Lightcast

Bridging Skills Demand and Vocational Qualifications: Linking Online Job Postings with National Skills and Qualification Frameworks

HARNESSING WEB DATA FOR NEXT-GENERATION SKILLS INTELLIGENCE

THESSALONIKI · 28–29 MAY 2026

Mauro Pelucchi · Sophie Gvasalia · Anna Clara Gatti
Ettore Colombo · Luca Agostoni · Simone Perego

What we will cover

01 **Motivation and research question**
Why a semantic bridge between OJPs and VET?

02 **Italian institutional landscape**
VET pathways anchored on the Atlante del Lavoro

03 **Data sources**
11.8 M postings · 5,160 courses · 43.5 K skill links

04 **Methodology**
SBERT + LLM matching, notation, and the VAI

05 **Results · skill level**
Common and Specialised mismatches

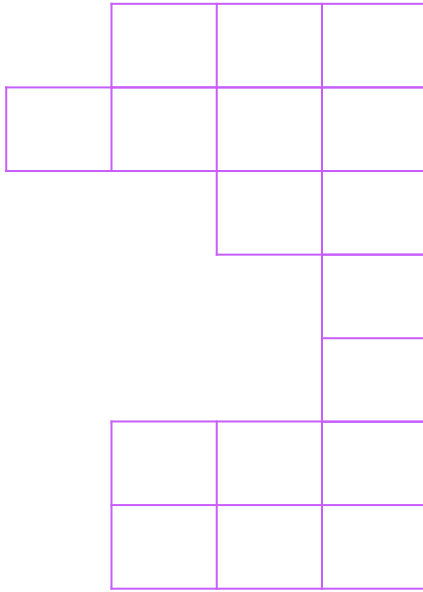
06 **Results · sectoral level**
VAI ranking and the four-quadrant matrix

07 **Policy implications**
A continuous-monitoring policy cycle

08 **Next steps and research agenda**
From offer to effective skill provision

Why a semantic bridge between OJPs and VET?

- OJP-based skills intelligence is organised around occupations (ISCO\ESCO, national qualification).
- VET systems are organised around qualifications and learning outcomes.
- The two universes do not share a common semantic layer.
- **Result: web evidence circulates without informing curriculum updates.**



How aligned is what we teach
with what employers demand?

And how can the answer be made:

- measurable at sector level,
- interpretable across skill types,
- and operational for VET policy?

Contribution

METHODOLOGICAL

A replicable framework

Two-stage NLP pipeline (multilingual SBERT + LLM filter) anchored on a national qualification repository through the Atlante del Lavoro activities layer.

ANALYTICAL

A composite indicator: VET Alignment Index (VAI)

Two-component index combining quantitative (top-K demand-weighted coverage) and qualitative (SBERT cosine similarity) terms, separately by skill type.

Roadmap

EMPIRICAL

First national-level evidence

Skill- and sector-level alignment for the full Italian VET system: 5k courses across IeFP / IFTS / ITS pathways and 23+1 Economic-Professional Sectors.

POLICY

From description to decision

A continuous-monitoring cycle and a four-quadrant priority matrix that translate the indicator into operational signals for sector boards and qualification authorities.

Italian VET: three pathways, one semantic infrastructure

Atlante del Lavoro e delle Qualificazioni: Reticular description of the world of work · 24+1 Economic-Professional Sectors (SEP) · 840 Areas of Activity (ADA) · indexed to CP2021 occupations (mappable to ISCO-08) and to ATECO 2007 sectoral codes (NACE Rev. 2 compatible)

1 IeFP

EQF 3–4 · 3–4 years

GOVERNANCE
Regional providers

TARGET QUALIFICATION
Vocational Operator Qualification (3 yrs) · Vocational Diploma (4 yrs)

COVERAGE IN CORPUS
Upper-secondary regional pathway · largest by volume

2 IFTS

EQF 4 · ~800–1,000 hours

GOVERNANCE
Regional partnerships (schools, providers, firms, universities)

TARGET QUALIFICATION
Higher Technical Specialization Certificate

COVERAGE IN CORPUS
Intermediate share · technician-level demand

3 ITS Academy

EQF 5 · 2 years (1,800–2,000 hrs)

GOVERNANCE
ITS Foundations (school, university, firm, training body)

TARGET QUALIFICATION
Higher Technician Diploma

COVERAGE IN CORPUS
Smaller volume · strategic technological areas (D.M. 203/2023)

ADA as the semantic pivot between demand and supply

DEMAND SIDE

Online job postings

Each posting carries an occupational code (CP2021 / ISCO-08) and an ATECO sectoral code attached by the Lightcast pipeline.

