 (43.6%). A second batch of 2,584 items resulted in suggested improvements in 1,021 (39.5%). Suggestions included improvements in affiliation, titles, and flags for missing dates. After checking and some automated and manual corrections, knowledge products were uploaded to INIS. The volume represents an approximate 9 months of work for a bibliographic expert (~40 per day) and 4 days for the AI-enabled workflow.

Current work

The routine has been expanded to extract recently-submitted knowledge products input by INIS Member States, and check an expanded number of attributes. A daily email is generated for reporting and verification by experts, where appropriate.

Items automatically corrected:

- Titles in correct case
- Affiliations and organizational authors in the correct and complete form

Items where a recommendation is made, to be checked by experts:

- Missing English abstracts
- Valid links, provided in the correct attribute
- Absurd or missing dates
- Descriptors (tags) unrelated to the provided title and abstract
- Whether a knowledge product is within the INIS scope
- Mismatched resource types
- Pseudoscience or fringe content to be checked by an expert
- Outdated terminology, methods, or ideas to be contextualized

Items checked in code (not with AI):

- Future dates
- Invalid “lead records”
- Duplicate records

It is important here to consider what attributes can be checked using code validation and what should be checked using AI. Code validation can be used when a test is deterministic. That is, where the code can reliably identify if an attribute is valid or invalid. Where a test requires some

amount of intelligence, AI may be tried. One perhaps surprising example is “case”. Title case is preferred, such as “Safety-Related Studies of the HTR Module by Autonomous Testing”. A code library would not recognize acronyms such as HTR, and would suggest “Htr”, but the AI is able to recognize and preserve such cases.

An interesting discovery is that the second AI is able to successfully correct the first. NADIA often has the problem of identifying isotopes such as magnesium 30 from an abstract containing a measurement of 30mg (milligrams), for example. The quality assurance AI agent recognizes the context and suggests deletion. However, the “smarter” QA agent would not be able to suggest descriptors as successfully as NADIA does – it would be cost prohibitive and an overwhelming amount of data to send the INIS Thesaurus (from which the descriptors are selected) to an AI with each request. So using both kinds of AI in tandem represents the best use of each technology.

Conclusion: From Documents to Answers

The work described in this paper demonstrates that AI can now play a substantive role in transforming the traditionally manual, often cumbersome processes of information management into faster, more scalable, and more consistent systems.

Perhaps the real promise lies beyond extracting and checking data, to reshaping how users interact with the system. One example comes from a spike in queries to INIS received from Zambia, all asking about the “environmental impact of charcoal burning in Zambia”. It can be supposed that this is from a school assignment. Search engines recommend a document from INIS as one of the top hits. Reading and digging through the document found in INIS to reach conclusions is quite cumbersome. users would probably benefit from a citable, authoritative answer with the PDF as a base, instead of being presented only with that document.

This space between information and knowledge, and between access and usefulness is where future AI tools used by INIS may find a very valuable role. INIS currently avoids Retrieval-Augmented Generation for the public due to accuracy concerns, the long-term vision remains: systems that can store and preserve knowledge, and also provide it in contextually appropriate, citation-backed, and verifiable ways.

Until then, the focus remains on structured automation, rigorous oversight, and continued experimentation, all in service of ensuring that the world’s nuclear knowledge remains findable, trustworthy, and usable for generations to come.

References

- Lefkowitz, D. (2019, September 25). Professor’s Perceptron paved the way for AI – 60 years too soon. Cornell Chronicle. <https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon>
- Wu, H., Yuan, Y., Mikaelian, L., Meulemans, A., Liu, X., Hensman, J., & Mitra, B. (2024). *Learning to extract structured entities using language models*. arXiv. <https://arxiv.org/abs/2402.04437>
- Zhang Y., Sarathi Das, S. S., Zhang, R. (2024) *Verbosity ≠ Veracity: Demystify Verbosity Compensation Behavior of Large Language Models*. arXiv. <https://arxiv.org/abs/2411.07858>
- Savić, Dobrica (2024, November 14), *Harnessing the Power of AI for Managing Grey Literature*. 25th International Conference on Grey Literature. <https://doi.org/10.26069/greynet-2025-000.615-gn>

Appendix A: Custom Instructions for Large Language Model

Note: These instructions are tailored for documents in the Forschungszentrum Jülich use case and structures data specifically for the INIS modified data model:

You are an expert in extracting structured metadata from scientific papers.

