



AI-Supported Educational Mobility Infrastructures in Europe: Opportunities and Limitations Using EMREX as a Reference Architecture

Guido Bacharach¹, Tor Fridell², Geir Vangen³, Janina Mincer-Daszkiewicz⁴, Jan-Joost Norder⁵, Minna Pylkkonen⁶, Igor Drvodelić⁷

¹EMREX Executive Committee, Germany (<https://orcid.org/0000-0002-7945-9118>)

²the Ladok consortium, Sweden

³Sikt – Norwegian Agency for Shared Services in Education and Research, Norway

⁴University of Warsaw, Poland (<https://orcid.org/0000-0003-0672-5291>)

⁵DUO, Dienst Uitvoering Onderwijs, the Netherlands

⁶CSC – IT Center for Science Ltd, Finland

⁷AZVO – Agency for Science and Higher Education, Croatia

guido.bacharach@hs-bochum.de, tor.fridell@liu.se,
geir.vangen@sikt.no, jmd@mimuw.edu.pl, janjoost.norder@duo.nl,
minna.pylkkonen@csc.fi, igor.drvodelic@azvo.hr

Abstract

European educational mobility increasingly relies on interoperable digital infrastructures capable of handling heterogeneous educational data across institutional and national boundaries. While systems such as EMREX have successfully supported standardized data exchange in Higher Education (EMREX Community, n.d.), comparable solutions for Secondary and Vocational Education remain limited due to fragmented standards, resource constraints, and strict data protection requirements.

This paper explores how AI can support educational mobility infrastructures, focusing on selected use cases and exploratory proof-of-concept experiments. Particular attention is given to the feasibility of these approaches under the General Data Protection Regulation (GDPR) and digital sovereignty constraints (European Union, 2016).

The experimental section is based on three exploratory proof-of-concept studies. The results indicate that AI can support productivity and interoperability-related tasks, while significant challenges remain in robustness, scalability, and document interpretation, particularly for handwritten records. The paper concludes by discussing architectural,

legal, and organizational implications and outlines directions for future research toward trustworthy, AI-supported European mobility ecosystems.

1 Introduction

European education systems are undergoing a continuous process of digital transformation, driven by increasing learner mobility, cross-border recognition of qualifications, and the need for interoperable information systems. In Higher Education, infrastructures such as EMREX have demonstrated that standardized digital exchange of student achievement data can significantly reduce administrative effort and improve transparency (EMREX Community, n.d.). However, comparable levels of interoperability and automation are largely absent in Secondary and Vocational Education, where data structures, processes, and governance models are considerably more fragmented.

At the same time, recent advances in artificial intelligence (AI), particularly in natural language processing and semantic data interpretation, have led to renewed interest in their potential application within educational infrastructures. While much of the current discourse on AI in education focuses on learning analytics, personalized tutoring, or content generation, comparatively less attention has been paid to the role AI could play in supporting *educational data interoperability*, *infrastructure scalability*, and *trust-building processes*—especially in mobility scenarios (Amershi, et al., 2019).

This paper explores how AI techniques can support the digital exchange and processing of educational data in the European higher but also Secondary and Vocational Education landscape, using EMREX as a concrete reference architecture. The term “reference architecture” is used here in a pragmatic sense, referring to EMREX as a well-established operational framework rather than a formally defined architectural reference model.

The focus is not on replacing existing governance or decision-making processes, but on augmenting them by increasing productivity, improving data quality, and enabling more scalable handling of heterogeneous educational records. Particular attention is given to AI-supported data transformation between educational standards, the interpretation of semi-structured and unstructured source documents, and the challenges of operating such systems in a GDPR-compliant and sovereignty-preserving manner.

The contributions of this paper are threefold:

1. It systematically identifies AI-supported use cases relevant to educational data exchange and mobility processes in secondary, higher and vocational education and categorizes them based on their potential impact and feasibility.
2. It reports on exploratory experimental work based on three proof-of-concept experiments, e.g. one with GDPR-compliant, open-source AI infrastructures and discusses observed limitations.
3. It critically examines architectural, legal, and organizational constraints, outlining open research and implementation challenges for future European mobility ecosystems.

From an institutional perspective, these AI-supported approaches offer clear operational benefits. They can reduce manual administrative workload, improve data consistency, and provide earlier detection of data quality issues. As such, they contribute directly to making mobility processes more scalable and cost-efficient.

2 Background and Related Work

2.1 EMREX and Educational Data Standards

EMREX is a European infrastructure designed to support the secure and standardized exchange of student achievement data between educational institutions (EMREX Community, n.d.). Its architecture is based on clearly defined roles and interfaces, enabling institutions to request and retrieve verified academic records directly from source systems (EMREX Technical Community, n.d.). At the core of EMREX lies the ELMO (European Learner Mobility) data model, which specifies a structured representation of learning achievements, courses, grades, and related metadata (EMREX Schemas Community, n.d.).