CP2021 codes index one or more ADA

SEMANTIC BRIDGE

Areas of Activity (ADA)

840

4-LEVEL HIERARCHY

Sectors → Processes → ADA → Qualifications

SUPPLY SIDE

Vocational qualifications

Each qualification in the INAPP National Repository is linked to one or more ADA in a logic of potential employment outlets.

Qualifications index one or more ADA

Each ADA is doubly referenced — via occupational codes (reachable from postings) and via qualifications (reachable from courses).

Routing demand and supply through a shared ADA layer enables a like-for-like comparison: this is the bridge the rest of the study exploits.

Three sources, one integrated dataset

CLASSIFICATIONS USED

SKILLS	Lightcast Open Skills taxonomy · Specialised (technical) vs Common (transversal) types
OCCUPATIONS	CP2021 (Italian) at posting level · mappable to ISCO-08 for international comparability
SECTORS	ATECO 2007 at posting level · compatible with NACE Rev. 2 · projected onto 25 SEP via the Atlante
QUALIFICATIONS	INAPP National Repository · 840 ADA · IeFP / IFTS / ITS pathways
EMBEDDINGS	Multilingual SBERT (sentence-transformers, 384-dimensional) for the qualitative component of the VAI

ITALIAN ONLINE JOB POSTINGS

Lightcast Italian database (11M job postings)) · collected since January 2023 · skills extracted by the Lightcast Open Skills NLP pipeline.

VOCATIONAL COURSES

IeFP · IFTS · ITS Academy · National Repository of Qualifications maintained by INAPP (~5k qualifications).

COURSE–SKILL LINKS

After matching · 8.4 Lightcast skills per course on average · 23+1 SEP sectors covered (~43k skills extracted).

The course corpus by pathway

PATHWAY	EQF	DURATION	GOVERNANCE	COURSES (N)	SHARE
IeFP — Initial vocational pathways	3–4	3–4 years	Regional (and IP)	≈ 3,600	**≈ 70%**
IFTS — Higher technical specialisation	4	≈ 800–1,000 h	Regional partnerships	≈ 1,200	**≈ 23%**
ITS Academy — Tertiary technical	5	2 years (1,800–2,000 h)	ITS Foundations	≈ 360	**≈ 7%**
ALL PATHWAYS	**3–5**	—	—	**5,160**	**100%**

Pathway weighting

IeFP dominates by volume (≈70% of courses), which shapes how the integrated dataset represents Italian VET.

Granularity

ITS Academy courses, though fewer, contribute disproportionately to the technical and digital skill coverage.

Why no enrolment weights

Enrolment counts are inconsistently reported across pathways and regions; **we measure the offer of training, not exposure.**

Three sources, one integrated dataset

DEMAND SIGNAL

Online job posting

Electrical installer

for residential buildings

Source
Lightcast Italian OJPs

Region
Lombardia · 2024

OCCUPATIONAL CODE

CP2021 (mappable to ISCO-08)

6.1.3.7.0

Electricians and installers of electrical systems in civil constructions

ISCO-08
7411 · Electrical fitters

EQF level
3–4

ATLANTE PIVOT

Sector + Area of Activity

SEP 10

Meccanica · ADA.10.04.05

ADA label
Installation of electrical systems

Process
Plant engineering — residential

QUALIFICATION

INAPP National Repository

Operatore elettrico

IeFP · EQF 3 · 3-year pathway

Channel
IeFP

Indexed to ADA
via the Atlante

SUPPLY SIGNAL

Course in the corpus

Course

INSTALLATORE E MANUTENTORE ELETTRICO ED ELETTRONICO (Lombardia)

Linked to ADA
via the qualification

Linking courses to skills through SBERT + LLM verification

WHY A TWO-STAGE DESIGN

Recall and precision trade off across thresholds. SBERT retrieval maximises recall on a wide candidate set; the LLM step recovers precision by rejecting semantically-plausible but contextually-irrelevant pairs.

01

Input

Free-text course descriptions and competence units in Italian. Sources: National Repository, IeFP profiles, ITS reference profiles (D.M. 203/2023).

02

Retrieval

A multilingual SBERT model (paraphrase-multilingual-mpnet-base-v2) encodes both course passages and Lightcast skill descriptions into a 384-dimensional space. Top-K candidate skills are retrieved per course.

03

LLM verification

A large language model classifies each candidate (skill, course) pair as relevant or non-relevant, using the course passage and the canonical skill name + description as input. Acts as a precision filter.

04

Thresholding

A cosine threshold (≥ 0.55 in baseline) is retained after LLM filtering. Validated against a manually-coded subsample of courses across pathways.

05

Output

43,545 validated course–skill links · 8.4 skills per course on average · attached to the ADA layer and to the SEP sector of the course.

Linking courses to skills through SBERT + LLM verification

WHY A TWO-STAGE DESIGN

SBERT alone would keep all eight skills above 0.55; the LLM filter rejects Power Distribution, Solar Panel Installation and Electrical Drafting as out of scope for an EQF-3 residential course.

