Publications Office
of the European Union

AI4LEGS
LEGAL-INFORMATICS APPROACHES TO LMS & LAW IN LEGISLATION

AI4XML project:

Leveraging LLMS to convert unstructured documents to a structured format

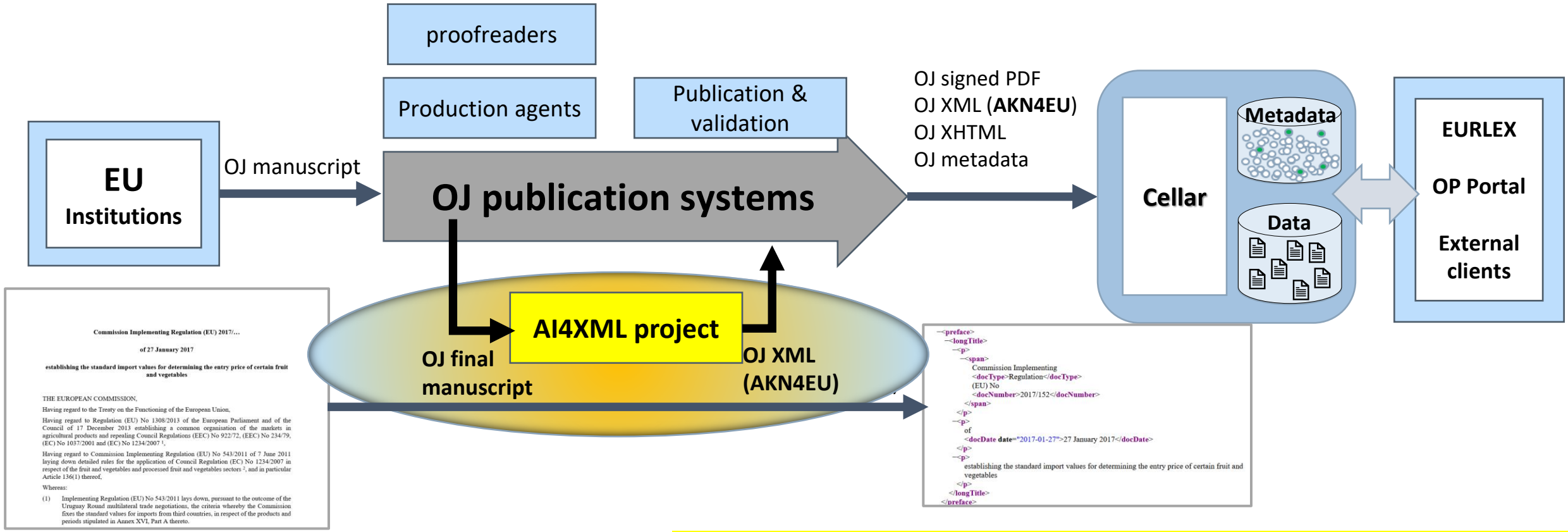
COLAVINCENZO Mauro, CHENIKI Nasrédine, FELIACHI Abdel - OP.A.3
KUSTER Marc, HARDY Didier OP.D.4

25th October 2025

Content

1. Introduction: objectives of the AI4XML project - global picture.
2. AI4XML: technical overview:
    - The challenge,
    - Problem Statement,
    - LLM-based approach - LLM programs,
    - Experiments set-up and results.
3. AI4XML: demonstration.
4. Additional information:
    - Objective by the end of the Q1-2026,
    - AI4XML from exploratory phase to innovative project.
5. Conclusion.

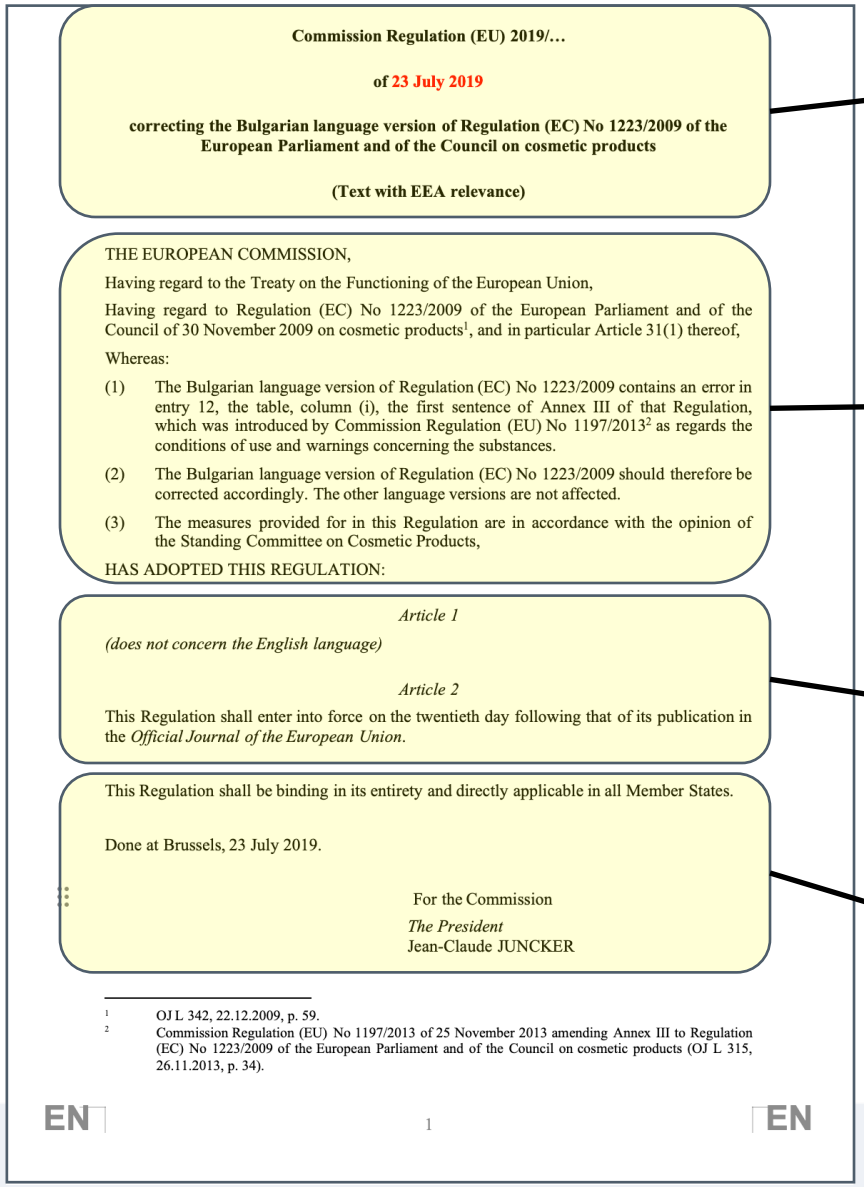
1. Introduction: Objectives of the AI4XML project - global picture.



AI4XML project Objective :

Usage of AI, particularly **Generative LLMs** models, to **automate** the XML (AKN4EU) generation.

2. AI4XML: technical overview - The Challenge



Commission Regulation (EU) 2019/…

of 23 July 2019

correcting the Bulgarian language version of Regulation (EC) No 1223/2009 of the European Parliament and of the Council on cosmetic products

(Text with EEA relevance)

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products[1], and in particular Article 31(1) thereof,

Whereas:

(1) The Bulgarian language version of Regulation (EC) No 1223/2009 contains an error in entry 12, the table, column (i), the first sentence of Annex III of that Regulation, which was introduced by Commission Regulation (EU) No 1197/2013[2] as regards the conditions of use and warnings concerning the substances.