ELMO itself is closely related to the European Learning Model (ELM), which provides a broader conceptual framework for describing learning opportunities, achievements, and credentials across different educational contexts (European Commission, European Learning Model (ELM) Browser, n.d.) (European Commission, European Learning Model: Data model for interoperability of learning opportunities, qualifications and credentials in Europe, n.d.). While ELM aims to cover a wide range of learning scenarios, ELMO represents a more operationalized subset tailored to mobility use cases. In practice, EMREX deployments must also interact with national or regional data standards, such as Germany's XSchule, further increasing the complexity of data integration and transformation. Recent work has also explored interoperability challenges and credential ecosystems in European Higher Education, highlighting the complexity of aligning standards and verification mechanisms across institutional contexts (Gottlieb & Bacharach, 2023).

Although EMREX has been primarily adopted in Higher Education, its architectural principles—decentralized data ownership, standardized exchange formats, and trust-based verification—are also relevant for Secondary and Vocational Education. However, these sectors exhibit significantly higher heterogeneity in document formats, grading systems, institutional processes, and levels of digital maturity, which poses substantial challenges for direct reuse of existing solutions.

2.2 AI in Educational Data Processing: Scope and Delimitation

Research on AI in education has expanded rapidly in recent years, with major strands focusing on adaptive learning systems, automated assessment, predictive analytics, and generative educational content. In contrast, the application of AI to *educational infrastructure services*—such as data transformation, interoperability support, and quality assurance—has received comparatively limited attention.

This paper deliberately narrows its scope to AI applications that support the *technical and organizational processes* underlying educational mobility. The focus is therefore on:

- semantic interpretation of educational documents and datasets,
- transformation between heterogeneous data standards,
- detection of inconsistencies and data quality issues, and
- support for explainability and auditability in data exchange processes.

Excluded from scope are AI-driven pedagogical interventions, learner profiling, or automated decision-making regarding academic recognition. In line with established best practices, the paper assumes that final validation and recognition decisions remain under human and institutional control, with AI acting as an assistive component rather than an autonomous authority (European Commission, Ethics guidelines for trustworthy AI, 2021).

3 Structural Challenges in Secondary and Vocational Education Mobility

Despite the existence of European frameworks and standards, the practical implementation of digital mobility processes in Secondary and Vocational Education faces a number of persistent structural challenges.

First, educational data in this sector is characterized by extreme heterogeneity. Certificates, transcripts, and achievement records vary widely not only between countries, but often between regions, institutions, and even individual schools. Many records still exist primarily in analog or semi-digital forms, such as scanned documents or PDFs without machine-readable structure. Mapping such data to standardized models like ELM or ELMO typically requires extensive manual effort and domain expertise.

Second, many institutions involved in Secondary and Vocational Education operate under severe resource constraints. Unlike large universities, they often lack dedicated IT teams with the capacity to continuously develop, maintain, and adapt complex data transformation pipelines. As a result, even when standards and specifications are available, their implementation remains partial, project-based, or unsustainable in the long term.

Third, mobility-related processes—such as the recognition of prior learning when students change institutions or educational tracks—require a high degree of trust, transparency, and explainability. Decision-makers must be able to understand how specific data elements were interpreted, transformed, and evaluated. Black-box solutions, particularly those involving external cloud-based AI services, are therefore problematic in a European public-sector context governed by strict data protection and accountability requirements (European Union, 2016).

Taken together, these challenges suggest that the primary bottleneck in extending digital mobility infrastructures to Secondary and Vocational Education is not the absence of standards, but the lack of scalable and trustworthy implementation mechanisms. This paper argues that AI-supported tools, if carefully designed and embedded within appropriate governance frameworks, can help address this bottleneck by increasing productivity, improving data quality, and enabling institutions to “do more with less” without compromising legal or ethical constraints.

Empirical studies on the digitalization of recognition and credit transfer processes further emphasize both the complexity of existing procedures and the potential for AI-supported tools to improve efficiency and transparency (Gilch, et al., 2025).

4 AI-Supported Use Cases in the EMREX Context

This section outlines concrete AI-supported use cases that are relevant for extending and operating digital mobility infrastructures in secondary, higher and vocational education. The use cases are organized along the lifecycle of educational data exchange, ranging from system development and data transformation to recognition-related support processes. The emphasis is on assistive AI functions that enhance scalability, consistency, and transparency, while preserving institutional responsibility and human oversight.

4.1 AI-Assisted Development and Conceptual Design

A fundamental challenge for many educational institutions and consortia, including the EMREX community, is the limited availability of specialized personnel capable of designing and maintaining complex interoperability solutions. Developing converters between educational data standards,

implementing evolving specifications, and documenting transformation logic requires expertise that is often scarce outside larger universities or national agencies.