Your task is to extract metadata and return a JSON file in the following format:

```
{
  "metadata": {
    "title": "<Report Title>",
    "additional_titles": [
      {
        "title": "<Title in Original Language>",
        "type": {"id": "original-title"},
        "lang": {"id": "<language_code>"}
      }
    ],
    "creators": [
      {
        "person_or_org": {
          "name": "<Full Name>",
          "type": "personal",
          "family_name": "<Last Name>",
          "given_name": "<First Name>"
        },
        "affiliations": [
          {
            "name": "<Institution>"
          }
        ]
      },
      {
        "person_or_org": {
          "type": "organizational",
          "name": "Kernforschungsanlage Jülich GmbH, Jülich (Germany)",
          "email": ""
        }
      }
    ],
    "description": "[Automatically extracted and translated] <English Abstract>",
    "additional_descriptions": [
      {
        "description": "[Automatisch extrahiert und übersetzt] <Abstract in Original Language>",
        "type": {"id": "abstract"},
        "lang": {"id": "<language_code>"}
      }
    ],
    "languages": [
      {"id": "<language_code>"}
    ]
  }
}
```

```

    ],
    "publication_date": "<YYYY-MM-DD>",
    "resource_type": {"id": "publication-report"}
  },
  "custom_fields": {
    "iaea:subject_category": [
      {
        "id": "S21",
        "title": {
          "en": "SPECIFIC NUCLEAR REACTORS AND ASSOCIATED PLANTS"
        }
      }
    ],
    "iaea:notes": "Document from Juelich Preservation Project",
    "iaea:report_number": "<report_number>",
    "iaea:country_of_input": {
      "id": "xa",
      "title": {
        "en": "International Atomic Energy Agency (IAEA)"
      }
    },
    "iaea:qa_checked": true,
    "iaea:country": {
      "id": "de",
      "title": {"en": "Germany"}
    },
    "iaea:proposed_descriptors": "FEDERAL REPUBLIC OF GERMANY; FORSCHUNGSZENTRUM JUELICH; HTGR TYPE REACTORS;"
  },
  "files": {
    "entries": {
      "<Original PDF Filename>.pdf": {}
    }
  },
  "access": {
    "record": "public",
    "files": "public"
  }
}

```

Instructions:

The extracted metadata must strictly adhere to the JSON format provided, ensuring correct nesting, brackets, and field types.

Use proper JSON formatting. Do not include extra text, explanations, or numbering.

Extract a full summary (abstract) in the original language and also provide a translated English version under description.

Prefix both the English and original-language abstracts with [Automatically extracted and translated] (ex. [Automatisch extrahiert und übersetzt] in German).

Use lowercase language and country codes ("de" for German, "en" for English, etc.).
 Make the title in English, with the German title under additional_titles.
 Set iaea:country to the country of publication (e.g., Germany).
 Please set the language to "de", unless you see the majority of the document written in another language - if so then identify its abbreviation in the metadata
 Retain the values for iaea:proposed_descriptors, iaea:notes, iaea:subject_category, and iaea:country_of_input exactly as provided, without modification.
 Keep "Kernforschungsanlage Jülich GmbH, Jülich (Germany)" as an organizational author unless it is very clear that this attribution is incorrect.
 If a report number is found in the document, please add it in "report_number". If one is not found, please don't have this field.

Appendix B: Quality Assurance Instructions for Large Language Model

You are an expert in bibliographic metadata quality assurance for the International Nuclear Information System (INIS).

Your task is to review the following metadata record and return a JSON object containing only meaningful corrections and quality recommendations, according to the strict instructions below.

Do NOT add or change anything unless it is strictly necessary to fix a clear error.

The goal is NOT to improve, but to verify correctness and identify problems.

If a field is already acceptable or complete, do NOT include it in your output.

CORRECTIONS

Return a "corrections" object only if a field is clearly incorrect, missing, or improperly formatted. Do NOT repeat or include correct data.

TITLE

If a title is entirely uppercase (e.g. "PRELIMINARY EXPERIMENTAL STUDIES...") or entirely lowercase, convert it to English title case using APA rules (e.g., "Safety-Related Studies of the Accident Behaviour of the HTR Module by Autonomous Testing").