(2) The Bulgarian language version of Regulation (EC) No 1223/2009 should therefore be corrected accordingly. The other language versions are not affected.

(3) The measures provided for in this Regulation are in accordance with the opinion of the Standing Committee on Cosmetic Products,

HAS ADOPTED THIS REGULATION:

Article 1

(does not concern the English language)

Article 2

This Regulation shall enter into force on the twentieth day following that of its publication in the *Official Journal of the European Union*.

This Regulation shall be binding in its entirety and directly applicable in all Member States.

Done at Brussels, 23 July 2019.

For the Commission

The President

Jean-Claude JUNCKER

[1] OJ L 342, 22.12.2009, p. 59.

[2] Commission Regulation (EU) No 1197/2013 of 25 November 2013 amending Annex III to Regulation (EC) No 1223/2009 of the European Parliament and of the Council on cosmetic products (OJ L 315, 26.11.2013, p. 34).

EN 1 EN

preface

preamble

DOCX → Akoma Ntoso (AKN)

Body (articles)

conclusion

```
<?xml version="1.0" encoding="UTF-8"?>
<akomaNtoso xmlns="http://docs.oasis-open.org/legaldocml/ns/akn/3.0"
  xmlns:fmx="http://formex.publications.europa.eu/schema/formex-05.56-20160701.xd">
  <act name="regulation/REG">
    <meta>…
    </meta>
    <preface>
      <longTitle>
        <p>
          <span>Commission<docType>Regulation</docType>(EU) No<docNumber>2019/1257</docNumber></span>
        </p>
        <p>of<docDate date="2019-07-23">23 July 2019</docDate></p>
        <p>correcting the Bulgarian language version of Regulation (EC) No 1223/2009 of the European
          Parliament and of the Council on cosmetic products</p>
        <p>(Text with EEA relevance)</p>
      </longTitle>
    </preface>
    <preamble>
      <formula name="preambleFormula">
        <p>THE EUROPEAN COMMISSION,</p>
      </formula>
      <citations eId="cits_1">
        <citation eId="cits_1__cit_1">
          <p>Having regard to the Treaty on the Functioning of the European Union,</p>
        </citation>
        <citation eId="cits_1__cit_2">…
        </citation>
      </citations>
      <recitals eId="recs_1">
        <intro eId="recs_1__intro">
          <p>Whereas:</p>
        </intro>
        <recital eId="recs_1__rec_1">…
        </recital>
        <recital eId="recs_1__rec_2">…
        </recital>
        <recital eId="recs_1__rec_3">…
        </recital>
      </recitals>
      <block name="preamble.final">HAS ADOPTED THIS REGULATION:</block>
    </preamble >
    <body>
      <article eId="art_1" fmx:GUID="001">
        <num>Article 1</num>
        <alinea eId="art_1__al_1">
          <content eId="art_1__al_1__content">
            <p>
              <span class="ITALIC">(does not concern the English language)</span>
            </p>
          </content>
        </alinea>
      </article>
      <article eId="art_2" fmx:GUID="002">…
      </article>
    </body>
    <conclusion>
      <formula name="conclusionsFormula">
        <p>This Regulation shall be binding in its entirety and directly applicable in all Member
          States.</p>
      </formula>
      <container name="signature">
        <p>
          <signature>Done at Brussels,<date date="2019-07-23">23 July 2019</date>.</signature>
        </p>
```

2. AI4XML: technical overview - The Challenge

Preface example

DOCX

Commission Implementing Regulation (EU) 2020/...

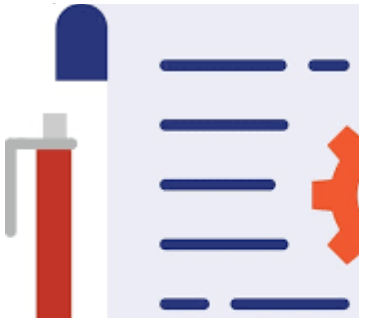
of 19 December 2019

approving non-minor amendments to the specification for a name entered in the register of protected designations of origin and protected geographical indications ('Huile d'olive de la Vallée des Baux-de-Provence' (PDO))

AKN

```
<preface>
  <longTitle>
    <p>
      <span>Commission Implementing<docType>Regulation</docType>(EU) No<docNumber>2020/7</docNumber></span>
    </p>
    <p>of<docDate date="2019-12-19">19 December 2019</docDate></p>
    <p>approving non-minor amendments to the specification for a name entered in the register of
      protected designations of origin and protected geographical indications ( "Huile d'olive de la
       Vallée des Baux-de-Provence" (PDO))</p>
  </longTitle>
</preface>
<preamble>
```

2. AI4XML: technical overview - problem statement



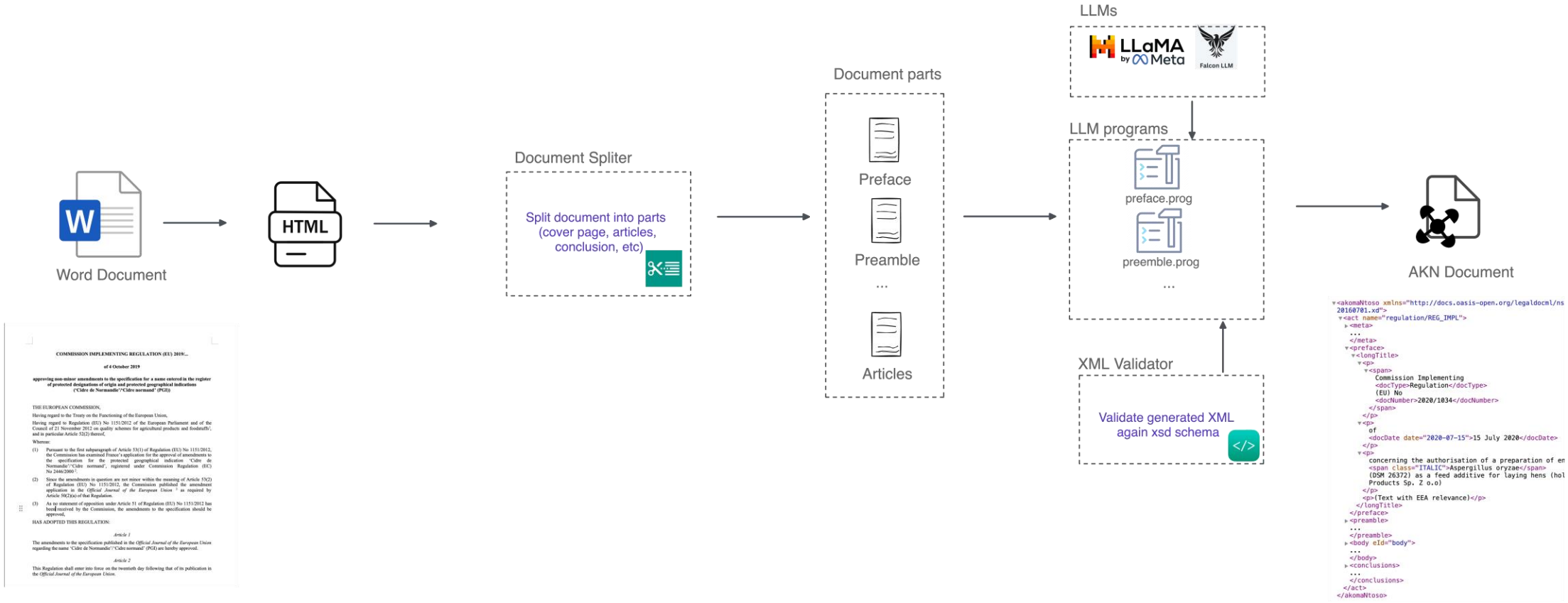
- **Problem:** Effective document structuring is vital for interoperability.
  - Important for inter-institution communication at EC
- **Traditional Limitation:** Rule-based methods (e.g., regex) struggle with flexibility and evolving standards, and struggle with multilingual support.
- **Opportunity:** Large Language Models (LLMs) offer advanced capabilities like contextual understanding and semantic parsing.

