Skills and AI Innovative Ecosystems: Evidence from Brazilian Microregions

Rafael de la Vega[*] Florencia Jaccoud[*] María de las Mercedes Menéndez[*]
Cecilia Seri[*] Tommaso Ciarli[*†]

[*]UNU-MERIT, United Nations University
[†]SPRU, University of Sussex

ESCoE Conference on Economic Measurement 2026
King's Business School, London, 19 May 2026

Motivation

- **Start-ups** pioneer the application of emerging technologies (Tushman and Anderson, 1986; Kolev et al., 2022), including AI (D'Alessandro et al., 2025).
- Require and generate new **competences** (Cai et al., 2025; Szulc-Wódarska and Kowalski, 2024): firms adjust **skill requirements** to leverage technological opportunities (Ciarli et al., 2021)
- Jobs consist of bundles of tasks carried out through **bundles of skills** by workers: regional bundles of skills (Neffke, 2019; Hosseinioun et al., 2025)
- **Regions** diversify into related products and technologies (Neffke et al., 2011; Boschma, 2017), building on existing knowledge and skills (Colombelli et al., 2023; Igna and Venturini, 2023; Xiao and Boschma, 2023): **path-dependence**
- **Technological paradigms shifts** open **windows of opportunities** that may enable lagging economies to catch-up and leapfrog (Perez, 2010)
  - Brazil emerging as AI leader in Latin America ▸ Patents ▸ Start-ups

This Paper

Research Question: Which skills are most strongly associated with the presence of AI start-ups across Brazilian microregions?

What we do / Contribution:

- Build a **novel region–year–skill dataset** linking Brazilian labor (RAIS) and occupation (CBO) data with ESCO
- Identify and map **AI** and **digital start-ups** across Brazilian microregions in 2006-21
- Use **Canonical Correspondence Analysis (CCA)** to reduce the regional skills dimensionality to AI startups

Contribution:

- Map **context-specific capabilities** relevant to AI and digital entrepreneurship in Brazil (beyond "AI skills")

Main Findings:

- Data architecture and ICT management most stably predict regional AI startups
- Large group of AI and advanced digital skills co-locate: path dependence
- AI-relevant skills are geographically rare and concentrate in a small number of highly diversified region

This Paper

Research Question: Which skills are most strongly associated with the presence of AI start-ups across Brazilian microregions?

What we do / Contribution:

- Build a **novel region–year–skill dataset** linking Brazilian labor (RAIS) and occupation (CBO) data with ESCO
- Identify and map **AI** and **digital start-ups** across Brazilian microregions in 2006-21
- Use **Canonical Correspondence Analysis (CCA)** to reduce the regional skills dimensionality to AI startups

Contribution:

- Map **context-specific capabilities** relevant to AI and digital entrepreneurship in Brazil (beyond "AI skills")

Main Findings:

- Data architecture and ICT management most stably predict regional AI startups
- Large group of AI and advanced digital skills co-locate: path dependence
- AI-relevant skills are geographically rare and concentrate in a small number of highly diversified region

This Paper

Research Question: Which skills are most strongly associated with the presence of AI start-ups across Brazilian microregions?

What we do / Contribution:

- Build a **novel region–year–skill dataset** linking Brazilian labor (RAIS) and occupation (CBO) data with ESCO
- Identify and map **AI** and **digital start-ups** across Brazilian microregions in 2006-21
- Use **Canonical Correspondence Analysis (CCA)** to reduce the regional skills dimensionality to AI startups

Contribution:

- Map **context-specific capabilities** relevant to AI and digital entrepreneurship in Brazil (beyond "AI skills")

Main Findings:

- Data architecture and ICT management most stably predict regional AI startups
- Large group of AI and advanced digital skills co-locate: path dependence
- AI-relevant skills are geographically rare and concentrate in a small number of highly diversified region

Related Literature

Start-ups and Regional Specialization

Start-ups as drivers of innovation

- Growing evidence: Start-ups deliver innovative solutions across industrial ecosystems — in both advanced and developing economies (IDB, 2022; Dechezleprêtre et al., 2023; European Commission, 2024; Mageste et al., 2024)
- Skills are critical for AI start-up creation (Colombelli et al., 2024; Szulc-Wódarska and Kowalski, 2024)

Regional capabilities as enablers of diversity

- New activities emerge from what regions already do well (Isaksen, 2014; Neffke et al., 2011; Tanner, 2016)
  - **Path dependence**: Existing knowledge shapes opportunities and directions of recombination (Schumpeter, 1934)
  - **Knowledge diffusion**: Proximity matters — technological, geographical, and institutional (Boschma, 2017)
- In Brazil, regions diversify into sectors matching local skills (Françoso et al., 2024; Galetti et al., 2021)

