Systematic Design of Competency-based HPC and AI Education for Transport Systems

Carmel Somers ✉[1] [0000−0002−8088−5640] and
Horacio González–Vélez[2] [0000−0003−0241−6053]

[1] Digital Technology Skills, Dublin, Ireland
https://digitaltechnologyskills.ie/
[2] Cloud Competency Centre, National College of Ireland, Dublin, Ireland
https://ncirl.ie/

Abstract. High Performance Computing (HPC) and Artificial Intelligence (AI) are increasingly shaping transport and mobility systems through digitalisation, modelling, optimisation, and data-intensive decision-making. However, postgraduate provision in these areas often remains anchored in traditional computer science pathways, creating misalignment with the applied and interdisciplinary capabilities now required across transport-sector roles. This paper presents a mixed-methods needs analysis developed within the GreenShift project to identify the capabilities required for advanced digital education in the European transport and mobility sector. The study integrates three complementary methods: LLM-assisted desk research in the EU27, a stakeholder survey with 58 valid responses, and nine semi-structured qualitative interviews with transport and mobility stakeholders. Throughout the desk research process, 156 potentially relevant sources were initially identified and 103 were retained after structured manual validation, resulting in a consolidated evidence base on emerging transport-related roles and associated skills. Rather than treating AI, sustainability, governance, and HPC as isolated specialist domains, the analysis shows that these operate as cross-cutting capability layers embedded across four curriculum-oriented Role Families: Transport Data Analytics; Modelling, Simulation and Digital Twin; Systems Integration and Platform; and Digital Innovation Leadership (Transport-specific). The paper argues that curriculum design for non-computer-science learners should be grounded in validated role-capability patterns and translated into flexible, interdisciplinary programme structures aligned with organisational readiness and labour market demand.

Keywords: Curriculum Design · Skills Needs Analysis · Digital Twins · Sustainability · Role-based Learning · High Performance Computing

1 Introduction

Digitalisation, automation, and data-driven operational decision-making are reshaping the European transport and mobility sector, increasing demand for advanced capabilities in AI, modelling, optimisation, simulation, digital infrastructure, and sustainability-oriented systems thinking [16]. Ideally complemented by HPC, these capabilities are no longer confined to specialist computing roles, but are increasingly embedded in smart cities, transport planning, operations, governance, systems integration, and innovation functions across public and private organisations [3, 27].

This creates an educational challenge for postgraduate programme design. Traditional HPC and AI curricula remain strongly oriented toward computer science and engineering pathways, whereas many emerging transport roles require applied computational understanding without assuming deep prior expertise in low-level parallel programming or algorithm engineering [12, 19, 24].

As a multi-million Euro ground-breaking initiative with 16 academic, industry and government partners located throughout Europe, the GreenShift project has identified the requirement for HPC and AI skills in the European transport industry for the eventual development of a postgraduate higher education curriculum.

This paper outlines GreenShift's *Needs Analysis*, an evidence-based academic account of how role-capability analysis can inform curriculum design for advanced digital education in transport and mobility. As a mixed-methods assessment of current and emerging skills needs within the European transport ecosystem, it provides a detailed account of source selection, inclusion criteria, validation, and coding procedures. It *i*) summarises the transition from analytically derived role families to curriculum-oriented role clusters; *ii*) integrates cross-method synthesis across desk research, survey, and interviews; *iii*) and makes explicit the translation from labour market evidence to capability areas, topic domains, and programme architecture.

1.1 Research question and contribution

The paper is guided by the following overarching research question: *how can curriculum-relevant HPC, AI, and advanced digital capabilities be systematically identified and translated into an academically robust, curriculum-oriented framework for transport and mobility education using mixed methods?*. Four subquestions structure its analysis:

- How can transport-sector roles and associated capability requirements be extracted from heterogeneous public sources in a traceable and methodologically defensible way?
- How can LLM-assisted extraction be incorporated into educational and labour market research without compromising evidential rigour?
- What do the results of the survey and interviews reveal about capability gaps, organisational barriers, and future training needs in the sector?
- How can validated role-capability findings be translated into curriculum architecture suitable for interdisciplinary and non-computer-science learner cohorts?

The paper contributes a fully articulated methodology for role-anchored curriculum design in an emerging digital-green domain. Its contribution lies not only in the identification of transport-relevant digital capabilities, but also in the demonstration of a transparent workflow that links evidence extraction, validation, synthesis, capability clustering, and curriculum design [22].

Beyond the transport domain, the study contributes a transferable methodology for competency discovery and curriculum development in emerging technology-intensive sectors [17]. The combination of bounded LLM-assisted extraction, structured human validation, multi-source triangulation, and curriculum traceability provides a replicable workflow that can be applied wherever role profiles, skills requirements, and curricular structures need to be kept aligned over time, including cybersecurity, energy, manufacturing, and health.

For ETACS and neighbouring cybersecurity education initiatives, the contribution lies in an end-to-end, role-anchored workflow that starts from documented work roles, incorporates competency frameworks such as the ENISA European Cybersecurity Skills Framework (ECSF) [15], and produces curriculum-ready capability structures with an explicit evidence trail for different skills frameworks [4]. In this sense, the methodological approach is particularly relevant to ETACS, because it offers a concrete model to systematically translate evolving cybersecurity workforce requirements into evidence-based educational design, rather than relying solely on technology-centred or expert-only syllabus updates.