AI-assisted development tools can partially mitigate this challenge by increasing productivity in software engineering and conceptual design tasks. Examples include:

- generating initial code skeletons for data converters between standards such as ELM, ELMO, or national schemas (European Commission, European Learning Model (ELM) Browser, n.d.) (EMREX Technical Community, n.d.),
- assisting in the formulation and documentation of mapping rules,
- supporting the exploration of alternative architectural approaches during early design phases.

In this context, AI does not replace professional software engineering practices but acts as a force multiplier for small teams. Particularly in early prototyping phases, AI-supported tooling can lower the entry barrier for institutions seeking to participate in interoperable mobility infrastructures. However, rigorous validation, testing, and long-term maintenance remain human responsibilities.

4.2 AI-Supported Educational Data Transformation

One of the most direct applications of AI in the EMREX context is the support of data transformation between heterogeneous educational standards. Even when institutions adopt European frameworks such as ELM or ELMO, practical implementations frequently require conversion between multiple schema versions or between European and national standards, such as Germany's XSchule.

Traditional rule-based converters can handle well-defined mappings but tend to be brittle when standards evolve or when source data deviates from expected structures. AI-supported approaches can assist by:

- identifying semantically corresponding elements across different schemas,
- highlighting ambiguous or incomplete mappings,
- proposing draft transformation rules that can be reviewed and refined by domain experts.

Initial prototype converters between ELM and ELMO have demonstrated the technical feasibility of such approaches (DC4EU Consortium, 2025). However, these prototypes typically require further development to reach production-level robustness, including comprehensive validation mechanisms and version management strategies.

A particularly challenging transformation scenario arises when educational records exist only in analog or semi-structured digital formats, such as scanned certificates or unstructured PDF documents. This situation is common especially in Secondary and Vocational Education, where historical records and legacy processes remain prevalent.

Experimental workflows combining optical character recognition (OCR) with AI-based semantic interpretation have shown promising results in initial proof-of-concept experiments within a ChatGPT environment. While classical OCR approaches such as Tesseract (Smith, 2007) provide a baseline for text extraction, recent AI-based methods extend these capabilities significantly, particularly for complex and semi-structured documents. In such workflows, extracted text is first normalized and then interpreted in context, allowing relevant information—such as subjects, grades, completion dates, and qualification types—to be mapped to structured representations like ELMO 2.0.

While successful experiments have been conducted with certain types of school-leaving certificates, the robustness of these approaches strongly depends on document quality, layout consistency, and language characteristics. As a result, AI-assisted transformation in this domain is best understood as a

semi-automated process, where human validation remains essential, particularly for legally or academically relevant outcomes.

The practical benefit of such transformations lies in reducing manual administrative workload in admission and mobility-related processes where legacy or non-standardized records are still common. While Higher Education institutions increasingly exchange structured digital transcripts, vocational and cross-sector admission workflows often still involve scanned or handwritten certificates that must be interpreted and entered manually. AI-assisted conversion into ELMO could therefore serve as an enabling pre-processing step, supporting faster digitization, improved consistency, and better integration of such records into existing mobility infrastructures. Importantly, these outputs would remain subject to institutional validation before any formal use.

4.3 Towards Semi-Automated Recognition Support: A Visionary Perspective

Beyond data exchange and transformation, AI has the potential to support recognition-related processes when students change institutions, educational tracks, or countries. In a visionary scenario, AI-supported tools could assist in:

- comparing learning outcomes across institutions,
- identifying potentially equivalent qualifications,
- flagging cases that require manual review due to ambiguity or missing information.

Within such a scenario, EMREX could evolve from a pure data exchange mechanism into a more comprehensive mobility support infrastructure. However, this vision raises significant technical, legal, and organizational questions. Automated recognition decisions are neither feasible nor desirable under current European governance frameworks. Instead, AI-supported recognition should be understood as a decision-support function embedded within clearly defined human-in-the-loop architectures.

The feasibility of such approaches depends not only on technical advances, but also on institutional acceptance, explainability requirements, and legal constraints. Consequently, this paper treats semi-automated recognition support as a long-term research direction rather than an immediately deployable solution.

A realistic pathway towards semi-automated recognition support would be the combination of structured mobility data (e.g., ELMO records exchanged via EMREX) with institutionally curated knowledge bases derived from study regulations, module catalogues, and accreditation documents. In such a setting, a large language model could support recognition officers by retrieving relevant rules, highlighting comparable learning outcomes, and generating explainable similarity assessments. Importantly, the AI component would remain strictly advisory, while final recognition decisions stay with authorized institutional bodies.