Related Literature

Start-ups and Regional Specialization

Start-ups as drivers of innovation

- Growing evidence: Start-ups deliver innovative solutions across industrial ecosystems — in both advanced and developing economies (IDB, 2022; Dechezleprêtre et al., 2023; European Commission, 2024; Mageste et al., 2024)
- Skills are critical for AI start-up creation (Colombelli et al., 2024; Szulc-Wódarska and Kowalski, 2024)

Regional capabilities as enablers of diversity

- New activities emerge from what regions already do well (Isaksen, 2014; Neffke et al., 2011; Tanner, 2016)
  - **Path dependence**: Existing knowledge shapes opportunities and directions of recombination (Schumpeter, 1934)
  - **Knowledge diffusion**: Proximity matters — technological, geographical, and institutional (Boschma, 2017)
- In Brazil, regions diversify into sectors matching local skills (Françoso et al., 2024; Galetti et al., 2021)

Data
Skills Specialization by Region

Unit of analysis: 510 Regiões Geográficas Imediatas (RGIs) of Brazil

Building regional specialisations in skills:

1. **RAIS** (Relação Anual de Informações Sociais): Brazilian formal labour market data
   - Covers $\sim$45MM workers per year (2006–2021) $\Rightarrow$ occupation intensity (FTE) of regions
2. **ESCO** (European Skills, Competences, Qualifications and Occupations)
   - Maps 13,939 skills to 3,039 occupations $\Rightarrow$ skill intensity of occupations

- Crosswalks provided by the Brazilian Ministry of Labour and ILO (CBO-02 $\rightarrow$ ISCO-88 $\rightarrow$ ISCO-08)
- Combining RAIS and ESCO: skill intensity of regions
- **Output:** skill specialisation (LQ, RCA) for each RGI–skill pair (2006–2021)

Specialisation eqs.

Data

Skills Specialization by Region

Unit of analysis: 510 Regiões Geográficas Imediatas (RGIs) of Brazil

Building regional specialisations in skills:

1. **RAIS** (Relação Anual de Informações Sociais): Brazilian formal labour market data
   - Covers ∼45MM workers per year (2006–2021) ⇒ occupation intensity (FTE) of regions
2. **ESCO** (European Skills, Competences, Qualifications and Occupations)
   - Maps 13,939 skills to 3,039 occupations ⇒ skill intensity of occupations

- Crosswalks provided by the Brazilian Ministry of Labour and ILO (CBO-02 ⟶ ISCO-88 ⟶ ISCO-08)
- Combining RAIS and ESCO: skill intensity of regions
- **Output:** skill specialisation (LQ, RCA) for each RGI–skill pair (2006–2021)

Specialisation eqs.

Data
Skills Specialization by Region

Unit of analysis: 510 Regiões Geográficas Imediatas (RGIs) of Brazil

Building regional specialisations in skills:

1. **RAIS** (Relação Anual de Informações Sociais): Brazilian formal labour market data
   - Covers $\sim$45MM workers per year (2006–2021) $\Rightarrow$ occupation intensity (FTE) of regions
2. **ESCO** (European Skills, Competences, Qualifications and Occupations)
   - Maps 13,939 skills to 3,039 occupations $\Rightarrow$ skill intensity of occupations

- Crosswalks provided by the Brazilian Ministry of Labour and ILO (CBO-02 $\longrightarrow$ ISCO-88 $\longrightarrow$ ISCO-08)
- Combining RAIS and ESCO: skill intensity of regions
- **Output:** skill specialisation (LQ, RCA) for each RGI–skill pair (2006–2021)

Specialisation eqs.

Data
AI and Digital Firms

3. **FactSet / Crunchbase**: AI and digital firms
   - Firm info: name, year of incorporation, location (city), industry
   - Crunchbase-specific tags that categorize firms according to their core activities

List of tags

     - **AI firms:** Artificial Intelligence, Generative AI, Intelligent Systems, Machine Learning, Natural Language Processing, Predictive Analytics, and Robotic Process Automation (RPA)
     - **Digital firms:** producers of digital goods (software, cloud, data infrastructure) (strict) and core-digital + platforms and digitally-enabled industries (broad)
     - **Coverage:** 25,684 firms in total, including 609 AI firms, 2,937 digital (strict), and 7,002 digital (broad)
   - Cleaning of location data and assignment of firms to RGIs $\longrightarrow$ 392 RGIs with firms retained, 75 have $\geq$ 1 AI firm (19%)