2 Literature Review

The study is located within the literature on HPC education, competency-based learning, workforce-informed curriculum design, and just transition. Its central premise is that advanced digital education for transport systems cannot be designed by transferring conventional computer science curricula to a new application area. Instead, curriculum design must account for the roles, organisational settings, and applied decisions through which HPC, AI, modelling, simulation, digital twins, data governance, and sustainability are now being used in transport and mobility.

HPC education has historically been grounded in Parallel and Distributed Computing (PDC), with emphasis on concurrency, parallel architectures, programming models, performance optimisation, and algorithmic scalability [24]. This disciplinary foundation remains necessary for specialist computing education, but it does not fully address the needs of professionals who engage with HPC through applied modelling, simulation, optimisation, digital twins, or AI-enabled decision-support environments. Therefore, recent work on HPC education points towards domain-driven and competency-based learning models, where learners acquire sufficient computational understanding to reason about scale, data movement, model behaviour, infrastructure constraints, and performance implications in applied contexts rather than necessarily becoming specialist parallel programmers [9, 12, 31].

This distinction is important for transport and mobility education. Transport professionals may need to understand when large-scale simulation, GPU-accelerated analytics, data-intensive optimisation, or distributed digital-twin infrastructures are appropriate, reliable, sustainable, and governable. Their educational requirements are therefore different from those of conventional computer science cohorts. The curriculum problem is not simply to transfer existing HPC modules into transport programmes, but to identify which computational concepts should be retained, reframed, scaffolded, and connected to transport-sector decisions. This supports a role-first approach in which HPC and AI are treated as enabling capabilities embedded within operational, analytical, governance, and innovation roles [1, 10].

The literature also indicates that the integration of AI and HPC into non-computing domains requires curriculum structures that combine technical fluency with contextual judgement. Domain professionals must interpret model outputs, assess data quality, understand uncertainty, evaluate sustainability implications, and collaborate across organisational boundaries. For this reason, advanced digital education should avoid a narrow separation between technical modules and contextual modules. Instead, it should organise learning around capability integration: data, modelling, AI, HPC, digital infrastructure, governance, sustainability, and cybersecurity should be presented as mutually dependent elements of digitally transformed systems [14, 19, 20].

Transport systems provide a clear example of this integration problem. Digitalisation, automation, and data-driven decision-making are increasingly embedded in infrastructure management, logistics, public transport, mobility services, urban planning, and policy-oriented functions [16]. In these settings, AI and HPC are not isolated technologies but part of wider socio-technical systems affected by legacy infrastructure, fragmented data environments, interoperability constraints, sustainability pressures, and uneven digital maturity [2, 3, 21]. Green and digital transition agendas also require education to address environmental performance, organisational readiness, and social distribution of capabilities, rather than treating technical development as a narrow software-training problem [23].

2.1 Existing approaches to workforce-informed curriculum design

The alignment of educational provision with evolving workforce requirements has been a longstanding concern within professional and vocational education. Traditional approaches to curriculum design have relied on occupational standards, competency frameworks, expert panels, and structured job analysis methods such as Developing A Curriculum (DACUM) [28] and task-based competency modelling [29]. These approaches provide valuable mechanisms for identifying role requirements, but are often resource-intensive, periodic in nature, and limited in their ability to respond rapidly to emerging technological change. Competency-based education provides a useful foundation because it directs attention to what learners must be able to understand and do in context, but it still requires a defensible process to identify which capabilities matter in a changing domain [22].

More recently, advances in labour market intelligence have enabled the use of skills analytics, occupational databases, job advertisement mining, and sectoral forecasting to identify emerging workforce trends. European frameworks such as the European e-Competence Framework (e-CF), ESCO, DigComp, GreenComp, and the European Cybersecurity Skills Framework (ECSF) provide structured reference models for competency identification and curriculum alignment [7]. These frameworks offer important guidance, but are not designed to identify emerging sector-specific capability requirements in rapidly evolving domains such as AI-enabled transport systems, digital twins, advanced simulation, or HPC-supported optimisation. They also do not automatically explain how a capability should be translated into learning outcomes, assessments, and modular programme structures.

The growing availability of Large Language Models (LLMs) has also led to increased interest in automated approaches to competency extraction and skills intelligence. Recent studies have explored AI-assisted analysis of job advertisements, occupational descriptions, and policy documents to identify skill demands and trends in the workforce. However, concerns remain about transparency, reproducibility, source fidelity, and the risk of generating inferred or unsupported findings. Consequently, AI-assisted approaches require robust validation mechanisms and should complement rather than replace human expertise and judgment [5, 8, 18].

Workforce-informed curriculum design therefore faces a double translation problem. First, heterogeneous evidence from the labour-market must be converted into stable role and capability patterns. Second, these patterns must be converted into curriculum structures that are teachable, assessable, and accessible to heterogeneous cohorts of learners. Existing curriculum design methodologies frequently address only one side of this problem. Workforce studies often identify emerging skills but provide limited guidance on programme structure, while educational frameworks can define learning outcomes without a transparent evidence trail linking them to labour market requirements. This creates challenges in demonstrating curriculum relevance in interdisciplinary fields where technological, organisational, sustainability, and governance considerations intersect.