5 Cross-Cutting AI-Supported Infrastructure Functions

Beyond individual use cases for data transformation or development support, AI techniques can provide a set of cross-cutting infrastructure services that enhance the robustness, transparency, and scalability of educational mobility systems. These functions are not tied to a single phase of the data lifecycle but instead support the overall operation and evolution of infrastructures such as EMREX.

5.1 AI-Supported Data Quality Assessment

Data quality is a critical prerequisite for trustworthy educational data exchange. In mobility scenarios, incomplete, inconsistent, or implausible data can lead to delays, manual rework, or incorrect interpretations. Traditional validation mechanisms typically focus on syntactic correctness and schema compliance but often fail to detect semantic inconsistencies.

AI-supported data quality assessment can complement existing validation approaches by:

- detecting missing or contradictory information across related data elements,
- identifying implausible combinations, such as incompatible grading scales or time spans,
- flagging outliers that deviate significantly from typical patterns observed in comparable datasets.

Importantly, such mechanisms are intended to support, rather than replace, established validation rules. AI-generated assessments should therefore be treated as advisory signals, enabling institutions to prioritize manual reviews where they are most needed (Amershi, et al., 2019).

5.2 Detection of Inconsistencies in Exchanged EMREX Datasets

In distributed infrastructures like EMREX, data is exchanged between autonomous systems that may differ in interpretation, implementation details, or update cycles of underlying standards. As a result, inconsistencies may emerge that are difficult to capture through static validation alone.

AI techniques can assist in identifying such issues by analyzing exchanged datasets across multiple transactions and institutions. Potential applications include:

- detecting recurring mapping errors between specific schema versions,
- identifying systematic omissions of optional but relevant data fields,
- recognizing patterns that indicate misinterpretation of standard definitions.

By providing early warnings of such issues, AI-supported monitoring can help improve overall ecosystem stability without introducing central control mechanisms.

5.3 Schema and Standard Evolution Support

Educational data standards such as ELM and ELMO are subject to ongoing evolution in response to new policy requirements, pedagogical models, and technical insights. For institutions, adapting existing systems and converters to new versions can be costly and error-prone.

AI-assisted analysis can support schema evolution by:

- comparing different versions of standards and identifying structural and semantic changes,
- proposing draft migration strategies for existing datasets,
- highlighting areas where manual intervention is likely to be required due to semantic shifts rather than simple structural modifications.

Such support is particularly valuable in environments where multiple standard versions coexist, as is often the case during transition phases.

5.4 Explainability Layers for Mobility Data Processing

Trust in digital mobility infrastructures depends not only on technical correctness, but also on explainability. Institutions must be able to understand how specific datasets were mapped, transformed, or interpreted—especially in cases involving disputes or audits.

AI-generated explainability layers can contribute by:

- producing human-readable explanations of transformation steps,
- documenting assumptions made during semantic interpretation,
- summarizing differences between source and target representations.

When integrated into EMREX-related workflows, such explanations can improve transparency without requiring users to inspect low-level technical artifacts such as transformation scripts or raw logs.

5.5 Synthetic Test Data Generation

Testing and validating mobility infrastructures is complicated by the sensitivity of real student data. Legal and ethical constraints often limit the availability of realistic datasets for development and quality assurance purposes (European Union, 2016).

AI-supported generation of synthetic mobility datasets offers a promising alternative. Such datasets can be designed to:

- reflect realistic distributions of educational records,
- include edge cases relevant for testing converters and validators,
- avoid the use of personal or identifiable information.

Synthetic data generation is particularly valuable for collaborative testing across institutions, where sharing real data may not be feasible.

5.6 Human-in-the-Loop Architectures

Across all cross-cutting functions, the integration of AI into educational mobility infrastructures must follow a human-in-the-loop paradigm. AI systems can assist in analysis, detection, and proposal generation, but final decisions, especially those with legal or academic consequences, must remain with authorized institutional actors.

Explicit design patterns that define the interaction between AI components and human decision-makers are therefore essential. Such patterns contribute to compliance, acceptance, and long-term sustainability, and are particularly relevant in regulated public-sector environments.

6 Experimental Work: Exploratory Proof-of-Concept Studies under GDPR Constraints

To assess the practical feasibility of AI-supported processes in educational mobility infrastructures under European data protection constraints, a set of exploratory proof-of-concept experiments was conducted. The objective was not to provide a large-scale quantitative evaluation, but rather to generate concrete evidence on what can currently be achieved with both state-of-the-art commercial AI systems and GDPR-compliant open-source deployments.

The experiments focused on three representative scenarios, chosen to cover a spectrum from basic data transformation to complex document interpretation and infrastructure deployment challenges: (1) AI-assisted schema mapping between educational standards, (2) interpretation and conversion of semi-structured educational documents into ELMO formats, and (3) feasibility limits of self-hosted open-source AI infrastructures.