Final Dataset

Data on skills specialization and presence of AI and digital firms in each of the 392 RGIs over 2006–2021

Representativeness I | Representativeness II | Start-up trends

Data

AI and Digital Firms

3. **FactSet / Crunchbase**: AI and digital firms
   - Firm info: name, year of incorporation, location (city), industry
   - Crunchbase-specific tags that categorize firms according to their core activities

     List of tags

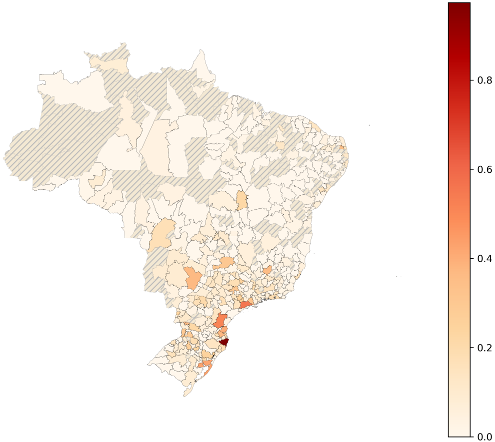
     - **AI firms:** Artificial Intelligence, Generative AI, Intelligent Systems, Machine Learning, Natural Language Processing, Predictive Analytics, and Robotic Process Automation (RPA)
     - **Digital firms:** producers of digital goods (software, cloud, data infrastructure) (strict) and core-digital + platforms and digitally-enabled industries (broad)
     - **Coverage:** 25,684 firms in total, including 609 AI firms, 2,937 digital (strict), and 7,002 digital (broad)
   - Cleaning of location data and assignment of firms to RGIs $\longrightarrow$ 392 RGIs with firms retained, 75 have ≥ 1 AI firm (19%)

Final Dataset

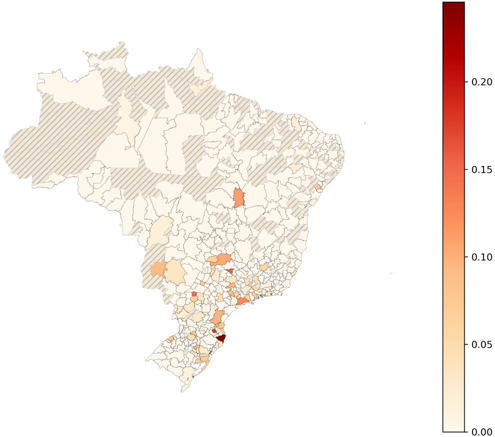
Data on skills specialization and presence of AI and digital firms in each of the 392 RGIs over 2006–2021

Representativeness I Representativeness II Start-up trends

Regional distribution of AI and digital firms relative to population



Digital (strict) firms per 10,000 people



AI firms per 10,000 people

Source: Own elaboration based on FactSet/Crunchbase

Correlation stocks
Correlation stocks pc

Methodology

Canonical Correspondence Analysis (CCA):

- Originally developed in quantitative ecology, recently introduced in economics (Nomaler and Verspagen, 2024)
- **Method of dimensionality reduction** of a two-way matrix through the identification of the principal eigenvectors, while **maximising correlation with pre-defined variables** ▸ Intuition
  - CCA extends correspondence analysis (CA) – the statistical foundation underlying the economic complexity framework – by incorporating external explanatory variables into the ordination
  - the extracted axes are chosen to maximally correlate with a set of pre-defined variable

This paper's application: Relate RCA skill profiles of Brazilian RGIs with other regional characteristics of interest

- Start from the 392 × 13,460 binary skill specialisation matrix
- Extract orthogonal canonical dimensions: each related to one explanatory variable and representing a distinct gradient along which regional skill specialisation varies
- First dimensions: regional GDP pc
- Other dimensions variables and models:
  - baseline: stock of **AI firms pc** (**M1**) (2021)
  - **M2**: add stock of **digital firms pc** (2021)
  - **M3**: add **innovation ecosystem** controls (M3)
  - M4: add stock of digital firms (2015)

Methodology

Canonical Correspondence Analysis (CCA):

- Originally developed in quantitative ecology, recently introduced in economics (Nomaler and Verspagen, 2024)
- **Method of dimensionality reduction** of a two-way matrix through the identification of the principal eigenvectors, while **maximising correlation with pre-defined variables** Intuition
  - CCA extends correspondence analysis (CA) – the statistical foundation underlying the economic complexity framework – by incorporating external explanatory variables into the ordination
  - the extracted axes are chosen to maximally correlate with a set of pre-defined variable