COURSE INPUT

Impianti elettrici civili (IeFP — Operatore elettrico, EQF 3)

ADA

09.05.07

SEP

09 · Construction

STAGE 1 + 2 · SBERT TOP-K WITH LLM VERIFICATION

CANDIDATE SKILL	COS	VERDICT	LLM RATIONALE
Electrical Installation	0.91	KEEP	*Core practical skill for residential wiring.*
Wiring	0.88	KEEP	*Directly taught in module 2 of the course.*
Electrical Safety	0.84	KEEP	*Explicit competence unit on safety norms.*
Circuit Testing	0.79	KEEP	*Listed under verification activities.*
Conduit Bending	0.71	KEEP	*Practical activity in the workshop block.*
~~**Power Distribution**~~	0.69	DROP	*Industrial scope; course is residential only.*
~~**Solar Panel Installation**~~	0.66	DROP	*Out of programme scope at EQF 3.*
~~**Electrical Drafting**~~	0.58	DROP	*Not in learning outcomes; auxiliary skill.*

OUTPUT · ATTACHED TO THE COURSE

5 validated skill links (5 KEEP / 3 DROP)

VALIDATED SKILL SET

- Electrical Installation
- Wiring
- Electrical Safety
- Circuit Testing
- Conduit Bending

The VET Alignment Index: the elementary indicators

PRIMITIVES

P	Universe of Italian online job postings	`|P| = 11,760,514`
C	Universe of vocational courses	`|C| = 5,160`
S	Universe of Lightcast skills (Spec + Common)	–
Σ	23+1 Economic-Professional Sectors	`|Σ| = 24`
T(s)	Skill type: Specialised | Common	–
skills(p)	Skills extracted from posting p	`from postings`
skills(c)	Skills attached to course c	`from courses`
s(p), s(c)	SEP sector of posting / course	`∈ Σ`

ELEMENTARY INDICATORS

Demand share of skill s

demand_share(s) =

```
| { p ∈ P : s ∈ skills(p) } | / |P|
```

Supply share of skill s

supply_share(s) =

```
| { c ∈ C : s ∈ skills(c) } | / |C|
```

Skill-level gap

gap(s) = demand_share(s) – supply_share(s)

Sectoral versions: restrict numerator / denominator to postings and courses in σ.

The VET Alignment Index and its two components

$$VAI(\sigma, T) = \alpha \cdot \min(Quant(\sigma, T), 1) + (1 - \alpha) \cdot Qual(\sigma, T)$$

baseline α = 0.2 (qualitative-dominant) · K = 30 (top-K cutoff, following OECD 2023) · evaluated separately for T ∈ {Specialised, Common}

QUANTITATIVE COMPONENT

Demand-weighted coverage

$D_K(\sigma, T)$ = top-K skills of type T in σ ranked by demand_share_σ

$taught(\sigma, T) = \{ s \in S : T(s) = T, supply\_share\_\sigma(s) > 0 \}$

Quant(σ, T) =

$$\Sigma_{s \in D_K(\sigma, T) \cap taught(\sigma, T)} demand\_share(s)$$

QUALITATIVE COMPONENT

Semantic similarity

$v(s) \in R^{384}$ = multilingual SBERT embedding of skill s

$sim(s, s') = cos(v(s), v(s'))$

Qual(σ, T) =

$$(1/|D_K|) \cdot \Sigma_{s \in D_K} \max_{s' \in taught} sim(s, s')$$

Calibration choices

WEIGHTING α

α = 0.2

ARGUMENT

Quant has limited discriminatory power once partial coverage is achieved — it cannot distinguish a sector teaching ten of the top demanded skills well from one teaching them poorly. Qual recovers that information.

ROBUSTNESS

Re-running with α ∈ {0.0, 0.2, 0.5, 0.8, 1.0} shifts sector ranks by at most two positions; the bottom-quartile priority ranking is preserved at every α.

TOP-K CUTOFF

K = 30

ARGUMENT

Follows the OECD (2023) precedent. Captures the demand-relevant tail without diluting the indicator with marginally-demanded skills.

ROBUSTNESS

Increasing K from 30 → 50 → 100 leaves the identity of the bottom-three and top-three sectors unchanged. Restricting to courses with explicit ADA metadata yields the same substantive ranking.

SUPPLY SIDE

Offer, not exposure

ARGUMENT

Each course counts once: VAI measures the structure of the training offer, not the volume of learners actually exposed to it. Enrolment data are not consistently available across pathways and regions.

ROBUSTNESS

Enrolment-weighted re-estimation deferred to future work. Qualitative findings are not expected to change; quantitative magnitudes might.

The VET Alignment Index: how to read

FIRST COMPONENT: BREADTH

Quant

What share of demand is taught at all?