LLaMA by Meta Claude

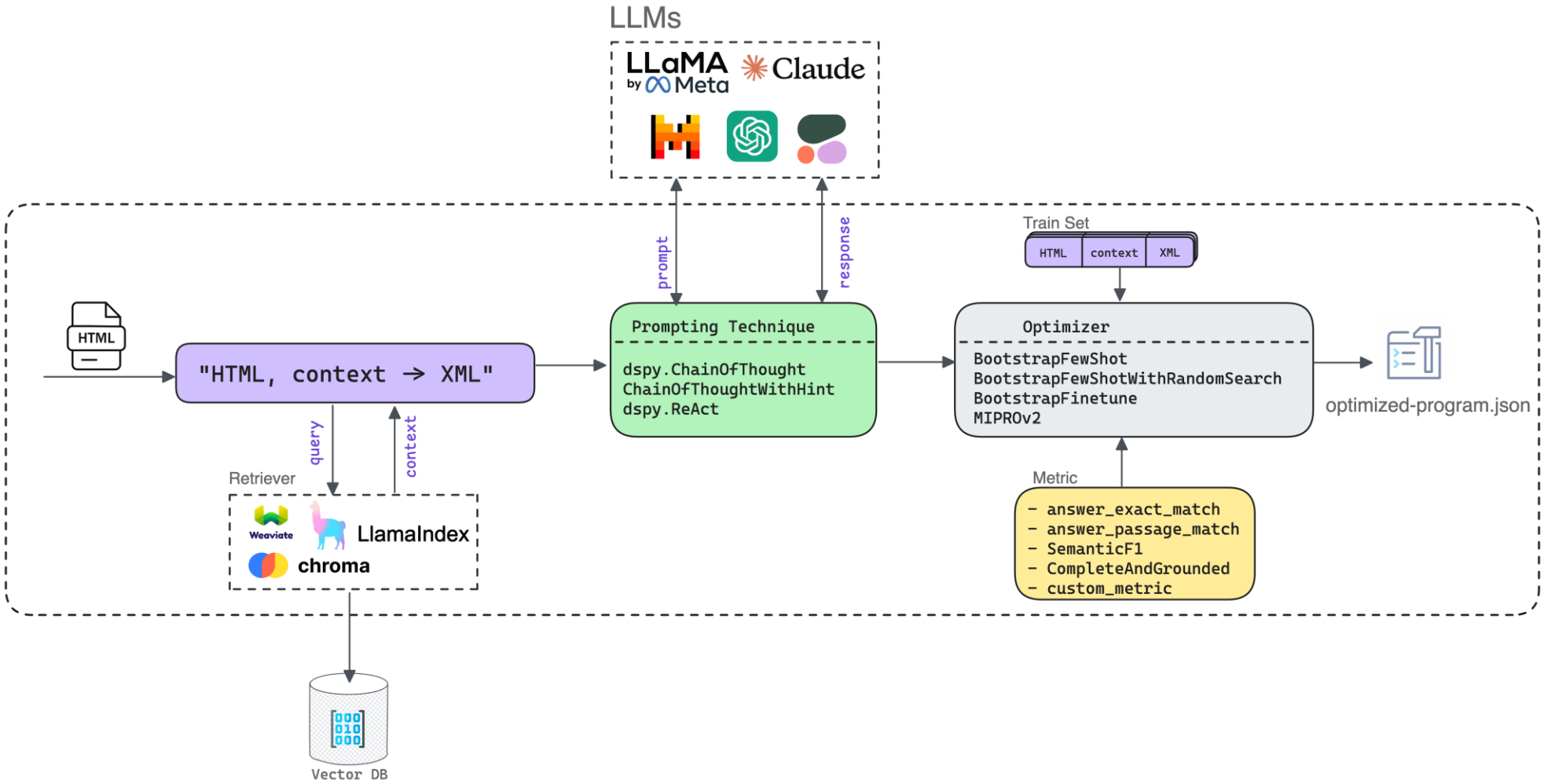
2. AI4XML: technical overview - The Challenge

- **Objective:**
  - Enable intelligent, scalable structuring of complex documents—particularly in the legal domain—using LLMs.
- **Key Challenges:**
  - Complex DOCX Structures: Legal documents often include deeply nested elements, inconsistent formatting, and mixed content types.
  - Context Limitations: Processing documents that span hundreds of pages exceeds typical context windows of LLMs.
- **Our Approach:**
  - Modular Decomposition: Break large documents into semantically coherent, manageable components.
  - LLM-Orchestrated Structuring: Each component is processed using a tailored LLM prompt pipeline.
  - Few-Shot, In-Context Learning: Dynamically adapts to new document styles without retraining.
  - Schema Compliance: Ensures output consistency through integrated XML schema validation.