This paper's application: Relate RCA skill profiles of Brazilian RGIs with other regional characteristics of interest

- Start from the 392 × 13,460 binary skill specialisation matrix
- Extract orthogonal canonical dimensions: each related to one explanatory variable and representing a distinct gradient along which regional skill specialisation varies
- First dimensions: regional GDP pc
- Other dimensions variables and models:
  - baseline: stock of **AI firms pc** (**M1**) (2021)
  - **M2**: add stock of **digital firms pc** (2021)
  - **M3**: add **innovation ecosystem** controls (M3)
  - M4: add stock of digital firms (2015)

CCA Results

CCA1 and CCA2 variable loadings and share of inertia across different models

	M1 (baseline)		M2 (+ digital)		M3 (full controls)		M4 (lagged)	
Variable	CCA1	CCA2	CCA1	CCA2	CCA1	CCA2	CCA1	CCA2
GDP per capita	0.805	−0.050	0.803	−0.097	0.783	0.043	0.820	−0.060
AI firms per capita	0.278	0.454	0.293	0.434	0.247	0.428	0.290	0.320
Strict digital firms	—	—	0.406	0.363	—	—	0.420	0.310
Incubators/tech parks	—	—	—	—	0.181	0.310	—	—
Bank agencies	—	—	—	—	0.624	−0.179	—	—
STEM programs	—	—	—	—	0.055	0.493	—	—
Datacenters	—	—	—	—	0.135	0.478	—	—
GDP growth	—	—	—	—	0.131	−0.236	—	—
Share of inertia	0.086	0.020	0.088	0.023	0.094	0.041	0.090	0.020

Axis loadings: association between each extracted dimension and the regional variables

Share of inertia: how much of the total variation in the skill specialisation matrix each dimension accounts for

What Skills Define the AI Gradient?

Skill rankings emerging from CCA2 scores, comparing across model specifications:

- **Group 1 — The stable core** (8 skills): data architecture and ICT management competencies, stable across all specifications
  - Foundational prerequisites always present where AI firms concentrate
  - Many also rank near the top of CCA1 ⇒ **underpin both general economic development and AI-specific activity**

Group 1 skills

- **Group 2 — AI and digital technical skills** (142 skills)
  - Rank 7th on CCA2, but below 54th on CCA1 ⇒ **more closely associated with AI than with income levels**
  - Spans system design, cloud computing, DevOps, cybersecurity, blockchain, and specialist programming frameworks
  - Association with AI remains strong when digital firms are introduced (7th in M2, 8th in M4)

Group 2 skills

- **Group 3 — AI regulation**
  - Regulatory (data ethics, GDPR, security) and software configuration management skills even more distinctive when including innovation system gradients in M3 (STEP, data centres, incubators)

- **AI and digital skill profiles not separable**

  ⇒ AI start-ups skill specialisation is related to digital start-ups skills (similar loadings/vectors)

Spearman correlation

What Skills Define the AI Gradient?

Skill rankings emerging from CCA2 scores, comparing across model specifications:

- **Group 1 — The stable core** (8 skills): data architecture and ICT management competencies, stable across all specifications
  - Foundational prerequisites always present where AI firms concentrate
  - Many also rank near the top of CCA1 ⇒ **underpin both general economic development and AI-specific activity**

Group 1 skills

- **Group 2 — AI and digital technical skills** (142 skills)
  - Rank 7th on CCA2, but below 54th on CCA1 ⇒ **more closely associated with AI than with income levels**
  - Spans system design, cloud computing, DevOps, cybersecurity, blockchain, and specialist programming frameworks
  - Association with AI remains strong when digital firms are introduced (7th in M2, 8th in M4)

Group 2 skills

- **Group 3 — AI regulation**
  - Regulatory (data ethics, GDPR, security) and software configuration management skills even more distinctive when including innovation system gradients in M3 (STEP, data centres, incubators)

- **AI and digital skill profiles not separable**

  ⇒ AI start-ups skill specialisation is related to digital start-ups skills (similar loadings/vectors)

Spearman correlation

What Skills Define the AI Gradient?