6.1 Case Study 1: Draft Mapping between XSchule and ELM

In a first experiment, to benchmark the capabilities of state-of-the-art commercial AI systems, ChatGPT-5.2 was prompted with excerpts from the German XSchule standard and corresponding elements of the European Learning Model (ELM). The task was to generate an initial mapping proposal that could serve as a starting point for converter development.

The model produced a draft mapping table aligning key data elements such as personal data, qualification types, subject-level grades, and completion dates to plausible ELM entities. Table 1 provides an excerpt of the mapping generated.

Person / Holder (credentialSubject)			
Semantic	XSchule (Abiturzeugnis)	ELM	Remarks
Identifier Person	allgemeineAngaben/schueler /.../schuelerId (falls vorhanden)	credentialSubject.id	If no DID exists: national Scholar-ID or temporarily URI
Type	implicitly „SchuelerAbiturzeugnis“	credentialSubject.type	The ELM schema provides for a fixed type (e.g., person).
Family Name	allgemeineAngaben/schueler /name/familienname	credentialSubject.familyName	Accept directly
Given Name (s)	.../schueler/name/vorname	credentialSubject.givenName	Merge multiple first names if necessary.
Date of Birth	allgemeineAngaben/schueler /geburt/geburtsdatum	credentialSubject.dateOfBirth	ISO-8601 Date.
Place of Birth	.../geburt/geburtsort	credentialSubject.placeOfBirth (if provided for in the profile)	Optional if the ELM profile recognizes the field; otherwise as an additional note.
Nationality or similar	If applicable, from XSchool attributes / code lists (XEducation)	credentialSubject.nationality (if applicable)	Only if specified in the target profile

Table 1 Example excerpt of AI-generated mapping proposal (XSchule → ELM)

The generated mapping covered approximately 15–20 core data elements. Manual inspection indicated that most mappings were semantically plausible, while a smaller subset required refinement due to ambiguity in source definitions and differences in schema granularity. While the mapping appeared promising as an initial engineering aid, it was not further validated in a production environment. The experiment nevertheless illustrates the potential of AI to accelerate early-stage interoperability work by generating structured draft artefacts that can subsequently be reviewed by domain experts.

6.2 Case Study 2: Conversion of a Handwritten Certificate into ELMO

A second experiment addressed one of the most challenging scenarios in Secondary and Vocational Education: the transformation of analog or scanned achievement documents into structured mobility data.

A historical handwritten German school-leaving certificate was processed using ChatGPT-5.2. The system was tasked with (1) interpreting the extracted content semantically and (2) generating corresponding structured output in ELMO 2.0, followed by a simplified conversion into ELMO 1.7.

A simplified excerpt of the generated ELMO-like representation is shown below:

```
...
<learningOpportunitySpecification>
<identifier>
<type>local</type>
<value>de-abitur-allgemeine-hochschulreife</value>
</identifier>
<name xml:lang="de">Allgemeine Hochschulreife (Abitur)</name>
<type>Degree programme</type>
<description xml:lang="de">Zeugnis der allgemeinen Hochschulreife
(Abitur).</description>
```

Manual inspection of the generated output did not reveal obvious semantic inconsistencies, and the resulting ELMO 1.7 representation was considered human-readable and structurally coherent. However, automated schema validation revealed some deviations of the AI-ELMO from the XSD schema. In a simplified validation step, several structural inconsistencies were identified, indicating that while the semantic interpretation was largely coherent, strict schema compliance was not consistently achieved, so the experiment should be understood as a proof of concept rather than a verified conversion pipeline. Figure 1 shows the process.

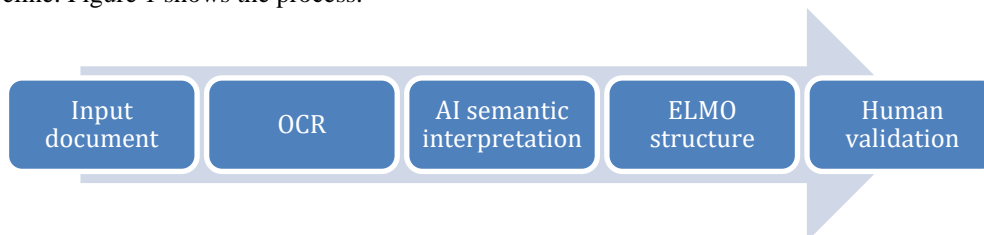


Figure 1: Case Study 2 Process

The experiment highlights the potential of large language models to support semantic interpretation beyond traditional OCR extraction, while also underlining the need for systematic validation mechanisms before operational use.

6.3 Case Study 3: Limits of a GDPR-Compliant Open-Source AI Environment

In order to explore GDPR-compliant alternatives to commercial cloud-based AI systems, GovPart GmbH conducted extensive experiments with a fully self-hosted open-source AI stack. The setup included:

- **Hardware:** A GX44 system equipped with an NVIDIA RTX 4000 SFF Ada Generation GPU.