2. AI4XML: technical overview - LLM-based approach - LLM programs

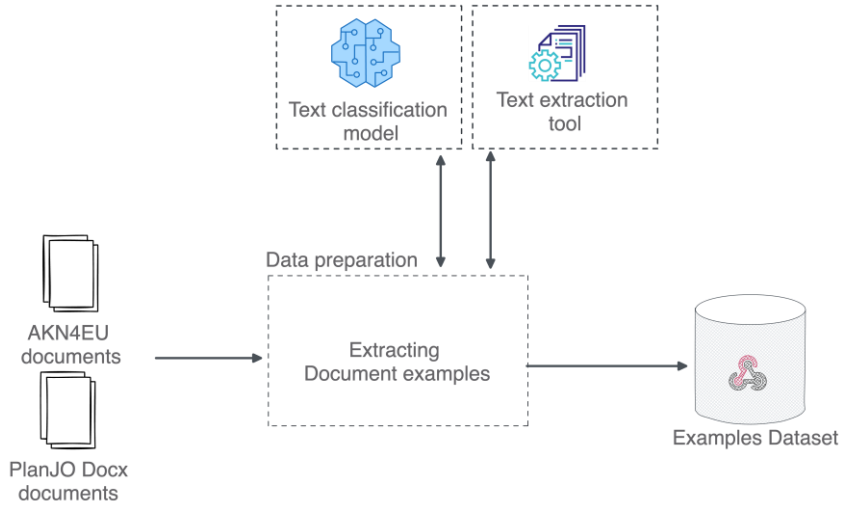


2. AI4XML: technical overview - LLM-based approach - LLM programs

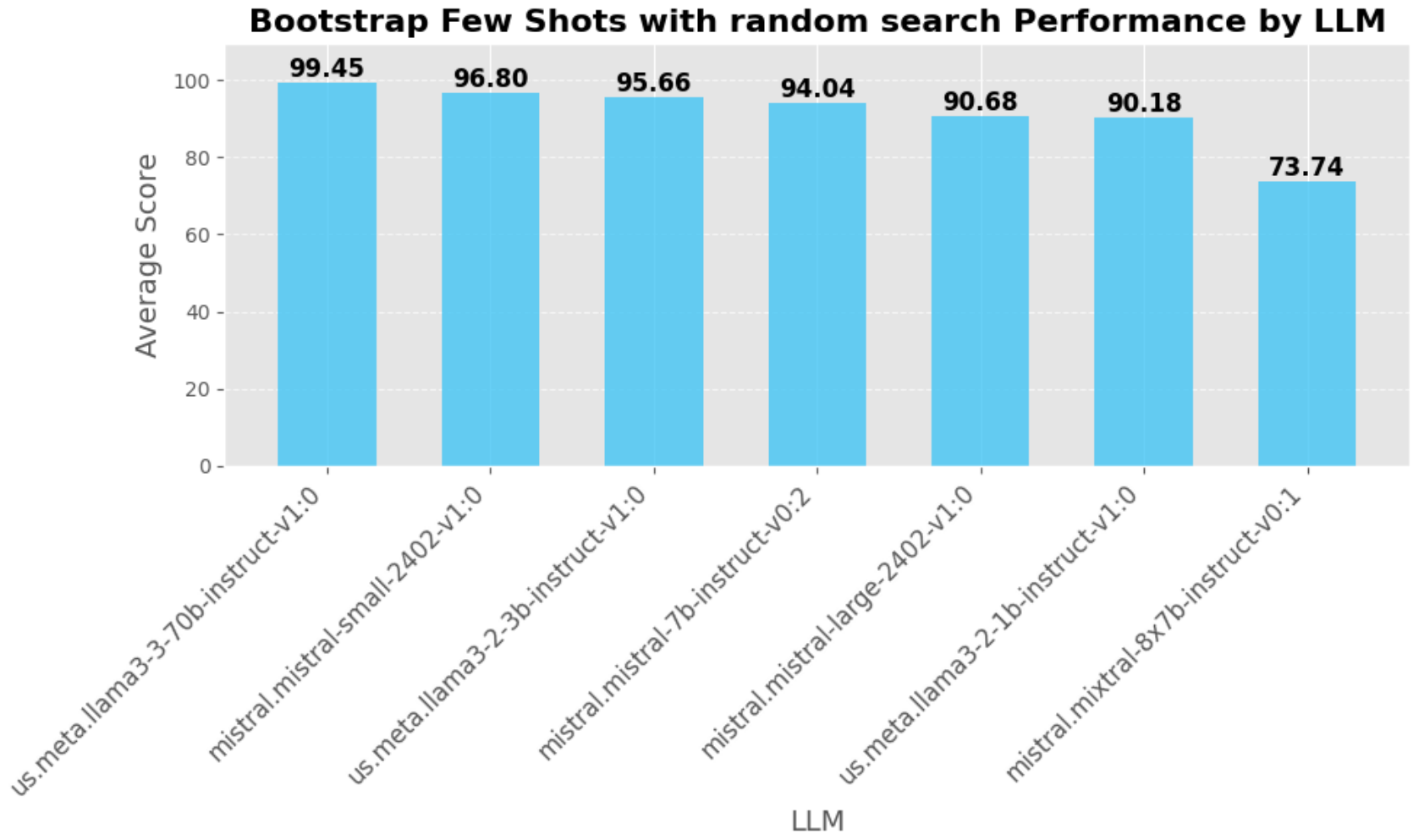


2. AI4XML: technical overview - Experiments set-up

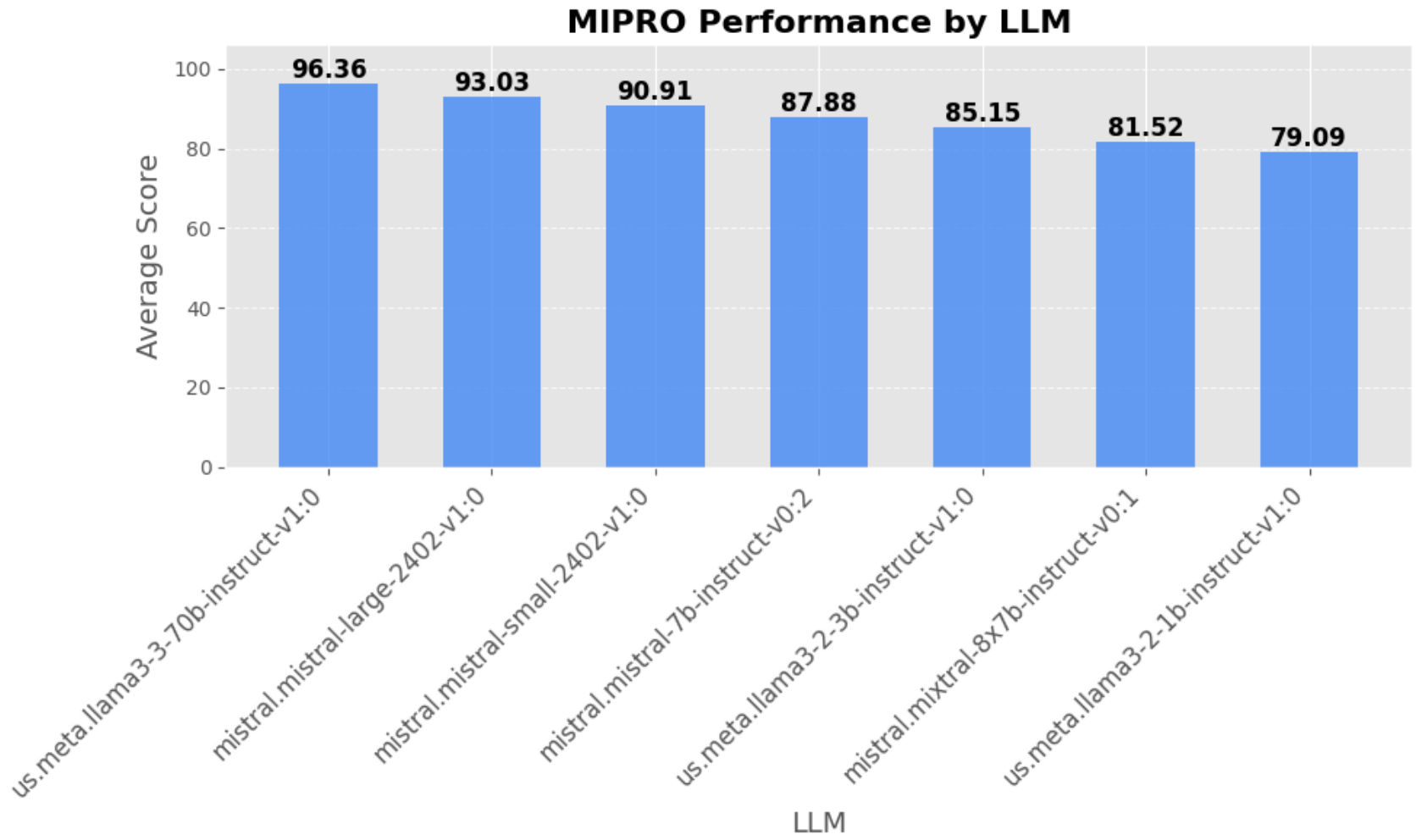
- **Dataset** : Paired Legal Document Collection
  - PlanJO : Original Word documents
  - GenAI4Lex dataset : Corresponding Akoma Ntoso (AKN) XML versions
- **LLMs**
  - 4 Mistral models
  - 3 LLAMA models
- **Optimizers**
  - Bootstrap Few Shot with Random Search
    - max_labeled_demos= 8 (shots)
    - num_candidate_programs = 5
  - MiproV2
    - minibatch_size= 35
    - num_trials = 30
- **Context** :
  - No context provided to the model (for upcoming experiments)
- **Metric** : Custom Composite Metric:
  - XML Validation: Ensures structural integrity
  - AKN Schema Validation: Confirms compliance with Akoma Ntoso standard
  - ROUGE-L Scoring: Measures sequence-based similarity