The proposed methodology seeks to address this gap through an integrated workflow that combines evidence extraction assisted by bounded LLM, structured human validation, stakeholder consultation, framework mapping, and curriculum traceability. Rather than relying on AI-generated interpretation, the approach constrains LLMs to the identification of explicitly evidenced roles and skills from publicly available sources. The extracted outputs are subsequently validated, normalised, triangulated through survey and interview data, and mapped to recognised competency frameworks before informing curriculum design decisions.

The novelty of the approach lies not in the use of LLMs alone, but in the systematic integration of AI-assisted extraction, expert validation, multi-method triangulation, and curriculum translation within a single methodological framework. This provides a transparent and replicable process for transforming diverse evidence from the labour market into curriculum-relevant capability structures. Although developed within the transport and mobility sector, the methodology has potential applicability in other domains facing rapid digital transformation, including cybersecurity, energy, manufacturing, and healthcare.

3 Methodology

Similarly to previous needs analyses to generate competence-based curricula in computing [30], the GreenShift Needs Analysis has adopted an exploratory mixed-methods design to generate an evidence-based and curriculum-oriented understanding of skills needs in European transport and mobility. Three methods were integrated: desk research, survey research, and qualitative interviews, with triangulation used to strengthen reliability and practical relevance.

The desk research systematically examined transport-related roles and associated skills related to AI, HPC, advanced digital technologies, and sustainability in the EU27. A standardised prompt framework was used to identify only explicitly stated roles and associated skills from verifiable public sources, including policy documents, occupational profiles, industry reports, and labour market analyses [5, 8]. The searches were conducted primarily in English, while relevant sources in the national-language were included where they could be reliably interpreted and validated. In the EU27, 156 potentially relevant sources were initially identified and manually screened for accessibility, relevance, and evidential quality. After validation, 103 sources were retained, while 53 were excluded due to inaccessibility, duplication, insufficient role–skill evidence, or lack of verifiability. This explicit inclusion and exclusion logic strengthens the academic robustness of the methodology.

3.1 LLM-assisted extraction and validation

LLMs have been used to support scalable extraction of transport roles and associated skills from public sources, but their use was strictly limited to retrieval and structured extraction rather than interpretation or synthesis. The extraction process was carried out using OpenAI's ChatGPT platform between December 2025 and January 2026 with web search enabled, and outputs were captured in standardised Excel templates containing role title, associated skills, source reference, and country. The prompt is presented in Appendix A.

Several platforms, including ChatGPT, Perplexity Pro, and Microsoft Copilot, were tested for adherence to detailed extraction rules, source fidelity, and Excel-compatible structured output. ChatGPT was retained because it showed the strongest adherence to prompt constraints and the most consistent structured outputs, while systems that produced more inferred or less controllable outputs were excluded from the final methodology [8].

Because transport-sector sources frequently contained evidence relating to more than one capability domain, extraction was conducted separately for each predefined skill category, including AI, HPC, cybersecurity, governance, sustainability, data, and systems integration. Skill categories were treated as controlled analytical lenses rather than mutually exclusive properties of a source. The same source could therefore be processed more than once using category-specific prompts where it contained explicit evidence for multiple capability areas. This design avoided forcing a single dominant label onto multi-capability roles and directly addressed the risk that AI-enabled optimisation, HPC-supported simulation, digital twins, data governance, or cybersecurity governance could otherwise be collapsed into a single technical category.

Each extraction pass used a conservative zero-row rule: if the source did not explicitly evidence both a role and an associated skill within the specified category, no record was created. This rule was particularly important for areas that overlapped, where plausible inference could otherwise be introduced too easily. Following extraction, outputs from different categories were merged, and duplicate role entries were reviewed manually. Cross-domain entries were retained only where the underlying source supported the association. During validation, records were classified as single-category evidence, multi-category evidence, or rejected inference. This allowed overlaps to be analysed as validated cross-cutting capability themes rather than as artefacts of prompt design.

Validation was multi-layered. All extracted entries were manually reviewed by a domain-aware researcher using a structured checklist to verify source accessibility, transport relevance, evidential alignment between roles and skills, and compliance with coding rules. Ambiguous entries were removed rather than rewritten, and a second reviewer examined the methodology, validation approach, and consolidated findings as part of quality assurance.

To support methodological transparency and future replication, the project maintained standardised prompt templates, coding structures, validation checklists, controlled vocabularies, inclusion and exclusion rules, and examples of accepted and rejected extraction records. Subject to consortium approval and project dissemination requirements, these materials will be made available through a public repository following completion of the project.

Data structuring and synthesis An Excel-based coding framework was developed to organise and analyse the extracted evidence. Key fields included country, source title, source type, source URL, role title, role category, skill category, associated skills, maturity level, sector or mode, organisation type, relevance, validation status, and analytic comments. This coding structure supported transparency, auditability, and aggregation across countries. Following validation, 28 distinct transport-related roles were identified in the dataset, with frequencies ranging from one to sixteen occurrences across country-level records. To support robust consolidation, only roles appearing five or more times were retained for detailed clustering, while lower-frequency roles were excluded from role-family analysis but still informed broader capability and topic mapping. This refinement supports methodological clarity and explains how role evidence was translated into stable curriculum-relevant patterns.