- **Model Serving and Runtime:** The *vLLM* framework was used for efficient inference of large language models.
- **User Interface and Orchestration:** *Open WebUI* was employed as an interaction and orchestration layer for prompt-based experimentation and workflow testing.
- **Language Model:** *Mistralai/Mistral-7B-Instruct-v0.3*, selected due to its open-source license, relatively modest hardware requirements, and suitability for instruction-following tasks, making it a representative choice for evaluating feasible sovereign deployments within typical institutional constraints.
- **Knowledge Preparation Pipeline:** Educational standards, technical specifications, and domain knowledge were extracted into textual representations using custom Python scripts. The extracted content was enriched with AI-friendly structural headers, chunked with metadata, and indexed to support retrieval-augmented generation (RAG) scenarios.

The system prompt and retrieval setup were iteratively optimized to improve domain adherence and reduce hallucinations, particularly in tasks involving data mapping and schema interpretation.

The system demonstrated basic capabilities in summarization and extraction of short technical passages. However, more demanding interoperability-related tasks, such as generating consistent mapping tables or performing multi-step semantic conversions, were not reliably achievable.

Observed limitations included:

- incomplete or structurally inconsistent outputs in schema-related tasks,
- failure to maintain context across multiple documents,
- inability to robustly interpret noisy OCR input, and
- lack of feasible open-source support for handwritten document recognition.

These results suggest that current 7B-scale open-source models, when deployed on modest hardware, may be insufficient for complex educational mobility workflows requiring high precision and trustworthiness. While basic extraction and summarization tasks could be reproduced consistently, more complex tasks such as schema mapping and multi-step transformations showed limited reliability. In repeated test runs, these tasks frequently resulted in incomplete outputs or structurally inconsistent representations, indicating that the evaluated setup is currently not robust enough for production-level interoperability workflows.

6.4 Summary of Experimental Insights

Taken together, the three proof-of-concept studies provide initial evidence that AI can already support interoperability engineering and document interpretation tasks in educational mobility contexts. At the same time, they confirm substantial limitations regarding validation, robustness, and the performance gap between commercial frontier models and currently feasible sovereign open-source deployments.

Future work should therefore focus on systematic evaluation with larger datasets, integration of schema validation layers, and experimentation with more capable open-source models on shared European infrastructures.

7 Discussion: Privacy, Digital Sovereignty, and Feasibility

The experimental results and identified use cases illustrate both the potential and the current limitations of AI-supported approaches in educational mobility infrastructures. Beyond technical feasibility, their adoption is shaped by legal, organizational, and strategic considerations that are particularly salient in the European public-sector context.

7.1 Data Protection as a Structural Constraint

Educational mobility processes inherently involve personal and often sensitive data, including academic performance, qualifications, and biographical information. Under GDPR, such data must be processed in a manner that ensures confidentiality, purpose limitation, and transparency (European Union, 2016).

While state-of-the-art commercial AI systems demonstrate impressive capabilities in tasks such as document interpretation and semantic reasoning, their use typically relies on cloud-based architectures that pose significant challenges with regard to data protection, jurisdiction, and accountability. For many institutions, especially in Secondary and Vocational Education, these challenges effectively preclude the use of such systems in operational mobility processes.

The experiments described in this paper confirm that GDPR-compliant, self-hosted AI infrastructures are technically feasible but currently involve trade-offs in terms of performance and scalability. This raises the question of whether existing regulatory and organizational frameworks sufficiently support innovation in this domain, or whether new models of compliant AI operation are required. In addition, emerging regulatory frameworks such as the European AI Act introduce further constraints and requirements for AI deployment in public-sector domains, reinforcing the importance of explainability, accountability, and risk management.

7.2 Digital Sovereignty and Open-Source AI

Closely related to data protection is the issue of digital sovereignty. Dependence on external vendors for core infrastructure components may conflict with strategic objectives of European educational systems, particularly where long-term sustainability and control over data and processes are concerned.

Open-source AI models and tools offer a promising alternative by enabling institutions to retain control over their data and deployment environments. However, the experiments reported here (e.g., Case Study 2 on handwritten certificates and Case Study 3 on open-source limits) indicate that open-source solutions currently struggle to match the performance of large proprietary systems, especially in complex tasks such as handwriting recognition or deep semantic interpretation.

This gap suggests that digital sovereignty should not be understood solely as a question of software licensing, but also as an issue of investment, collaboration, and shared infrastructure. Initiatives that pool resources at national or European level may be necessary to make sovereign AI solutions viable for demanding educational use cases.