Σ *demand_share over taught top-K skills*

WHAT IT MEASURES

The fraction of demand-weighted skills covered by the training offer - coverage of what employers actually ask for.

RATIONALE

A sector can teach many skills, but if none are demanded, alignment is illusory. Quant disciplines the indicator with employer-weighted demand and penalises curricula that ignore high-frequency skills.

SECOND COMPONENT: DEPTH

Qual

How close are the skills that ARE taught?

mean cosine to closest taught skill, over top-K demanded

WHAT IT MEASURES

Semantic proximity in SBERT space between every demanded skill and the closest skill actually taught - the depth of the match.

RATIONALE

Two sectors with identical Quant can differ sharply: one teaches the demanded skill, the other a vague cousin. Qual recovers this distinction that Quant cannot make.

READING A VALUE · VAI ∈ [0, 1]

Lightcast

The VET Alignment Index: decomposition for Construction

INPUT · TOP-K DEMANDED SPECIALISED SKILLS, σ = 09 (CONSTRUCTION)

#	SPECIALISED SKILL	DEMAND	TAUGHT?	COS	CONTRIB. Q	CONTRIB. SIM
1	**Wiring**	8.4%	✓	0.92	8.4	0.92
2	**Electrical Installation**	6.1%	✓	0.89	6.1	0.89
3	**Working Drawings**	5.7%	✓	0.81	5.7	0.81
4	**Structural Steel**	4.2%	—	0.42	0.0	0.42
5	**Plumbing**	3.9%	✓	0.78	3.9	0.78
6	**Welding**	3.5%	✓	0.74	3.5	0.74
7	**Safety on Site**	3.2%	✓	0.88	3.2	0.88
…	(further skills, ranks 8–30)	…	…	…	…	…
	TOTAL OVER K = 30				**48.7%**	**0.76**

COMPUTATION · SPECIALISED

STEP 1 · QUANTITATIVE

Quant(σ, T) = Σ demand_share over taught top-K skills

Quant = 0.487

STEP 2 · QUALITATIVE

Qual(σ, T) = mean cosine to closest taught skill, over top-K demanded

Qual = 0.762

STEP 3 · COMPOSITE (α = 0.2)

VAI = α · min(Quant, 1) + (1 − α) · Qual

= 0.2 · 0.487 + 0.8 · 0.762

VAI(Spec) = 0.609

Lightcast

Four findings
Skill-level evidence and sectoral patterns

21% / 28%	Asymmetric universes	*The supply universe is structurally narrower than the demand universe; only 21% of the Specialised skill set and 28% of the Common skill set are covered by both sides (Jaccard).*
20.4 pp	Qualitative transversal gap	*The transversal skills employers most consistently demand are essentially absent from formal curricula: a sector-invariant, systematic and severe under-supply.*
20k vs 398	Dispersed Specialised mismatch	*Demand for technical skills is spread over a much larger universe than for transversal skills; the Specialised mismatch is sector-specific and concentrated on cross-occupational competences.*
18 / 24	Structural type asymmetry	*Manufacturing-intensive sectors codify technical content; care and service sectors codify transversal content. The asymmetry is institutional, not pedagogical.*

Demand and supply universes, by skill type

QUANTITY	SPECIALISED	COMMON
Distinct skills demanded (postings)	20,791	398
Distinct skills taught (courses)	4,986	143
Skills both demanded and taught	4,476	119
Overlap (% of union)	21%	28%

Size of the demand and supply universes by skill type. |P| = 11,760,514 postings, |C| = 5,160 courses.