Skill rankings emerging from CCA2 scores, comparing across model specifications:

- **Group 1 — The stable core** (8 skills): data architecture and ICT management competencies, stable across all specifications
  - Foundational prerequisites always present where AI firms concentrate
  - Many also rank near the top of CCA1 $\Rightarrow$ **underpin both general economic development and AI-specific activity**

Group 1 skills

- **Group 2 — AI and digital technical skills** (142 skills)
  - Rank 7th on CCA2, but below 54th on CCA1 $\Rightarrow$ **more closely associated with AI than with income levels**
  - Spans system design, cloud computing, DevOps, cybersecurity, blockchain, and specialist programming frameworks
  - Association with AI remains strong when digital firms are introduced (7th in M2, 8th in M4)

Group 2 skills

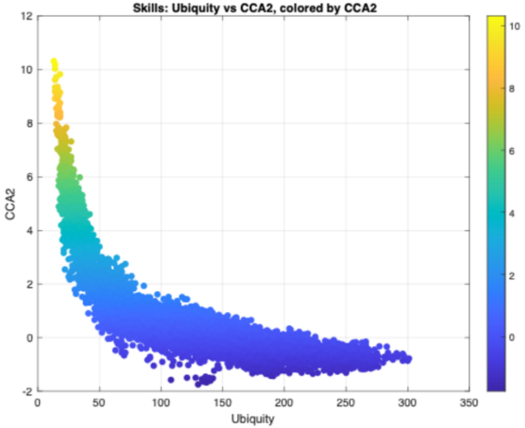
- **Group 3 — AI regulation**
  - Regulatory (data ethics, GDPR, security) and software configuration management skills even more distinctive when including innovation system gradients in M3 (STEP, data centres, incubators)
- **AI and digital skill profiles not separable**

  $\Rightarrow$ AI start-ups skill specialisation is related to digital start-ups skills (similar loadings/vectors)

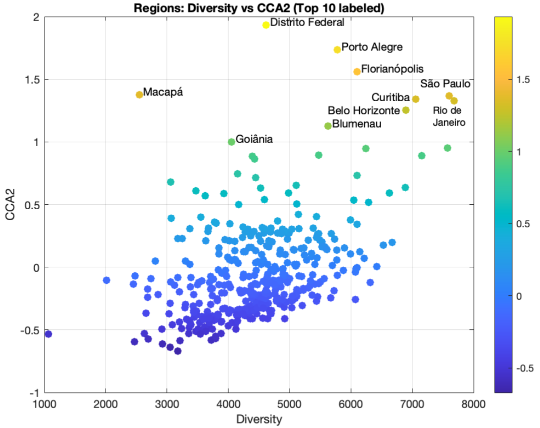
Spearman correlation

The "sophistication" of AI Skills

Top AI-start-ups skills concentrate in few, highly diversified, regions ⇒ sophisticated skills found in regions that are able to produce most products and services



Notes: Skills ranked by CCA2 score (M1) against their ubiquity



Notes: Regions ranked by CCA2 score (M1) against their diversity

Role of Pre Existing Digital Capabilities on AI Start-ups

Estimate the probability that AI firms emerge in regions with high specialisation of digital-AI skills in $t-1$ (LQ of top 5% skills on the digital-AI gradient in 2015)

Stock of AI firms – Poisson Pseudo-Maximum Likelihood (PPML):

$$E[AI_{rt} \mid CoreSkill_{r,t-1}, \alpha_r, \gamma_t] = \exp\left(\beta\, CoreSkill_{r,t-1} + \alpha_r + \gamma_t\right) \times pop_{rt}$$

where

- AI_{rt} is the stock of AI firms in region r, year t;
- $CoreSkill_{r,t-1}$ is the regional specialisation in core skills in $t-1$;
- pop_{rt}: population; α_r: region FE; γ_t: year FE.

Entry of new AI firms (LPM):

$$has\_new\_ai_{rt} = \beta\, CoreSkill_{r,t-1} + \alpha_r + \gamma_t + \varepsilon_{rt}$$

where

- $has\_new\_ai_{rt}$ = 1 if at least one new AI firm appears in region r, year t (three-year moving average); 0 otherwise.

Pre-existing Capabilities and AI Start-ups

- 10 additional digital-AI skills in $t-1$ associated with 2.2% more AI start-ups in t (average region specializes in ∼37 core skills).
- 1-SD increase in LQ associated with $\approx 64\%$ more AI start-ups.

Regional specialisation in core skills and subsequent AI firm activity

	AI Firm Stocks (PPML)		**AI Firm Flows (Extensive Margin)**	
	(1)	(2)	(1)	(2)
Core skill spec (count) (t-1)	0.0022**		0.0008***	
	(0.0010)		(0.0002)	
Core skill spec (avg LQ) (t-1)		1.5872**		0.0280
		(0.6737)		(0.0233)
Mean dep. var.	0.612	0.612	0.067	0.067
Region FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Population exposure	Yes	Yes	—	—
Clusters	74	74	392	392
Observations	1110	1110	5880	5880

Notes: Stocks: PPML with region and year FE, population exposure.
Flows: LPM with region and year FE.