Survey and interview methods The stakeholder survey was designed as a structured quantitative instrument aligned to the GreenShift analytical framework, covering digital maturity, AI and digital twin adoption, infrastructure readiness, barriers to adoption, sustainability capability needs, and programme preferences. It was administered online using Microsoft Forms between 3 March 2026 and 27 March 2026, distributed through consortium networks and open dissemination channels such as LinkedIn, and generated 58 valid responses for analysis. The survey sample is exploratory and not statistically representative of the full EU27 sector. Nine semi-structured interviews were also conducted with purposively selected stakeholders from industry, research, public sector, and technology contexts. Interviews were analysed using a hybrid thematic approach and used to validate and deepen understanding of organisational practices, implementation barriers, and future capability requirements.

4 Results

Our results indicate that the European transport and mobility sector is undergoing uneven but accelerating digital transformation, with foundational capabilities more established than advanced data-intensive, simulation, or integrated decision-support practices [23].

The identified four Role Families are:

1. **Transport Data Analytics**
2. **Modelling, Simulation and Digital Twin**
3. **Systems Integration and Platform**
4. **Digital Innovation Leadership (Transport-specific)**

These role families are designed as pedagogically actionable structures rather than as narrow occupational categories. AI capability, sustainability, governance, cybersecurity, systems thinking, and HPC should be treated as cross-cutting capability layers embedded across multiple role families rather than as isolated standalone domains [26]. This better reflects the convergence of technical, operational, sustainability, and governance demands in real transport systems.

Table 1 consolidates the role-family interpretation against external competence frameworks. The table is intentionally role-family based rather than role-title based. This avoids overfitting the curriculum to unstable job titles and shows how the evidence can be connected to recognised European reference frameworks while retaining sector-specific transport meaning.

Current capabilities and skills gaps. The analysis identifies a baseline of existing digital and operational capability across much of the sector, but also highlights major gaps in data integration, advanced analytics deployment, modelling and optimisation, systems integration, and digitally enabled systems management.

Table 1: Role-Family Alignment with Competence Frameworks and Curriculum Implications

Role family	**Primary capability focus**	**Relevant framework alignment**	**Curriculum implication**
Transport Data Analytics	Transport data pipelines, analytics, visualisation, and evidence-based decision support	ESCO data and transport occupations; e-CF data and analysis competences; DigComp information and data literacy; GreenComp sustainability data interpretation	Include transport data management, data quality, applied analytics, visualisation, and interpretation of operational and sustainability indicators.
Modelling, Simulation, and Digital Twin	Scenario modelling, simulation, optimisation, digital-twin reasoning, and HPC-enabled analysis	e-CF architecture and solution development competences; ESCO modelling and engineering roles; GreenComp systems thinking and futures literacy	Develop learning outcomes on simulation design, model assumptions, validation, uncertainty, computational scale, and interpretation of digital-twin outputs.
Systems Integration and Platform	Interoperability, infrastructure integration, data platforms, operational systems, and AI/HPC deployment contexts	e-CF systems architecture, integration, and service delivery competences; ECSF-adjacent architecture, implementation, and risk-management profiles; ESCO ICT and infrastructure roles	Embed systems architecture, interoperability, platform governance, operational resilience, and secure-by-design integration tasks into practical assessments.
Digital Innovation Leadership	Digital transformation, governance, procurement awareness, sustainability strategy, organisational change, and accountability	GreenComp sustainability competences; e-CF governance and strategy competences; ECSF-adjacent policy, compliance, and risk-management profiles; ESCO management and policy roles	Include governance, sustainability, risk, cybersecurity awareness, organisational readiness, and strategic roadmap assessments for AI- and HPC-enabled transport.

Advanced capabilities such as large-scale simulation, intelligent decision support, digital twins, and HPC-supported optimisation remain concentrated in larger organisations and specialised teams.

The analysis places a strong emphasis on the uneven diffusion of AI and HPC. The survey respondents recognised the importance of HPC for future use cases, such as optimisation and simulation, but the current deployment remained limited outside of research-intensive or large infrastructure settings. Similarly, the demand for AI capabilities exceeded the organisational readiness, with implementation often fragmented into pilots rather than embedded operational systems. Cross-cutting gaps also emerged in transversal skills. These include systems thinking, interdisciplinary collaboration, governance awareness, and the ability to connect technical tools to domain-specific operational and strategic decisions. These are not secondary soft skills, but central requirements for effective digital transformation in transport contexts.

Organisational barriers and enabling conditions. Major barriers include siloed structures, legacy systems, limited internal training pathways, resource constraints, fragmented data environments, and challenges in moving from innovation activity to operational deployment. SMEs appear particularly constrained by lower access to specialist expertise, compute infrastructure, and internal digital capability development [6, 11]. At the same time, the evidence points to important enablers. Stakeholders consistently emphasised the value of practical, flexible, and industry-relevant learning pathways, including modular study, micro-credentials, applied projects, and work-integrated approaches [13,25]. This finding directly supports a curriculum design model that balances conceptual depth with accessibility for learners entering from varied disciplinary and professional backgrounds.