STRUCTURAL ASYMMETRY

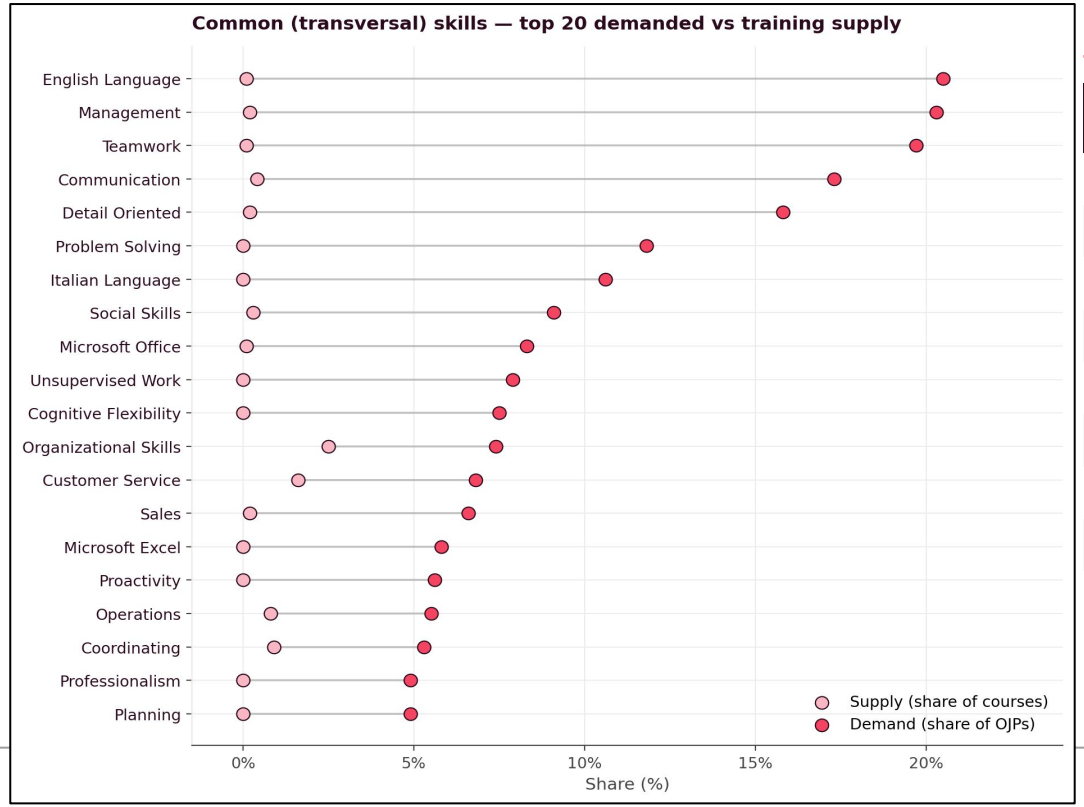
Specialised demand spans 20,791 distinct skills; the supply universe covers 4,986 — narrower by design, since VET cannot teach the full long tail.

TYPE COMPRESSION

Common demand spans only 398 skills, of which 143 are taught — but with a far more concentrated demand distribution (top ten skills > 50%).

Common skills: a systematic and uniform under-supply

Five of the top fifteen Common skills are taught in 0% of the 5k vocational courses analysed

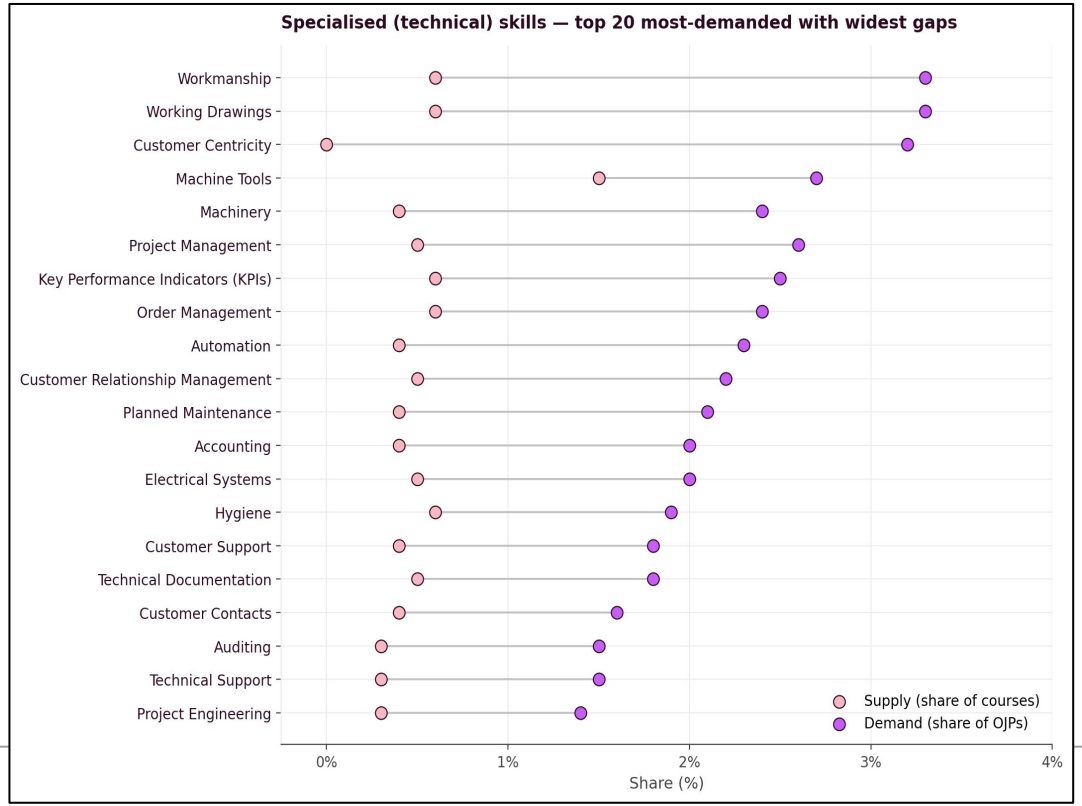


TOP 10 COMMON-SKILL GAPS (PERCENTAGE POINTS)