Standard errors clustered at region level. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Core skills distribution

Robustness Checks

Alternative AI/Digital classifications:

Results

- Exclude RPA-only firms from AI category (20 firms reassigned)
  - Loading structure virtually unchanged: AI firms per capita remains the anchor of CCA2
- Broad definition of digital firms (7,002 vs 2,937 strict)
  - AI loading on CCA2 stable at 0.43; digital firms load on both CCA1 and CCA2

Permutation tests:

Results

- Evaluate CCA results against a null of no association: randomly reshuffle the rows of the regional variables matrix (AI firms pc, GDP pc) 1,000 times and compare observed results to this random baseline

  $\Rightarrow$ skill ordering obtained is statistically different from random ($p = 0.001$)
  $\Rightarrow$ variable loadings on each CCA dimension are statistically different from random
  $\Rightarrow$ share of inertia explained by the actual model is statistically higher than random

Conclusions

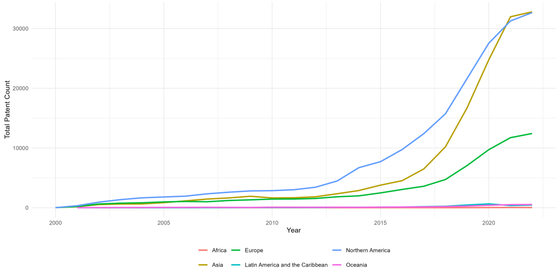
- We adopt a **data-driven approach** to identify the **skills most closely associated with** the presence of **AI start-ups** in Brazilian microregions
- **Two groups of AI-relevant skills** emerge: a stable core of **data architecture and ICT management** competencies, and a broader cluster of **AI related digital skills** (cloud, DevOps, security, blockchain)
- When including AI innovation system, **AI regulation and software configuration management** skills are most distinctive
- AI-relevant skills are **geographically rare** and concentrate in a small number of **highly diversified regions** — consistent with the nestedness principle from economic complexity
- **Prior specialisation** in core skills predicts both a **larger stock** of AI firms and a higher **probability of new entry**
- **AI and broader digital skill overlap** $\Rightarrow$ AI-related specialisation is embedded within the wider digital economy
- **Policy implication:** strengthening the underlying **digital capability base** — ICT skills, digital infrastructure, technology management — may be essential for spurring AI investment