Cross-method synthesis. The cross-method synthesis shows convergence across desk research, survey, and interviews. AI-related capability demand was widely evidenced in role documentation, but survey and interview data showed low implementation maturity and fragmented organisational uptake. Data integration emerged across all three methods as a foundational and cross-cutting challenge, while sustainability appeared as both a strategic priority and an operational capability gap. These differences are analytically important rather than problematic. The desk research revealed structural role and capability patterns, the survey provided directional evidence on adoption and barriers, and the interviews revealed the practical realities of implementation, including organisational silos, cultural barriers, and uneven readiness. Together, the three methods produce a richer basis for curriculum design than any source could provide alone.

The synthesis shows a consistent pattern: capability demand is outpacing organisational readiness. Although AI-, HPC-, and data-related roles are widely recognised, surveys and interviews indicate that adoption is constrained by organisational, infrastructural, and workforce factors. These shortages therefore reflect broader transformation challenges, not only technical skills gaps. Educational interventions must address both technological competencies and adoption capacity.

5 Discussion

The analysis strengthens the case for moving from technology-centred to capability-based and role-anchored curriculum design. Existing approaches frequently separate workforce analysis from curriculum development, creating limited traceability between identified skills needs and educational provision. Our methodology addresses this challenge by establishing an explicit evidence trail linking role identification, capability analysis, stakeholder validation, and curriculum design. In this model, curriculum is not derived from abstract disciplinary content alone, but from validated evidence about what professionals in digitally transforming transport systems need to understand and do.

5.1 From labour market evidence to curriculum architecture

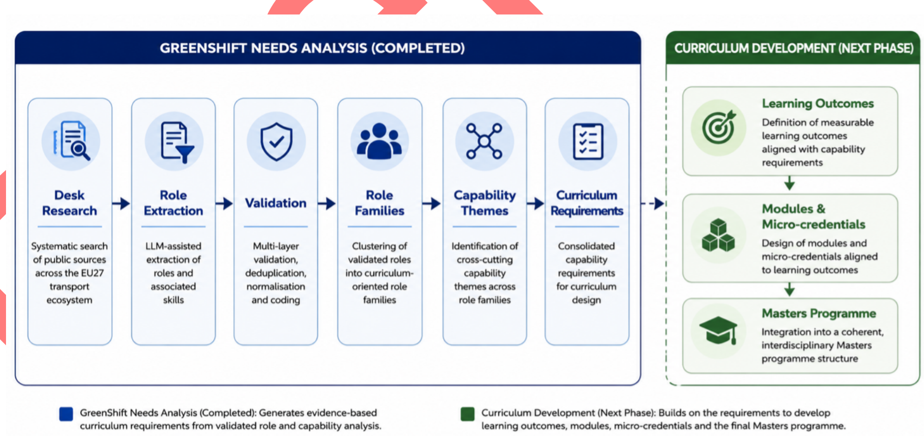

Fig. 1: Evidence-to-curriculum traceability framework. The GreenShift Needs Analysis establishes a pathway from desk research and validated role extraction to role families, capability themes, and curriculum requirements. The next phase translates these requirements into learning outcomes, modules, micro-credentials, and the Master's programme architecture.

A central contribution of the study is the curriculum traceability framework linking desk research roles to role families, capability themes, and curriculum requirements. The framework provides a structured pathway from labour market evidence to subsequent curriculum development activities, including learning outcome definition, module design, and programme architecture.

Figure 1 illustrates the curriculum traceability process developed within GreenShift. The framework provides an explicit link between labour market evidence and curriculum architecture, ensuring that programme design decisions can be traced back to validated role and capability requirements identified through desk research, stakeholder surveys, and interviews. This approach increases transparency, supports curriculum relevance, and provides a reproducible methodology for workforce-aligned educational design.

The role-family analysis should be read as an intermediate artefact in curriculum design, rather than as a final occupational taxonomy. Its purpose is to make role-evidence pedagogically usable. The desk research produced heterogeneous role titles across countries and organisations, but direct translation of these titles into modules would have produced an unstable and overly fragmented curriculum. The synthesis therefore moved from individual role titles to role families, and from role families to capability clusters. This enables curriculum design to respond to recurring work requirements without becoming dependent on the exact vocabulary of particular employers or national labour markets.

The four role families support different curriculum emphases. Transport Data Analytics roles require data acquisition, data management, transport analytics, visualisation, and evidence-based decision support. Modelling, Simulation and Digital Twin roles require understanding of simulation methods, scenario analysis, validation, model limitations, and the computational implications of large-scale modelling. Systems Integration and Platform roles require attention to interoperability, architecture, data platforms, infrastructure integration, and the operational embedding of AI- and HPC-enabled tools. Digital Innovation Leadership roles require governance, sustainability, risk, procurement awareness, cross-organisational coordination, and the ability to translate digital capabilities into organisational change.