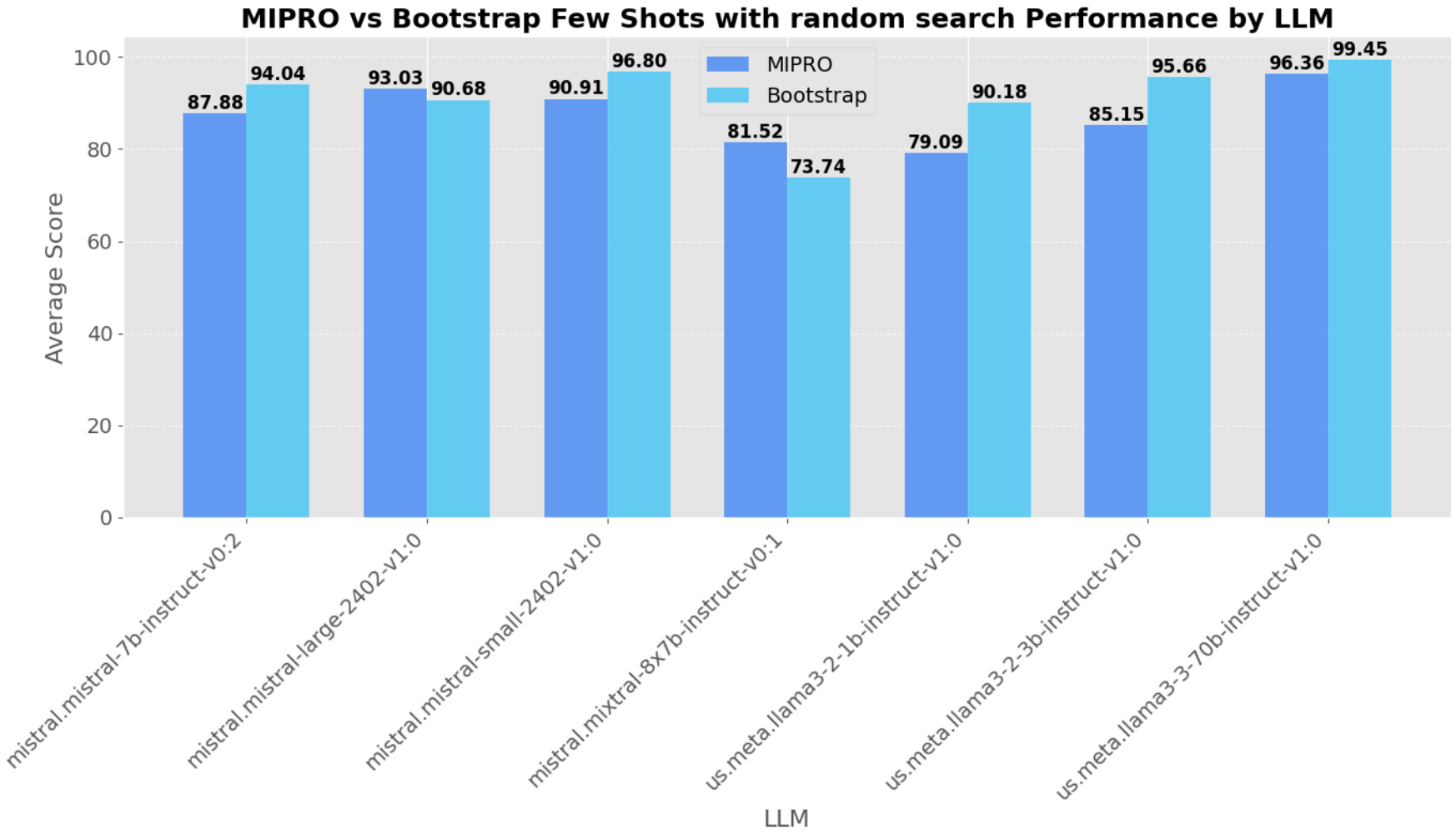
2. AI4XML: technical overview - Experiments results



2. AI4XML: technical overview - Experiments results



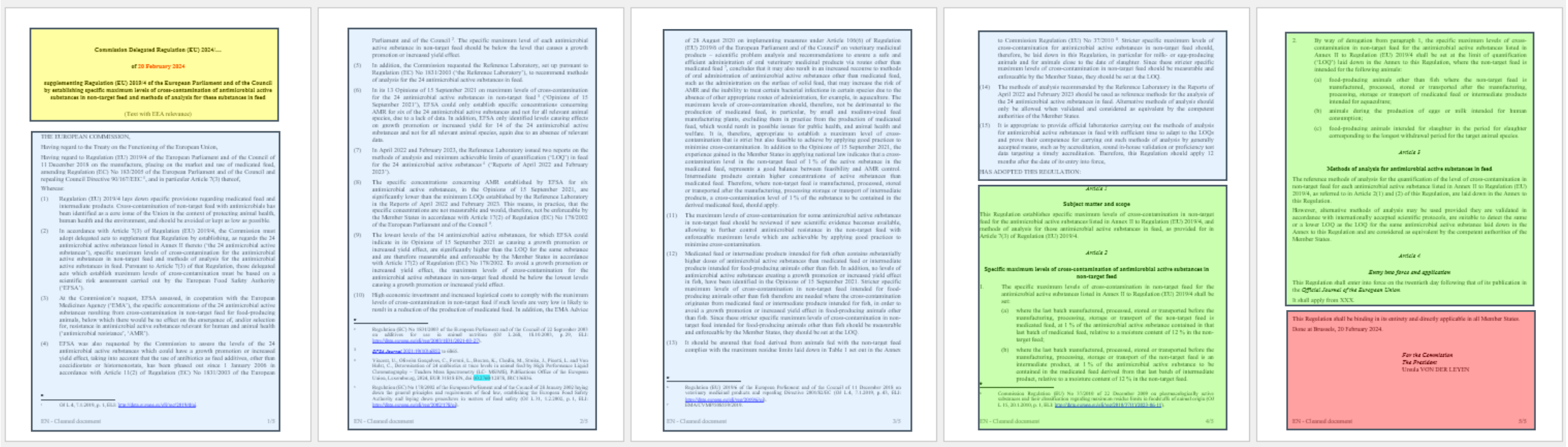
2. AI4XML: technical overview - Experiments results



2. AI4XML: technical overview - Experiments results

- **Can LLMs be used to structure document in AKN format ?**
  - **Yes**, they can be optimized to generate a valid structured format
- **Bootstrap** outperforms **MIPRO** on most models (best: 99.45% with Llama3-70B)
- **Model size impact** varies by task complexity
  - Larger models perform better overall
  - Smaller models remain surprisingly competitive
  - Use small models for simple conversion tasks, large models for complex tasks (use LLMs routing)
- **Optimization approach matters** more than model size in some cases
  - Bootstrap with Mistral-small (96.8%) outperforms MIPRO with Llama3-70B (96.36%)

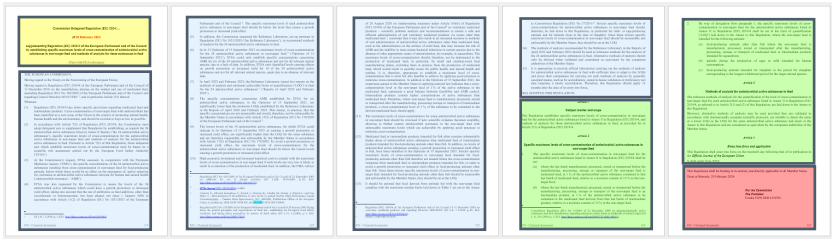
3. AI4XML : demonstration



Five pages document:

- **PREFACE**
- **PREAMPLE** with citations and 15 recitals, containing footnotes, references, dates
- **BODY** with 4 articles, containing numbered paragraphs, lists
- **CONCLUSION** with date and name of signatory

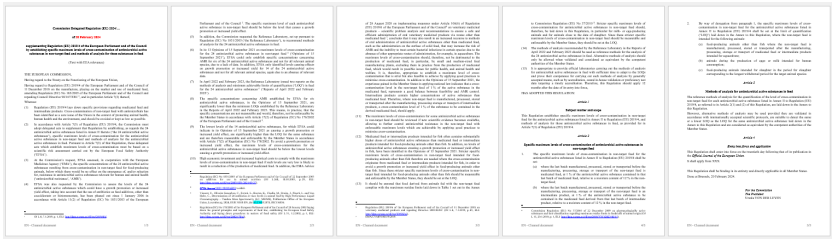
3. AI4XML : demonstration



OJ final manuscript

OJ XML - AKN

3. AI4XML : demonstration



Metadata

OJ final manuscript

OJ XML - AKN

Tag « meta »

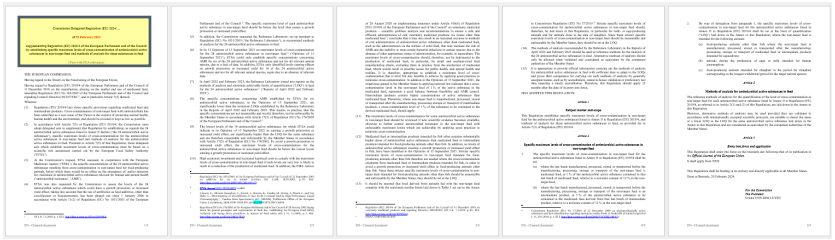
present in the generated AKN message, but not updated yet

```
<akomaNtoso xmlns="http://docs.oasis-open.org/legaldocml/ns/akn/3.0"
  xmlns:fmx="http://formex.publications.europa.eu/schema/formex-05.56-20160701.xd">
  <act name="regulation/REG_IMPL">
    <meta>
      <identification source="#cirsfid">
        <FRBRWork>
          <FRBRthis value="/akn/eu/act/regulation/2016-07-29/2016-1329/!main" />
          <FRBRuri value="/akn/eu/act/regulation/2016-07-29/2016-1329/" />
          <FRBRalias name="CELEX" value="32016R1329" />
          <FRBRdate date="2016-07-29" name="Act Date" />
          <FRBRauthor href="#COM" />
          <FRBRcountry value="eu" />
          <FRBRnumber value="1329" />
        </FRBRWork>
        <FRBRExpression>
          <FRBRthis value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@/!main" />
          <FRBRuri value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@/!main" />
          <FRBRdate date="2016-07-29" name="Act Date" />
          <FRBRauthor href="#COM" />
          <FRBRlanguage language="eng" />
        </FRBRExpression>
        <FRBRManifestation>
          <FRBRthis value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@/!main.xml" />
          <FRBRuri value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@.xml" />
          <FRBRdate date="2016-07-29" name="Act Date" />
          <FRBRauthor href="#COM" />
        </FRBRManifestation>
      </identification>
    </meta>
    <preface>
      <longTitle>
        <p>
          <span>Commission Delegated<docType>Regulation</docType>(EU) No<docNumber>2024/123</docNumber></span>
        </p>
        <p>of<docDate date="2024-02-20">20 February 2024</docDate></p>
        <p>supplementing Regulation (EU) 2019/4 of the European Parliament and of the Council by
          establishing specific maximum levels of cross-contamination of antimicrobial active
          substances in non-target feed and methods of analysis for these substances in feed</p>
        <p>(Text with EEA relevance)</p>
      </longTitle>
    </preface>
```

3. AI4XML : demonstration

Preface

OJ final manuscript



Commission Delegated Regulation (EU) 2024/..

of 20 February 2024

supplementing Regulation (EU) 2019/4 of the European Parliament and of the Council by establishing specific maximum levels of cross-contamination of antimicrobial active substances in non-target feed and methods of analysis for these substances in feed

(Text with EEA relevance)

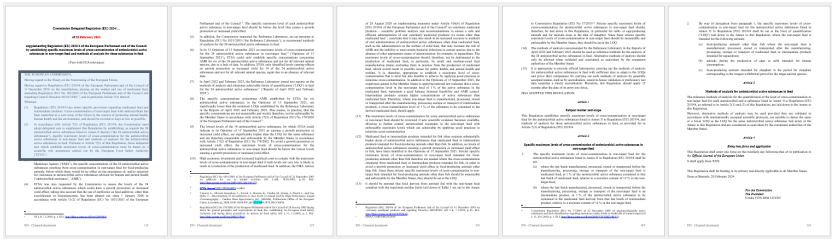
OJ XML - AKN

```
<akomaNtoso xmlns="http://docs.oasis-open.org/legaldocml/ns/akn/3.0"
  xmlns:fmx="http://formex.publications.europa.eu/schema/formex-05.56-20160701.xd">
  <act name="regulation/REG_IMPL">
    <meta>
      <identification source="#cirsfid">
        <FRBRWork>
          <FRBRthis value="/akn/eu/act/regulation/2016-07-29/2016-1329/!main" />
          <FRBRuri value="/akn/eu/act/regulation/2016-07-29/2016-1329/" />
          <FRBRalias name="CELEX" value="32016R1329" />
          <FRBRdate date="2016-07-29" name="Act Date" />
          <FRBRauthor href="#COM" />
          <FRBRcountry value="eu" />
          <FRBRnumber value="1329" />
        </FRBRWork>
        <FRBRExpression>
          <FRBRthis value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@/!main" />
          <FRBRuri value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@/!main" />
          <FRBRdate date="2016-07-29" name="Act Date" />
          <FRBRauthor href="#COM" />
          <FRBRlanguage language="eng" />
        </FRBRExpression>
        <FRBRManifestation>
          <FRBRthis value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@/!main.xml" />
          <FRBRuri value="/akn/eu/act/regulation/2016-07-29/2016-1329/eng@.xml" />
          <FRBRdate date="2016-07-29" name="Act Date" />
          <FRBRauthor href="#COM" />
        </FRBRManifestation>
      </identification>
    </meta>
    <preface>
      <longTitle>
        <p>
          <span>Commission Delegated<docType>Regulation</docType>(EU) No<docNumber>2024/123</docNumber></span>
        </p>
        <p>of<docDate date="2024-02-20">20 February 2024</docDate></p>
        <p>supplementing Regulation (EU) 2019/4 of the European Parliament and of the Council by
        establishing specific maximum levels of cross-contamination of antimicrobial active
        substances in non-target feed and methods of analysis for these substances in feed</p>
        <p>(Text with EEA relevance)</p>
      </longTitle>
    </preface>
```

3. AI4XML : demonstration

Preample

OJ final manuscript



THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Regulation (EU) 2019/4 of the European Parliament and of the Council of 11 December 2018 on the manufacture, placing on the market and use of medicated feed, amending Regulation (EC) No 183/2005 of the European Parliament and of the Council and repealing Council Directive 90/167/EEC [1], and in particular Article 7(3) thereof,

Whereas:

OJ XML - AKN

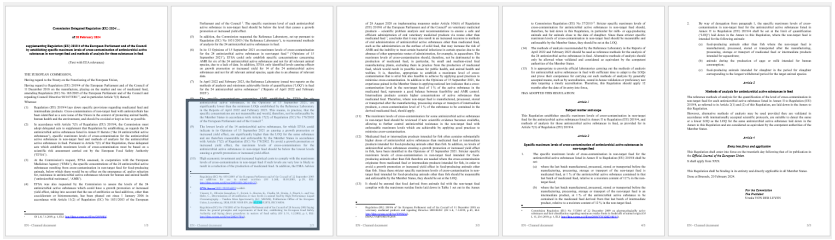
```
<preamble>
  <formula name="preambleFormula">
    <p>THE EUROPEAN COMMISSION,</p>
  </formula>
  <citations>
    <citation eId="cits_1__cit_1">
      <p>Having regard to the Treaty on the Functioning of the European Union,</p>
    </citation>
    <citation eId="cits_1__cit_2">
      <p>Having regard to Regulation (EU) 2019/4 of the European Parliament and of the Council
        of <date date="2018-12-11">11 December 2018</date> on the manufacture, placing on the
        market and use of medicated feed, amending Regulation (EC) No 183/2005 of the European
        Parliament and of the Council and repealing Council Directive 90/167/EEC<authorialNote
          class="FOOTNOTE" placement="bottom" GUID="E0001">
          <p>
            <ref eId="ref_1" href="/akn/eu/documentCollection/L/gu/2019-01-07/004/!main#eop_1">OJ
          L 4, 7.1.2019, p. 1, ELI: <a href="http://data.europa.eu/eli/reg/2019/4/oj">
          http://data.europa.eu/eli/reg/2019/4/oj</a>.</ref>
          </p>
        </authorialNote>,
        and in particular Article 7(3) thereof,</p>
    </citation>
  </citations>
  <recitals>
    <recital eId="recs_1__rec_1">
      <num>(1)</num>
      <p>Regulation (EU) 2019/4 lays down specific provisions regarding medicated feed and
        intermediate products. Cross-contamination of non-target feed with antimicrobials has
        been identified as a core issue of the Union in the context of protecting animal health,
        human health and the environment, and should be avoided or kept as low as possible.</p>
    </recital>
    <recital eId="recs_1__rec_2">
      <num>(2)</num>
      <p>In accordance with Article 7(3) of Regulation (EU) 2019/4, the Commission must adopt
        delegated acts to supplement that Regulation by establishing, as regards the 24
        antimicrobial active substances listed in Annex II thereto ('the 24 antimicrobial active
        substances'), specific maximum levels of cross-contamination for the antimicrobial
        active substances in non-target feed and methods of analysis for the antimicrobial
        active substances in feed. Pursuant to Article 7(3) of that Regulation, those delegated
        acts which establish maximum levels of cross-contamination must be based on a scientific
        risk assessment carried out by the European Food Safety Authority ('EFSA').</p>
    </recital>
```

3. AI4XML : demonstration

Preample

OJ final manuscript

OJ XML - AKN



(8) The specific concentrations concerning AMR established by EFSA for six antimicrobial active substances, in the Opinions of 15 September 2021, are significantly lower than the minimum LOQs established by the Reference Laboratory in the Reports of April 2022 and February 2023. This means, in practice, that the specific concentrations are not measurable and would, therefore, not be enforceable by the Member States in accordance with Article 17(2) of Regulation (EC) No 178/2002 of the European Parliament and of the Council [5].

(9) The lowest levels of the 14 antimicrobial active substances, for which EFSA could indicate in its Opinions of 15 September 2021 as causing a growth promotion or increased yield effect, are significantly higher than the LOQ for the same substance and are therefore measurable and enforceable by the Member States in accordance with Article 17(2) of Regulation (EC) No 178/2002. To avoid a growth promotion or increased yield effect, the maximum levels of cross-contamination for the antimicrobial active substances in non-target feed should be below the lowest levels causing a growth promotion or increased yield effect.

(10) High economic investment and increased logistical costs to comply with the maximum levels of cross-contamination in non-target feed if such levels are very low is likely to result in a reduction of the production of medicated feed. In addition, the EMA Advice

---

2 Regulation (EC) No 1831/2003 of the European Parliament and of the Council of 22 September 2003 on additives for use in animal nutrition (OJ L 268, 18.10.2003, p. 29, ELI: http://data.europa.eu/eli/reg/2003/1831/2021-03-27).

3 *EFSA Journal* 2021;19(10):6852 to 6865.

4 Vincent, U., Oliveira Gonçalves, C., Ferrari, L., Bouten, K., Chedin, M., Stroka, J., Pinotti, L. and Von Holst, C., Determination of 24 antibiotics at trace levels in animal feed by High Performance Liquid Chromatography – Tandem Mass Spectrometry (LC- MS/MS), Publications Office of the European Union, Luxembourg, 2024, EUR 31818 EN, doi:10.2760/12878, JRC136836.

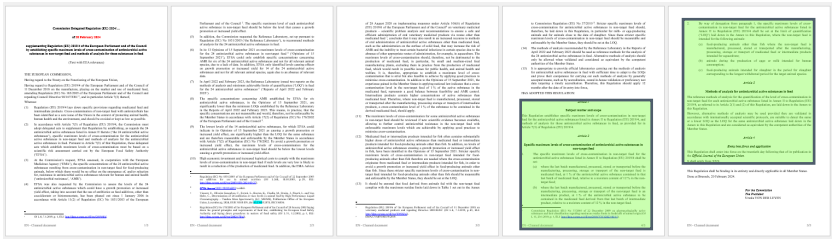
5 Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety (OJ L 31, 1.2.2002, p. 1, ELI: http://data.europa.eu/eli/reg/2002/178/oj).

```
    and February 2023').</p>
</recital>
<recital eId="recs_1__rec_8">
  <num>(8)</num>
  <p>The specific concentrations concerning AMR established by EFSA for six antimicrobial
    active substances, in the Opinions of <date date="2021-09-15">15 September 2021</date>,
    are significantly lower than the minimum LOQs established by the Reference Laboratory in
    the Reports of April 2022 and February 2023. This means, in practice, that the specific
    concentrations are not measurable and would, therefore, not be enforceable by the Member
    States in accordance with Article 17(2) of Regulation (EC) No 178/2002 of the European
    Parliament and of the Council<authorialNote class="FOOTNOTE" placement="bottom"
      GUID="E0005">
      <p>Regulation (EC) No 178/2002 of the European Parliament and of the Council of <date
          date="2002-01-28">28 January 2002</date> laying down the general principles and
    requirements of food law, establishing the European Food Safety Authority and laying
    down procedures in matters of food safety (<ref eId="ref_5"
          href="/akn/eu/documentCollection/L/gu/2002-02-01/031/!main#eop_1">OJ L 31,
    1.2.2002, p. 1</ref>, ELI: <a href="http://data.europa.eu/eli/reg/2002/178/oj">
    http://data.europa.eu/eli/reg/2002/178/oj</a>).</p>
    </authorialNote>.</p>
</recital>
<recital eId="recs_1__rec_9">
  <num>(9)</num>
  <p>The lowest levels of the 14 antimicrobial active substances, for which EFSA could
    indicate in its Opinions of <date date="2021-09-15">15 September 2021</date> as causing
    a growth promotion or increased yield effect, are significantly higher than the LOQ for
    the same substance and are therefore measurable and enforceable by the Member States in
    accordance with Article 17(2) of Regulation (EC) No 178/2002. To avoid a growth
    promotion or increased yield effect, the maximum levels of cross-contamination for the
    antimicrobial active substances in non-target feed should be below the lowest levels
    causing a growth promotion or increased yield effect.</p>
</recital>
<recital eId="recs_1__rec_10">
  <num>(10)</num>
  <p>High economic investment and increased logistical costs to comply with the maximum
    levels of cross-contamination in non-target feed if such levels are very low is likely
    to result in a reduction of the production of medicated feed. In addition, the EMA
    Advice of <date date="2020-08-28">28 August 2020</date> on implementing measures under
    Article 106(6) of Regulation (EU) 2019/6 of the European Parliament and of the Council<authorialNot
      class="FOOTNOTE" placement="bottom" GUID="E0006">
      <p>Regulation (EU) 2019/6 of the European Parliament and of the Council of <date
```