SKILL	DEMAND	SUPPLY	GAP (PP)
English Language	20.5%	0.1%	20.4
Management	20.3%	0.2%	20.1
Teamwork	19.7%	0.1%	19.6
Communication	17.3%	0.4%	16.9
Detail Oriented	15.8%	0.2%	15.6
Problem Solving	11.8%	0%	11.8
Italian Language	10.6%	0%	10.6
Social Skills	9.1%	0.3%	8.8
Microsoft Office	8.3%	0.1%	8.2

Specialised skills: a sector-specific mismatch

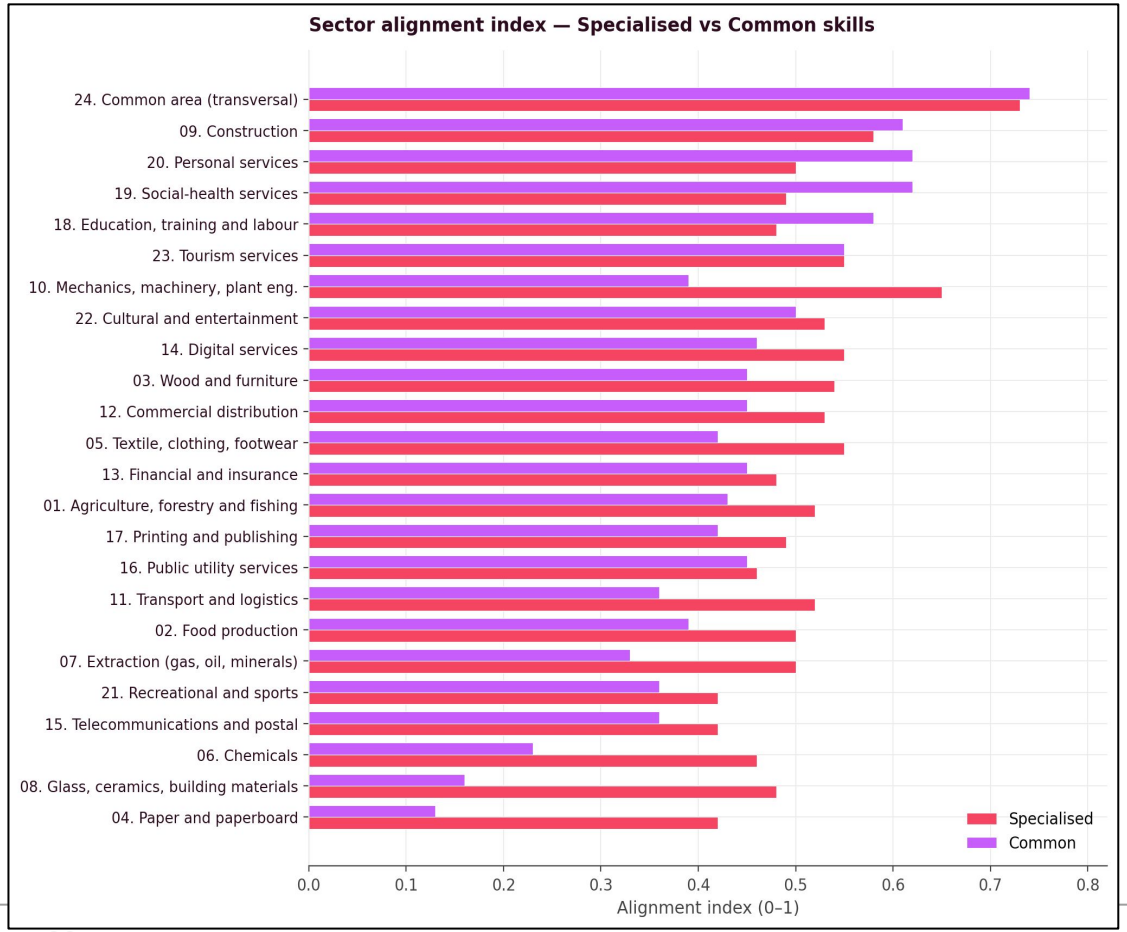
Specialised gaps are spread thinly: no single skill exceeds 5% of postings, but a long tail of mid-frequency competences accumulates.