These role families imply six curriculum topic groups. First, learners require a foundation in transport systems and domain context, including infrastructure, mobility systems, operations, planning, and sustainability pressures. Second, they require data management and analytics for transport, including transport data types, data pipelines, quality, integration, and analytical interpretation. Third, they require modelling, simulation, and HPC-enabled analysis, including scenario modelling, optimisation, computational scale, and interpretation of model outputs. Fourth, they require digital twins and advanced transport modelling, with emphasis on integrating data, models, simulations, and operational feedback loops. Fifth, they require systems architecture and integration for AI and HPC in transport, including interoperability, platforms, infrastructure, and links between analytical systems and operational environments. Sixth, they require decision support, strategy, and governance in AI-enabled transport, including accountability, risk, sustainability, cybersecurity, and organisational adoption.

This translation provides a concrete response to the curriculum-design problem. The expected next step is not merely to list skills, but to transform validated role-capability evidence into learning outcomes, module specifications, micro-credentials, assessment patterns, and learner pathways. For example, evidence of digital-twin and simulation roles can inform learning outcomes concerned with model interpretation, scenario design, computational constraints, and validation. Evidence of systems-integration roles can inform assessment through architecture design tasks or interoperability case studies. Evidence of digital innovation leadership roles can inform governance-oriented assessments, such as digital transformation roadmaps, sustainability impact analyses, or risk-aware deployment plans.

Table 2: Examples of Translation from Evidence to Curriculum Requirements

Evidence Source	**Key Finding**	**Curriculum Requirement**
Desk Research	Emerging transport roles increasingly require an understanding of transport systems alongside AI, HPC, digital twins, data analytics, and sustainability concepts. Many prospective learners are expected to enter from non-transport and non-computing backgrounds.	Introduce a foundational Transport Systems and Digital Transformation module providing an overview of the transport sector and its relationship with AI, HPC, sustainability, modelling, simulation, and digital transformation.
Survey	Respondents identified low organisational maturity in AI adoption despite recognising its future importance.	Include introductory AI and data-driven decision-making content to support learners entering with limited prior AI experience.
Interviews	Stakeholders consistently highlighted systems integration, interoperability, and cross-domain collaboration challenges.	Embed systems thinking, interoperability, and interdisciplinary problem-solving concepts across curriculum topics and practical activities.

The analysis also identifies a structured topic development process involving capability mapping, cross-family integration, learning architecture development, and strategic alignment with GreenShift priorities in AI, HPC, sustainability, and transport systems transformation. This supports a curriculum model that is both evidence-based and sufficiently flexible to accommodate future changes in technology and workforce demand.

The role families developed through the needs analysis provide the foundation for subsequent curriculum design activities within GreenShift. The next phase involves translating capability clusters into learning outcomes, module specifications, micro-credentials, assessment strategies, and learner pathways. Particular emphasis is placed on flexible entry routes for non-computer-science learners and on integrating AI, HPC, sustainability, governance, and cybersecurity as cross-cutting themes across the curriculum rather than as isolated specialist topics. This process has already informed the initial identification of curriculum topics, capability areas, and programme design considerations, demonstrating a direct pathway from workforce evidence to educational planning. While this paper focuses on the evidence-to-curriculum requirements phase, subsequent GreenShift work is now translating these requirements into detailed programme structures and pilot learning activities.

Implications for non-computer-science learners Our results reinforce the importance of designing for heterogeneous learner populations. Many prospective learners are likely to engage with HPC, AI, modelling, and digital infrastructure through applied contexts such as simulation, optimisation, emissions analysis, and data-driven decision support rather than through specialist software engineering roles.

This implies that curriculum should provide a strong foundation in transport systems and domain context while offering multiple access points into advanced digital concepts. It also suggests the need for modularity, scaffolded learning, practical cases, and programme structures that support both full Master's participation and shorter upskilling pathways.

Educational and labour market implications The analysis indicates growing demand for hybrid digital-green professionals who can interpret systems, integrate data, understand governance and sustainability constraints, and apply computational tools to operational transport challenges.

This has implications not only for GreenShift, but for advanced digital education more broadly. It suggests that AI and HPC education in application domains should be organised around capability integration, contextual use, and interdisciplinary problem solving rather than around purely technical specialisation. The same role-anchored, evidence-to-curriculum workflow can be adopted to organise cybersecurity education around validated role profiles and evolving capability requirements, providing a structured alternative to static, technology-centred syllabi.

Although developed in the context of transport systems, the methodological approach has direct relevance for cybersecurity education. The challenge of rapidly evolving role requirements, fragmented competency definitions, and emerging technology adoption is equally visible across cybersecurity domains. The GreenShift workflow therefore offers a potential model for identifying and validating curriculum requirements in distinct computing contexts.

6 Conclusion

This study contributes both substantive findings on HPC and AI capability requirements in transport systems and a transferable methodology for the design of evidence-based curriculum. The principal contribution lies in demonstrating how bounded LLM-assisted extraction, human validation, framework mapping, stakeholder engagement, and curriculum traceability can be integrated into a coherent and academically defensible workflow.

Our results show that the digital transformation of the transport and mobility sectors is real but uneven and that capability development has become a critical condition to translate AI, HPC, and advanced digital technologies into operational and strategic value [23]. The study also demonstrates that curriculum design in such contexts should be grounded in validated role-capability evidence, organised through curriculum-oriented role clusters, and delivered through flexible, interdisciplinary learning architectures that are accessible to non-traditional and non-computer-science learner cohorts.