Thank you for your attention

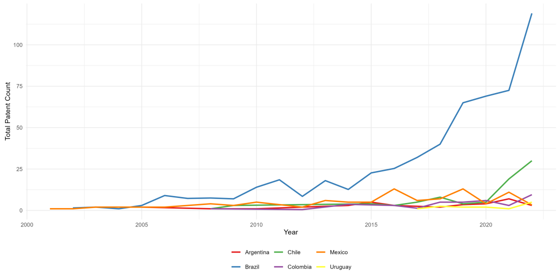
Comments welcome

ciarli@merit.unu.edu

Brazil emerges as a regional leader: patenting ▶ Back



Global regions

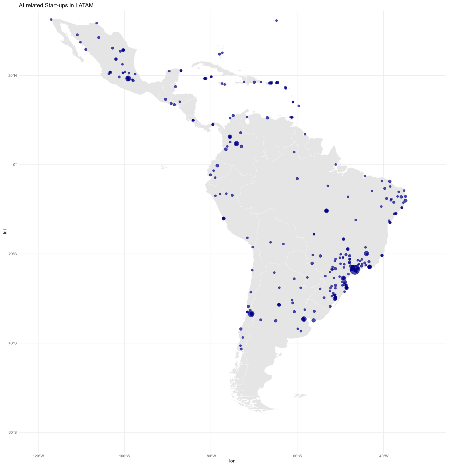


LAC countries

Source: Own elaboration based on PATSTAT and WIPO

Motivation

Brazil emerges as a regional leader: patenting ▸ Back



Source: Own elaboration based on FactSet/Crunchbase

List of industries included in AI and Digital Firms groups

AI Tags	Digital Strict Tags	Additional Broad Digital Tags
Artificial Intelligence (AI)	3D Technology	Ad Exchange
Generative AI	A/B Testing	Ad Network
Intelligent Systems	Apps	Ad Retargeting
Machine Learning	App Discovery	Ad Server
Natural Language Processing	Consumer Applications	Ad Targeting
Predictive Analytics	Enterprise Applications	Advertising Platforms
Robotic Process Automation (RPA)	Mobile Apps	Affiliate Marketing
	Web Apps	Mobile Advertising
	Cloud Computing	Native Advertising
	Cloud Data Services	SEM
	Cloud Management	Social Media Advertising
	Cloud Security	Video Advertising
	Data and Analytics	Digital Marketing
	Analytics	Digital Signage
	Application Performance Mgmt.	Email Marketing
	Big Data	Lead Generation
	Business Intelligence	Marketing Automation
	Data Collection and Labeling	Social CRM
	Data Governance	Social Media Management
	Data Integration	Social Media / Social Networks
	Data Management	E-Commerce / Marketplaces
	Data Mining	Retail Technology
	Data Visualization	Online Auctions, Coupons
	Database	Price Comparison, Group Buying
	Facial Recognition	Digital Media
	Image Recognition	Video Streaming / VoD
	Speech Recognition	Blogging, Podcasts, Internet Radio
	Text Analytics	Messaging, Video Chat, Conferencing
	SaaS	Payments, FinTech, Crypto
	DevOps	InsurTech, Online Payments
	IT Infrastructure	
	IT Management	
	Virtualization	

Go back

Core AI-Startup Skills (Group 1)

Skill	M1	M2	M3	M4
Create data sets	1	4	1	4
Identify processes for re-engineering	2	6	11	6
Evaluate cost of software products	2	24	11	24
Review development process of an organisation	2	6	11	6
Information architecture	3	15	5	15
Implement a management system	4	31	10	31
Provide user documentation	5	3	9	3
Develop information standards	6	14	22	14

Notes: Rank on CCA2 dimension across model specifications.

Go back

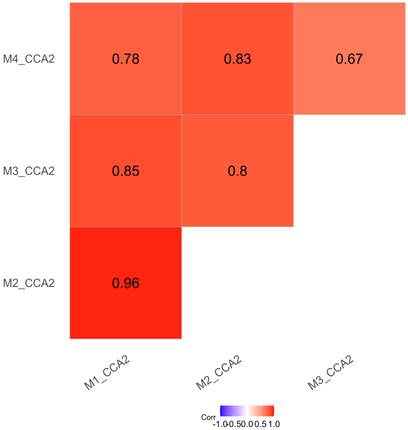
AI-Digital Technical Skills (Group 2)

Category	Representative skills
System & software architecture	System design; enterprise architecture; database physical structure
Distributed & cloud computing	Distributed computing; manage ICT virtualisation; deploy cloud resource
DevOps tooling	Chef; Octopus Deploy; automated software tests; ICT system integration
Security & forensics	Cybersecurity incidents; Wireshark; Metasploit; computer forensics
Blockchain & distributed ledger	Distributed ledger technology; smart contract; Solidity
Programming & data platforms	SPARK; Angular; SQL; NoSQL; data warehouse; deep learning

Notes: All 142 skills receive identical CCA2 scores as they are co-present and co-absent in exactly the same set of regions.

Go back

Spearman Correlation of CCA2 Across Model Specifications



Notes: Spearman rank correlation of CCA2 skill scores across the four model specifications, restricted to the top 5% of skills by M1 CCA2 score.