TOP UNDER-SUPPLIED SPECIALISED SKILLS

SKILL	DEMAND	SUPPLY	GAP (PP)
Loading and Unloading	4.9%	0.0%	4.9
Packaging and Labeling	3.6%	0.0%	3.6
Workmanship	3.3%	0.6%	2.7
Working Drawings	3.3%	0.6%	2.7
Customer Centricity	3.2%	0.0%	3.2
Results Orientation	3.2%	0.5%	2.7
Project Management	2.6%	0.5%	2.1
Invoicing	2.4%	0.0%	2.4
Order Management	2.4%	0.6%	1.8
Customer Relationship Mgmt	2.2%	0.5%	1.7

Sectoral alignment and the type-asymmetry divide



WORST ALIGNED (LOW VAI)

04. Paper and paperboard	**0.276**	n=15
08. Glass, ceramics, building materials	**0.322**	n=108
06. Chemicals	**0.350**	n=68

BEST ALIGNED (excl. Common area)

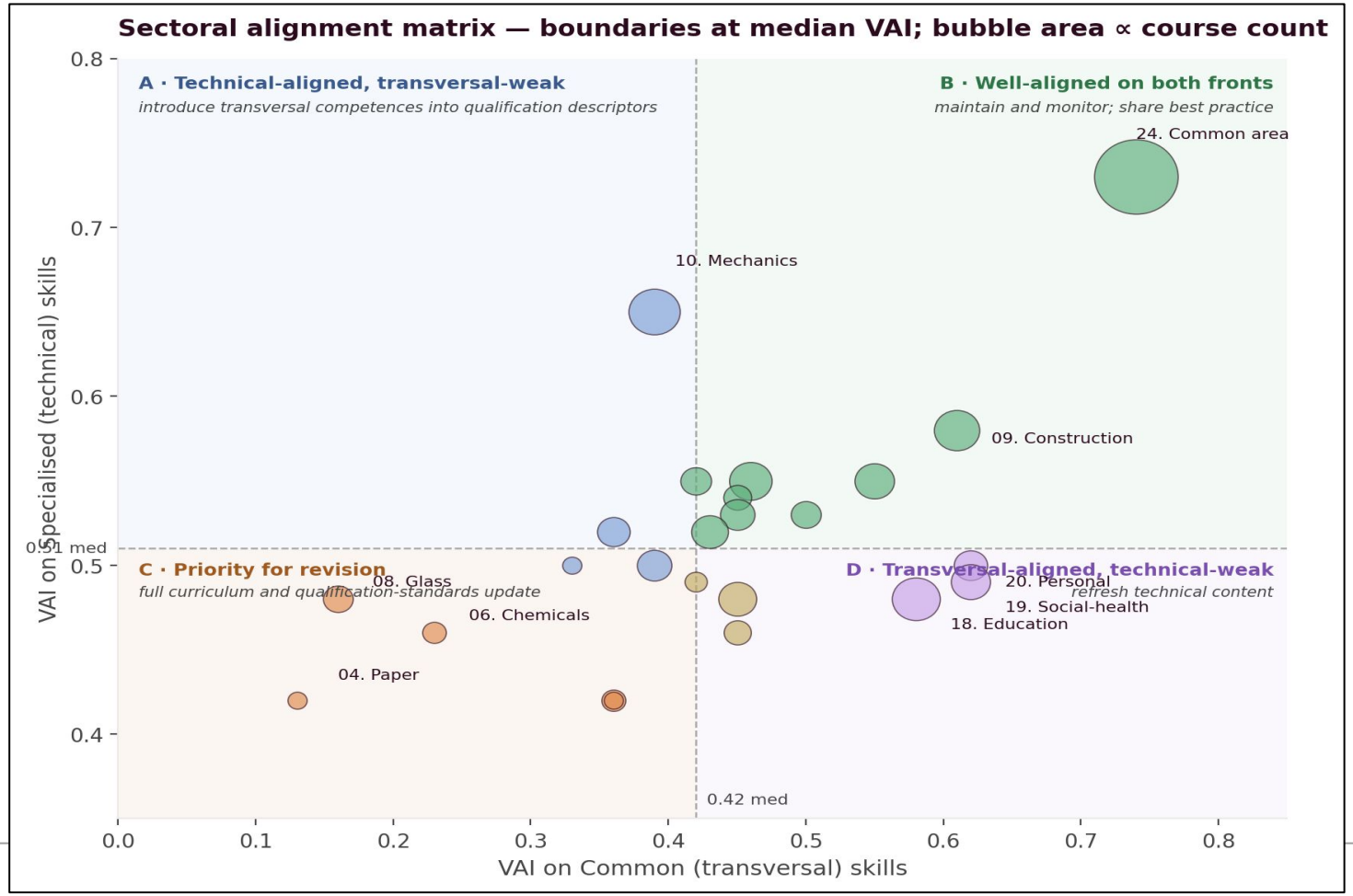
09. Construction	**0.584**	n=250
20. Personal services	**0.560**	n=136
19. Social-health services	**0.558**	n=187

STRUCTURAL DIVIDE

Eighteen of twenty-four substantive sectors show VAI(Spec) > VAI(Common), with a mean asymmetry of +0.13. The asymmetry is institutional, not pedagogical — manufacturing codifies technical work; care and service sectors codify transversal work.