Some limitations should be acknowledged. First, the study is exploratory and does not claim statistical representativeness across the entire European transport ecosystem. Second, desk research was dependent on publicly available sources whose quality and coverage varied considerably between countries. Third, although LLM outputs were manually validated, extraction quality remains partially dependent on source accessibility and prompt design. Fourth, the rapid evolution of AI, HPC, and transport technologies means that identified capability requirements should be regarded as indicative rather than permanent. Finally, while the methodology was developed within the transport domain, further validation would be required before generalising results to other sectors.

In summary, the GreenShift methodology provides a transferable model for curriculum design in sectors undergoing rapid technological change. The approach demonstrates how role identification, competency analysis, framework mapping, stakeholder validation, and curriculum development can be integrated within a transparent and reproducible workflow. This contribution is potentially applicable to domains such as cybersecurity, energy, manufacturing, and healthcare, where emerging technologies are reshaping workforce requirements faster than traditional curriculum design processes can respond.

Acknowledgements

This work is developed as part of "GreenShift: EU MSc in AI & HPC for Green Digital Innovation in Transport", a project funded by the European Commission, grant no. 101226222; `https://www.thegreenshift.eu/`.

The authors thank the following individuals who collected and validated results: D. Boyle, M. Castelli, M. Danelutto, D. Harrison, D. Janetzko, I. Kovačević Vukas, C. Landolfo, B. Lu, G. Meijer, G. Mencagli, G. O'Grady, and R. Talevska.

Conflicts of Interest The authors declare that they have no competing interests.

Appendix A Representative LLM Extraction Prompt

Listing 1.1: Representative category-specific LLM prompt for evidence extraction

```
A. HPC - High Performance Computing

Search string:
"[Country] high performance computing transport"

LLM Search Request:
Identify explicitly evidenced roles and skills related to the use of High Performance
     Computing (HPC) in the transport sector in [COUNTRY].

Search government policy, national HPC centres, transport authorities, academic research, and
      sector bodies where transport use cases of HPC are clearly stated. Relevant use cases
     include large-scale simulation, transport modelling, infrastructure optimisation,
     traffic systems, logistics modelling, digital twins, and scenario analysis.

Extraction rules:
- Extract only roles and skills explicitly mentioned in the source.
- Do not infer a role from a technology description.
- Do not infer a skill from a project description unless the skill is explicitly stated.
- Use approved role dropdowns only.
- Record skills exactly as evidenced.
- Assign maturity level only where the source provides sufficient evidence.
- Assign SME or MNC relevance only where the source provides sufficient evidence.
- Exclude generic computing strategies, marketing content, and non-transport HPC use cases.
- Exclude inaccessible sources and sources without stable bibliographic or organisational
     information.
- Output Sheet-02-ready records with accessible URLs only.

Skill Category:
HPC - High Performance Computing only.

Zero-row rule:
If the source does not explicitly evidence both a transport-related role and an associated
     HPC skill, return zero rows. Do not create a plausible or inferred entry.

Output fields:
Country | Source title | Source type | Source URL | Role title | Role category | Skill
     category | Associated skill(s) | Transport mode or sector | Organisation type |
     Evidence note | Validation status

Additional searches using the same rules:
1. "[Country] HPC simulation infrastructure"
2. "[Country] national HPC centre transport"
3. "[Country] large-scale simulation transport modelling"
4. "[Country] digital twin transport HPC"
```

The listing shows the controlled prompt structure used for category-specific extraction. The prompt illustrates the conservative extraction logic rather than the complete prompt library.

Sources containing evidence for more than one category were processed through repeated category-specific passes. Records were retained only where the source explicitly evidenced both the role and the associated skill. The same pattern was used for AI, HPC, cybersecurity, governance, sustainability, data, modelling, systems integration, innovation, production, business, leadership, and related capability categories.