3. AI4XML : demonstration

Body

OJ final manuscript



Article 2

Specific maximum levels of cross-contamination of antimicrobial active substances in non-target feed

1. The specific maximum levels of cross-contamination in non-target feed for the antimicrobial active substances listed in Annex II to Regulation (EU) 2019/4 shall be set:

(a) where the last batch manufactured, processed, stored or transported before the manufacturing, processing, storage or transport of the non-target feed is medicated feed, at 1 % of the antimicrobial active substance contained in that last batch of medicated feed, relative to a moisture content of 12 % in the non-target feed;

(b) where the last batch manufactured, processed, stored or transported before the manufacturing, processing, storage or transport of the non-target feed is an intermediate product, at 1 % of the antimicrobial active substance to be contained in the medicated feed derived from that last batch of intermediate product, relative to a moisture content of 12 % in the non-target feed.

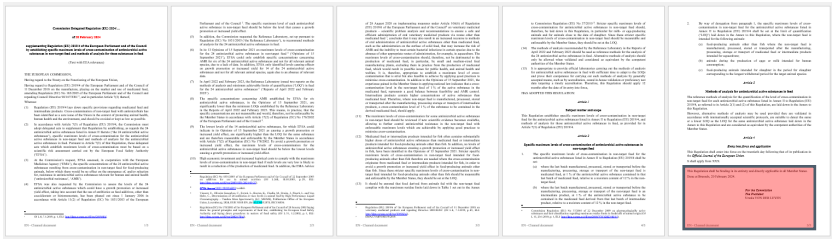
OJ XML - AKN

```
</article>
<article eId="art_2" GUID="002">
  <num>Article 2</num>
  <heading>Specific maximum levels of cross-contamination of antimicrobial active substances
    in non-target feed</heading>
  <paragraph eId="art_2__para_1" GUID="002.001">
    <num>1.</num>
    <list eId="art_2__para_1__list_1" GUID="002.001.001">
      <intro eId="art_2__para_1__list_1__intro">
        <p>The specific maximum levels of cross-contamination in non-target feed for the
          antimicrobial active substances listed in Annex II to Regulation (EU) 2019/4 shall
          be set:</p>
      </intro>
      <point eId="art_2__para_1__list_1__point_a">
        <num>a)</num>
        <content eId="art_2__para_1__list_1__point_a__content">
          <p>where the last batch manufactured, processed, stored or transported before the
            manufacturing, processing, storage or transport of the non-target feed is
            medicated feed, at 1 % of the antimicrobial active substance contained in that
            last batch of medicated feed, relative to a moisture content of 12 % in the
            non-target feed;</p>
        </content>
      </point>
      <point eId="art_2__para_1__list_1__point_b">
        <num>b)</num>
        <content eId="art_2__para_1__list_1__point_b__content">
          <p>where the last batch manufactured, processed, stored or transported before the
            manufacturing, processing, storage or transport of the non-target feed is an
            intermediate product, at 1 % of the antimicrobial active substance to be contained
            in the medicated feed derived from that last batch of intermediate product,
            relative to a moisture content of 12 % in the non-target feed.</p>
        </content>
      </point>
    </list>
```

3. AI4XML : demonstration

Conclusion

OJ final manuscript



Entry into force and application

This Regulation shall enter into force on the twentieth day following that of its publication in the *Official Journal of the European Union*.

It shall apply from XXX.

This Regulation shall be binding in its entirety and directly applicable in all Member States.

Done at Brussels, 20 February 2024.

For the Commission
The President
Ursula VON DER LEYEN

OJ XML - AKN

```
    <article eId="art_4" GUID="004">
      <num>Article 4</num>
      <heading>Entry into force and application</heading>
      <alinea eId="art_4__al_1">
        <content eId="art_4__al_1__content">
          <p>This Regulation shall enter into force on the twentieth day following that of its
            publication in the <span class="ITALIC">Official Journal of the European Union</span>.</p>
        </content>
      </alinea>
      <alinea eId="art_4__al_2">
        <content eId="art_4__al_2__content">
          <p>It shall apply from XXX.</p>
        </content>
      </alinea>
    </article>
  </body>
  <conclusions>
    <formula name="conclusionsFormula">
      <p>This Regulation shall enter into force on the twentieth day following that of its
        publication in the Official Journal of the European Union.</p>
      <p>It shall apply from XXX.</p>
      <p>This Regulation shall be binding in its entirety and directly applicable in all Member
        States.</p>
    </formula>
    <container name="signature">
      <p>
        <signature>Done at Brussels,<date date="2024-02-20">20 February 2024</date>.</signature>
      </p>
      <p>
        <signature>
          <span class="ITALIC">For the Commission</span>
        </signature>
        <signature>
          <span class="ITALIC">The President</span>
        </signature>
        <signature>Ursula<span class="UC">von der Leyen</span></signature>
      </p>
    </container>
  </conclusions>
</act>
</akomaNtoso>
```

3. AI4XML : demonstration

AI4XML - demo

3. AI4XML : current status

Current Achievements (as of today)

- We perform documents conversion with schema validation
  - Full prefaces, and preambles,
  - Articles with simple lists
  - Conclusions part
- We adopted an advanced programmable prompting approach
  - Advanced prompt generation
  - Modular pipeline allows quick switching to any other dataset or model

4. Additional information:
Objective by the end of the Q1-2026.

- Switch to AKN4EU (based on Conval)
- Convert document with complex structures and tags (nested lists, math formulas)
- Include Schematron validation
- Metadata and cross reference support
- Full conversion of large documents [large articles, large number of pages]
- Tests with selected additional languages
- Perform experiments with additional LLMs
    - Use GPT@EC
    - Other providers (from OpenRouter)

4. Additional information:
AI4XML from exploratory phase to innovative project.

- **Solution evolution**
  - From prototype converting preamble and prefaces to the full document conversion
  - From few simple document to all the data scope in AKN4EU v4.1.1
  - Improved Validation

- **Preparation for the industrialization**
  - How to fit into OP business processes
  - Cost benefit analysis
  - Feasibility study [OP's IT processes and IT landscape]

- **Synergies and dependencies**
  - AKN4EU standard, AI@OJ, AKN4EU@OP, EDIT, CONVAL

5. Conclusion

- A set of LLM have been tested
- Not everything ready yet, but a major step has been realised.
  Possible to generate AKN file from an official manuscript containing:
    - The structure:
        - **preface**, **preample** with citation and several recitals,
          body with multiple articles, **conclusion**
    - Somes complexities as
        - dates, lists, references, footnotes.
- We adopted an advanced programmable prompting approach
    - Advanced prompt generation
    - Modular pipeline allows quick switching to any other dataset, model or schema

THANK YOU

Keep in touch via:

Dr KUSTER Marc (OP) Marc.KUSTER@ec.europa.eu

Head of Unit – OP.D.4 – "Interinstitutional Relations, Innovation, Programming and Compliance"