Specialised vs Common VAI by sector. Source: own elaboration.

From evidence to action: a four-quadrant priority matrix



QUADRANT A

Technical-aligned, transversal-weak

Mechanics, Glass, Chemicals · integrate transversal competences into qualification descriptors.

QUADRANT B

Well-aligned on both

Construction, Tourism, Common area · maintain and disseminate as benchmark cases.

QUADRANT C

Priority for revision

Paper, Telecoms, Recreational · full curriculum and qualification-standards update.

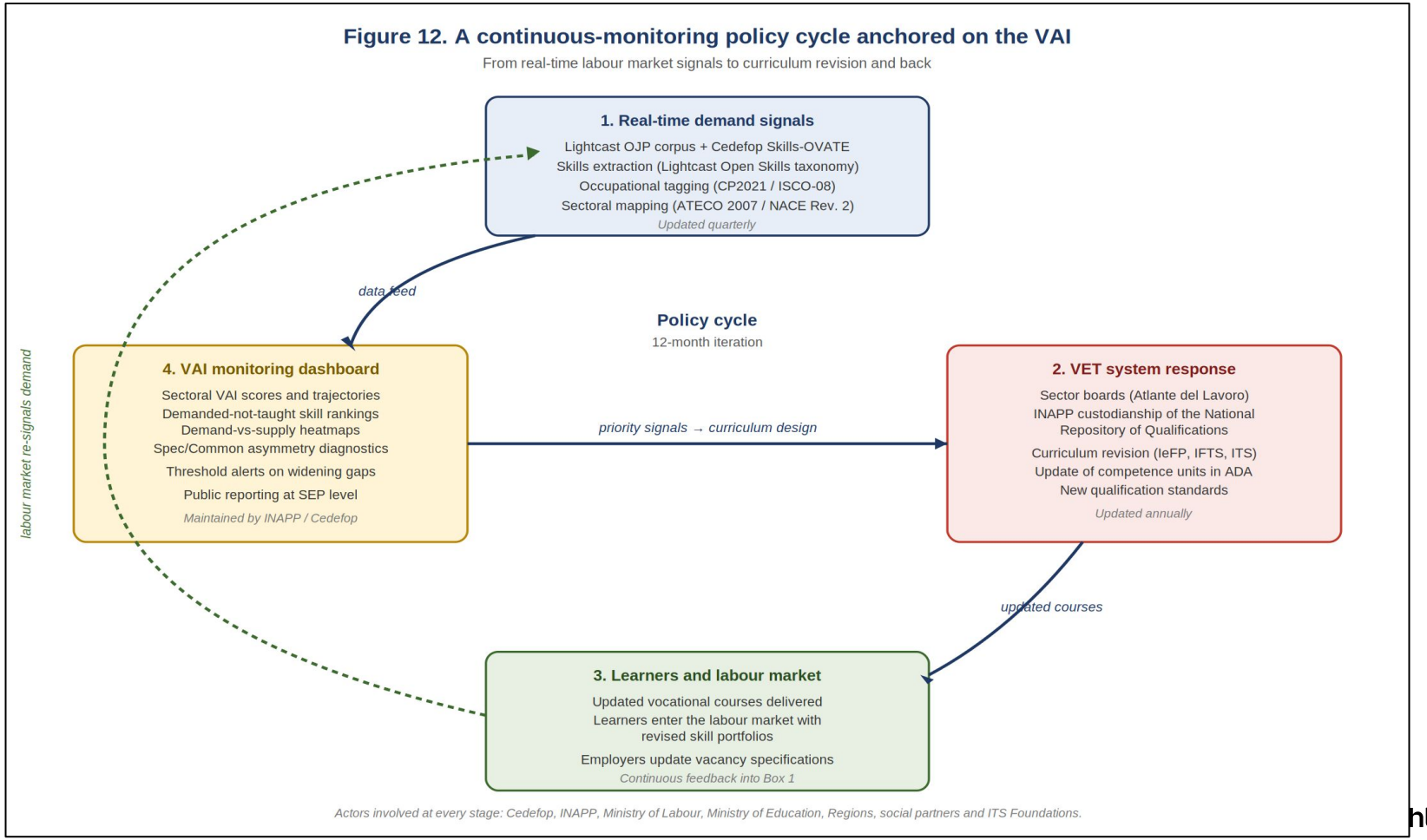
QUADRANT D

Transversal-aligned, technical-weak

Personal services, Social-health, Education · refresh technical content; preserve transversal.

Lightcast

A continuous-monitoring policy cycle



Thank you

LIGHTCAST.IO

Mauro Pelucchi
Head of Global Data Science
mauro.pelucchi@lightcast.io