References

1. Aggarwal, D., Bokshi, A., Lad, D.: A hands-on approach for scalable parallel applications development: From testbed to petascale. In: HiPC'22. pp. 36–43. Bengaluru (Dec 2022)
2. Aljohani, N.R., Aslam, A., Khadidos, A.O., Hassan, S.U.: Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature. Journal of Innovation & Knowledge **7**(3), 100190 (2022)
3. Allam, Z., Dhunny, Z.A.: On big data, artificial intelligence and smart cities. Cities **89**, 80–91 (2019)
4. Almeida, F.: Comparative analysis of EU-based cybersecurity skills frameworks. Computers & Security **151**, 104329 (2025)
5. Amershi, S., et al.: Guidelines for Human-AI interaction. In: CHI '19. pp. 1–13. ACM, Glasgow (May 2019)
6. Autor, D.H.: Why are there still so many jobs? the history and future of workplace automation. Journal of Economic Perspectives **29**(3), 3–30 (2015)
7. Bacigalupo, M., Binasco, A.: Teaching different key competences at once. EUR Scientific and Technical Research Report JRC136734, Publications Office of the European Union, Luxembourg (2024). https://doi.org/10.2760/013612
8. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: Can language models be too big? In: FAccT '21. pp. 610–623. ACM, Virtual Event, Canada (Mar 2021)
9. Bouvry, P., et al.: The European master for HPC curriculum. Journal of Parallel and Distributed Computing **201**, 105081 (2025)
10. Carneiro Neto, J.A., Alves Neto, A.J., Moreno, E.D.: A systematic review on teaching parallel programming. In: EATIS '22. ACM, Aveiro (Jun 2022)
11. Cedefop: European Centre for the Development of Vocational Training: Skills in transitions – the way to 2035. Tech. Rep. 4213, Publications Office of the European Union, Luxembourg (Sep 2023), `https://data.europa.eu/doi/10.2801/438491`
12. Chen, J., Ghafoor, S., Impagliazzo, J.: Producing competent HPC graduates. Communications of the ACM **65**(12), 56–65 (2022)
13. Council of the European Union: Council recommendation of 16 june 2022 on a European approach to micro-credentials for lifelong learning and employability. Official Journal of the European Union **C**(243), 10–25 (Jun 2022)
14. Danelutto, M.: HPC master design: experience from Pisa. In: PDP 2025. pp. 357–363. IEEE, Torino (Mar 2025)
15. ENISA: European Cybersecurity Skills Framework (ECSF). CC-BY 4.0 report, European Union Agency for Cybersecurity (ENISA) (Sep 2022). https://doi.org/10.2824/859537
16. Fetting, C.: The European Green Deal. Report, European Sustainable Development Network (ESDN), Vienna (Dec 2020), `https://www.esdn.eu/fileadmin/ESDN_Reports/ESDN_Report_2_2020.pdf`
17. González-Vélez, H., Feenan, D., Bradford, M., Henriques, R., Panciroli, C., Cvetkovska, M., Chis, A.E., Somers, C.: Digital4Business: Mastering advanced digital skills for European business transformation. In: WorldCIST 2025. LNNS, vol. 1583, pp. 377–390. Springer, Florianopolis (Mar 2025)
18. Kasneci, E., et al.: ChatGPT for good? on opportunities and challenges of large language models for education. Learning and Individual Differences **103**, 102274 (2023)

19. Kiesler, N.: Approaching the concept of competency. In: Modeling Programming Competency, chap. 1, pp. 17–36. Springer, Cham (2024)
20. Li, C., Guo, X., Yang, W., Chen, J.: Interdisciplinary practice for HPC competency cultivation. In: CSE 2025. pp. 1–8. IEEE, Guiyang (Nov 2025)
21. Mhaske, P., Bhattacharjee, B., Haldar, N., Upadhyay, P., Mandal, A.: Bridging digital skill gaps in the global workforce: A synthesis and conceptual framework building. Research in Globalization p. 100311 (2025)
22. Mulder, M. (ed.): Competence-based Vocational and Professional Education: Bridging the Worlds of Work and Education, Technical and Vocational Education and Training: Issues, Concerns and Prospects, vol. 23. Springer, Cham, 1 edn. (2017)
23. OECD: OECD employment outlook 2025: Can we get through the demographic crunch? Tech. rep., OECD Publishing, Paris (2025), `https://doi.org/10.1787/194a947b-en`
24. Raj, R.K., et al.: High performance computing education: Current challenges and future directions. In: ITiCSE-WGR '20. pp. 51–74. ACM, Trondheim (2020)
25. Ricci, S., Dzurenda, P., Somers, C., González-Vélez, H., Moravek, L.J.: Cybersecurity micro-credentials and career path design: The Digital4Security good practices. In: ARES 2025, EU Projects Symposium Workshops. LNCS, vol. 15999, pp. 77–95. Springer, Ghent (Aug 2025)
26. Righi, R., Lopez-Cobo, M., Alaveras, G., Samoili, S., Cardona, M., Baillet, M.V.P., Ziemba, L., De-Prato, G.: Academic offer of advanced digital skills in 2019-20. international comparison. focus on Artificial Intelligence, High Performance Computing, Cybersecurity and Data Science. JRC Research Reports JRC121680, Joint Research Centre (Sep 2020), `https://ideas.repec.org/p/ipt/iptwpa/jrc121680.html`
27. Silva, B.N., Khan, M., Han, K.: Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. Sustainable Cities and Society **38**, 697–713 (2018)
28. Sinnett, W.E.: The application of DACUM in retraining and post-secondary curriculum development. Tech. Rep. ED119013, Education Resources Information Center (Jan 1976), `https://eric.ed.gov/?id=ED119013`, Revised Second Edition
29. Soderquist, K.E., Papalexandris, A., Ioannou, G., Prastacos, G.: From task-based to competency-based: A typology and process supporting a critical HRM transition. Personnel Review **39**(3), 325–346 (2010)
30. Somers, C., Feenan, D., Fitzgerald, D., Henriques, R., Martignoni, M., Parletta, D.A., Cibin, E., Chis, A.E., González-Vélez, H.: Systematic needs analysis of advanced digital skills for postgraduate computing education: The DIGITAL4Business case. In: AIED 2024. CCIS, vol. 2150, pp. 179–191. Springer, Recife (Jul 2024)
31. Stevens, C., Anderson, S., Carlson, A.: Integrating high performance computing into higher education and the pedagogy of cluster computing. In: PEARC '24. ACM, Providence (Jul 2024)