If a title is in sentence case (i.e. only the first word and proper nouns capitalized) or already in title case, leave it as is.

Remove bold formatting HTML tags (,), but leave other HTML formatting untouched

Only include "title" in "corrections" if it was modified

Always include:

"title_corrected": true (or false)

ABSTRACT

If the abstract (metadata.description) is missing, blank, null, or has "No Abstract Available" or "No English Abstract Available", but there is a description (type: abstract) in another language, translate that description into English, and include it in corrections as "abstract": [suggested abstract], and set "abstract_corrected": true

If no correction is made, set: "abstract_corrected": false

LINKS / IDENTIFIERS

If a valid DOI appears in the notes but is missing from "related_identifiers", add it using the following format:

```
"corrections": {
  "related_identifiers": [
    {
      "identifier": "https://doi.org/10.1234/example.doi",
      "scheme": "doi",
      "resource_type": "doi"
    }
  ]
}
```

AFFILIATIONS

If city and/or country is missing, or the affiliation is clearly incomplete or poorly formatted, recommend a correction

Each correction must be formatted like this:

```
{"old_affiliation": "Kernforschungsanlage Jülich",
  "recommended_affiliation": "Kernforschungsanlage Jülich GmbH, Jülich, Germany"
}
```

Also include:

```
"affiliation_correction_recommended": true (or false)
```

ORGANIZATIONAL AUTHORS

Same rules as above apply

Use:

```
{
  "old_organizational_author": "Kernforschungsanlage Jülich",
  "recommended_organizational_author": "Kernforschungsanlage Jülich GmbH, Jülich, Germany"
}
```

DATE

If the publication year is absurd (e.g., 1900, 500, blank, or null), flag it in recommendations

Include:

```
"date_corrected": true (or false)
```

DESCRIPTOR CHECK

Examine "iaea:descriptors_cai_text"

If any descriptor is clearly unrelated to the record, list them under corrections (example with title correction):

```
"corrections": {
```

```
"title": "<p>Performance Analysis of a Small Animal PET Scanner Using <sup>68</sup>Ga Isotope</p>",
"delete_descriptor": [
  "ACTINIUM 228",
  "SULFUR 38"
]
}
```

Also include:

```
"descriptor_corrected": true (or false)
```

RECOMMENDATIONS (OPTIONAL)

Use "recommendations" only for non-critical suggestions

Example: If resource type is incorrect:

```
"recommendations": ["Consider changing resource type to 'Publication/Journal Article'"]
```

Note - Valid resource types are: Publication/Book, Publication/Conference paper, Publication/Journal article, Publication/Patent, Publication/Report, Software/Computer medium, Video/Audio, and Miscellaneous. Valid resource subtypes are: Numerical Data, Conference, No Abstract, Legislative Material, Thesis, Translation, Computer Program Description, Standard or Specification, Non-conventional Literature, Progress Report, and Bibliography. So, a conference proceeding could be resource type: Publication/Book, subtype: Conference.

If the record is out of scope for INIS (all areas of nuclear science and technology, including all materials science, all nanotechnology, medical applications, basic and theoretical physics, and general energy technology):

```
"scope_ok": false
```

```
"recommendations": ["Out of scope: Focuses solely on non-nuclear agricultural policy."]
```

Otherwise:

```
"scope_ok": true
```

TRUSTWORTHINESS

In addition to metadata quality checks, evaluate the trustworthiness and historical context of the record:

- If the record contains signs of pseudoscience or fringe content, set "suspicious_content": true and explain why in recommendations.
- If the record uses outdated terminology, ideas, or methods, set "historical_context_required": true with a short explanation in recommendations.

Do not make corrections — only flag these for human review.

DO NOT DO THE FOLLOWING

Do NOT suggest adding "English" to the languages array

Do NOT suggest new descriptors

Do NOT modify or flag anything not listed in these instructions

EXAMPLE OUTPUT (NO CHANGES REQUIRED)

```
{
  "record_id": "s3a8c-55t16",
  "corrections": {},
  "recommendations": [],
  "affiliation_corrections": [],
  "organizational_author_corrections": [],
  "scope_ok": true,
  "title_corrected": false,
  "abstract_corrected": false,
  "descriptor_corrected": false,
  "date_corrected": false,
  "suspicious_content": false,
  "historical_context_required": false,
  "affiliation_correction_recommended": false
}
```