7.3 Feasibility and Acceptance in Institutional Contexts

Technical feasibility alone does not guarantee institutional acceptance. Educational mobility processes are embedded in legal frameworks, established practices, and trust relationships between institutions. Any AI-supported system must therefore be explainable, auditable, and aligned with existing governance structures.

Human-in-the-loop architectures play a crucial role in this regard. By clearly defining the boundaries between automated assistance and human decision-making, such architectures can help reconcile

efficiency gains with accountability requirements. In practice, this implies that AI outputs should be framed as recommendations or indicators rather than binding decisions.

Acceptance also depends on the perceived added value relative to implementation effort. For institutions with limited resources, AI-supported tools must demonstrably reduce workload or improve quality in order to justify their adoption.

7.4 EMREX as Part of a Broader Mobility Ecosystem

Finally, the discussion raises the question of how AI-supported functions should be embedded within existing mobility infrastructures. EMREX provides a well-established framework for standardized data exchange and trust-based communication between institutions. As such, it represents a natural integration point for AI-assisted services related to data transformation, quality assessment, and explainability.

EMREX is particularly suitable as a reference architecture because it already operationalizes trust-based exchange of achievement data through a lightweight protocol and a well-scoped data model (ELMO). Compared to more comprehensive credential ecosystems, this relative simplicity increases feasibility for incremental AI-supported extensions, such as mapping assistance, quality checks, and explainability services, without redesigning the core mobility infrastructure.

At the same time, EMREX was not originally designed as a comprehensive end-to-end mobility solution. Extending its scope to include AI-supported functions would require careful consideration of architectural boundaries, governance responsibilities, and interoperability with other European initiatives (European Commission, Ethics guidelines for trustworthy AI, 2021). Rather than positioning EMREX as a monolithic solution, it may be more appropriate to view it as a central building block within a modular ecosystem of services that collectively support educational mobility.

8 Conclusion and Outlook

This paper has examined how AI techniques can support digital educational mobility processes in Europe, with a focus on Higher Education as well as Secondary and Vocational Education, using EMREX as a reference infrastructure. The analysis focused on infrastructure-oriented use cases such as data transformation, quality assurance, explainability, and interoperability across heterogeneous educational standards.

The findings indicate that AI can already provide tangible benefits as an assistive technology. In particular, AI-supported tools can increase productivity in the development and maintenance of data converters, support the interpretation of semi-structured educational records, and enhance data quality monitoring in distributed mobility infrastructures. Experimental work with GDPR-compliant, self-hosted open-source AI systems demonstrates that such approaches are technically feasible and suitable for exploratory and support-oriented tasks.

At the same time, significant limitations remain. Current open-source AI solutions struggle with complex semantic interpretation, and performance gaps compared to large commercial systems are substantial. Scaling sovereign AI infrastructures therefore requires considerable investment in hardware, models, and operational expertise.

Beyond technical aspects, successful adoption depends on governance, trust, and institutional acceptance. Human-in-the-loop architectures, explainability mechanisms, and clear responsibility boundaries are essential to ensure compliance with legal frameworks and maintain confidence in mobility-related decisions. Accordingly, AI should be viewed as an enabling component within existing infrastructures rather than an autonomous decision-making system.

Future research should focus on systematic evaluation of larger open-source models, improved OCR pipelines for historical and handwritten documents, and the integration of AI-supported services into

modular European mobility ecosystems. European-level initiatives combining shared infrastructure, open standards, and coordinated governance may be key to balancing digital sovereignty and technical performance.

In this context, EMREX appears well positioned as a foundational building block for AI-supported mobility services, provided that extensions are introduced incrementally and with careful attention to architectural and organizational constraints (EMREX Community, n.d.). The results aim to contribute to an informed discussion within communities such as EUNIS on how AI can be responsibly integrated into future European educational infrastructures.

Many of the concepts outlined remain exploratory. Initial discussions within the EMREX community indicate that AI-supported functions—such as mapping assistance, quality assessment, and explainability—may represent a potential future research direction alongside the continued evolution of the EMREX infrastructure.

9 AI Usage Statement

Portions of this paper were supported by AI-based tools for language refinement and structuring. All technical content, experimental design, and interpretations were developed and validated by the authors.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Founrey, A., Nushi, B., Collisson, P., & Horvitz, E. (2019). Guidelines for human–AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. doi:<https://doi.org/10.1145/3290605.3300233>
- DC4EU Consortium. (2025). *Deployment and Testing Scenarios Results Library (D5.2, Version 3.0, Section: Elements produced/provided by DC4EU – ELMO to ELM converter)*. Digital Credentials for Europe (DC4EU). Retrieved March 30, 2026, from <https://www.dc4eu.eu/reports/>
- EMREX Community. (n.d.). Retrieved January 28, 2026, from EMREX: Supporting student mobility: <https://emrex.eu>
- EMREX Schemas Community. (n.d.). *ELMO schemas*. Retrieved January 28, 2026, from <https://github.com/emrex-eu/elmo-schemas>
- EMREX Technical Community. (n.d.). Retrieved January 28, 2026, from Technical guide to EMREX: <https://emrex.eu/technical/>
- European Commission. (2021). *Ethics guidelines for trustworthy AI*. Publications Office of the European Union. Retrieved January 28, 2026, from <https://digital-strategy.ec.europa.eu/>
- European Commission. (n.d.). *European Learning Model (ELM) Browser*. Retrieved January 28, 2026, from <https://europa.eu/europass/elm-browser/>
- European Commission. (n.d.). *European Learning Model: Data model for interoperability of learning opportunities, qualifications and credentials in Europe*. Retrieved January 28, 2026, from <https://github.com/european-commission-empl/European-Learning-Model/>
- European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. Retrieved January 28, 2026, from <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- Gilch, H., Wannemacher, K., Stratmann, F., Wittke, A., von der Heyde, M., & Hürten, P. (2025). Potential of AI-supported recognition tools for higher mobility and more permeability in the higher education system. *Proceedings of the EUNIS 2024 Annual Congress, 105*, pp. 11-22. Athens. doi:<https://doi.org/10.29007/bvsvx>

- Gottlieb, M., & Bacharach, G. (2023). Cleaning Up: Interplay Between European Standards and Verifiable Credentials for Higher Education Institutions. In M. Auer, W. Pachatz, & T. Rüttemann (Eds.), *Learning in the Age of Digital and Green Transition* (Lecture Notes in Networks and Systems ed., Vol. 634). Cham: Springer. doi:https://doi.org/10.1007/978-3-031-26190-9_81
- Smith, R. (2007). An overview of the Tesseract OCR engine. *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*. doi:<https://doi.org/10.1109/ICDAR.2007.4376991>

Author biographies



Tor Fridell, M. Sc. in Computer Science and Engineering. Currently Head of Student Information System at Linköping University and coordinator for international affairs for the national Swedish Ladok Consortium. Previous jobs include Operations manager for the Ladok Consortium, IS manager for Linköping Institute of Technology, and programmer. Tor has been employed by the university since 1996. Tor has long been involved in international cooperation regarding exchange of student data and is also active in standards work and development of student information systems. Tor has been President of the European Campus Card Association and Chairman of the National Swedish Standards Committee SIS TK450, the national body for CEN TC 353, working with Learning Technologies. Tor has been involved in the EMREX project since the start and is also active in development of student information systems. Tor is currently Chair of the EMREX User Group Executive Committee.



Geir Vangen has more than 20 years' experience in developing nationwide systems within Higher Education. He is head of development at Sikt – the Norwegian Agency for Shared Services in Education and Research. He participates in national and international standardisation work. He has been member of national committees appointed by the Ministry of Education and Research and has led projects on behalf of the Ministry. Geir Vangen graduated from University of Oslo, Institute of Informatics in 1989.



Janina Mincer-Daszkiewicz graduated in computer science from the University of Warsaw, Poland, and obtained a Ph.D. degree in maths from the same university. She is an associate professor in Computer Science at the Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw. Since 1999, she leads a project for the development of a student management information system [USOS](#), which is used in more than 100 Polish Higher Education Institutions. She has been involved in EMREX, Erasmus Without Paper, European Digital Student Service Infrastructure and European Student Card Initiative projects.



Jan Joost Norder works at the Dienst Uitvoering Onderwijs, part of the Dutch Ministry of Education, Culture and Science. In his role as Product Owner he is responsible for the Dutch Diplomaregister and he also has been Chair of the Executive Committee of EMREX. He has many years of experience in improving the digital enrolment process and exchange of student data in Higher Education. Since 2016 he has been involved in international projects.



Minna Pylkkönen works at CSC – IT Center for Science as a Development Manager specializing in the digitalization of higher education. CSC is a company entrusted with special state assignment and owned by the state of Finland and Finnish higher education institutions. She has extensive experience in the higher education sector as well as in various services and digitalization projects related to higher education.



Guido Bacharach is an independent researcher focusing on digital education infrastructures and European student mobility systems. His work is closely connected to interoperability initiatives in the context of EMREX and educational data standards such as ELM and ELMO. His current interests include AI-assisted methods for educational data transformation, data quality assurance, and privacy-preserving architectures under GDPR and digital sovereignty constraints in regulated public-sector environments.



Igor Drvodelić is the Assistant Director of the Agency for Science and Higher Education. He is the head of the Central Admissions Office since its foundation in 2009. He actively promotes the introduction of new services into Croatian Higher Education, such as graduate tracking, recognition of foreign qualifications, student's guidance, electronic data exchange and verifiable credentials.