Go back

Permutation Tests

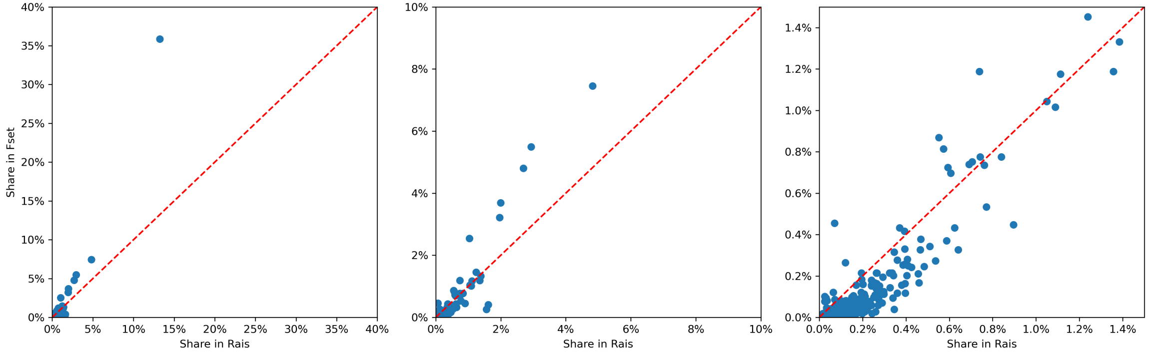
Permutation Tests Results

	M1	*p-val*	M2	*p-val*
Test 1				
Sum of eigenvalues	0.204	0.001	0.222	0.001
Test 2				
Dimension 1	0.167	0.001	0.171	0.001
Dimension 2	0.038	0.001	0.045	0.001
Test 3				
GDP per capita – Dimension 1	0.805	0.001	0.803	0.001
GDP per capita – Dimension 2	-0.050	0.909	-0.097	0.672
AI firms – Dimension 1	0.278	0.112	0.293	0.015
AI firms – Dimension 2	0.454	0.001	0.434	0.001
Strict digital firms – Dimension 1			0.406	0.001
Strict digital firms – Dimension 2			0.363	0.015

Notes: Permutation *p*-values based on 1,000 permutations. The global statistic equals the sum of canonical eigenvalues. Test 3 is two-sided.

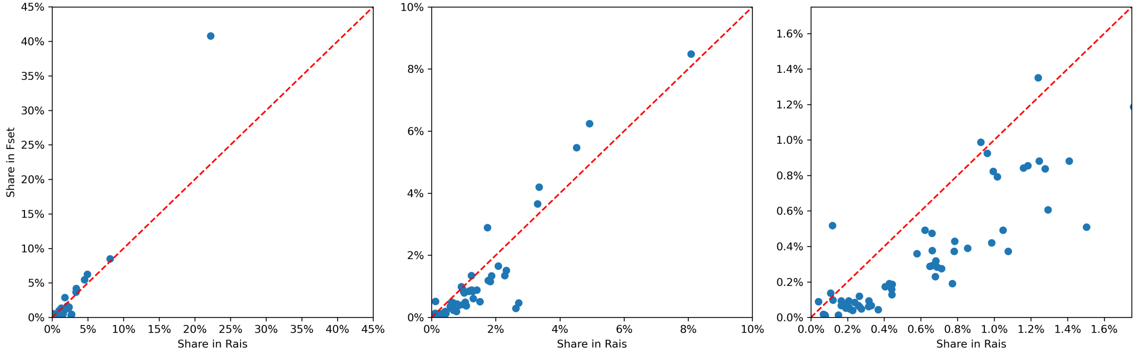
Go back

Distribution of Firms FactSet vs RAIS



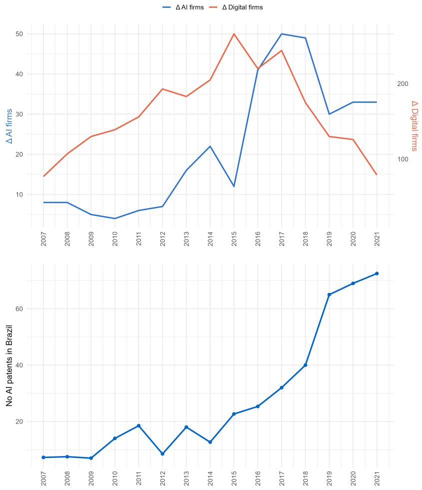
Go back

Distribution of Firms FactSet vs RAIS (RGIs With 10 or More Firms)



Go back

Trends in new digital and AI firms (strict)



Notes: Source: Own elaboration based on FactSet, Patstat, ORBIS-IP and (WIPO, 2022).

Go back

What is Canonical Correspondence Analysis (CCA)?

Analogy: constrained PCA

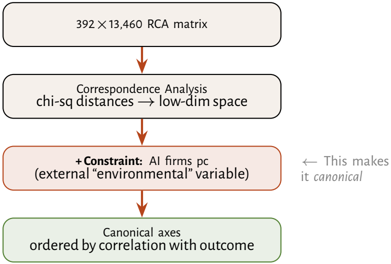
PCA finds axes that maximise *variance* in the skill matrix – atheoretical.

CCA maximises variance *while staying maximally correlated with the outcome* (AI firms pc).

$\Rightarrow$ **regression + factor model simultaneously** — the data finds the skill bundle most correlated with AI startups.

Ecological roots, economic application: CCA finds which *species* inhabit which *environments*. Here: which *skill species* inhabit AI start-up ecosystems.

Step by step



Output: skill & region scores on canonical axes ⟶ rank skills by AI start-up association, net of GDP per capita.

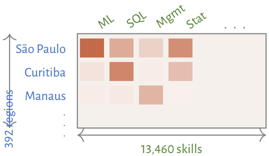
What is Canonical Correspondence Analysis (CCA)?

The setting

13,460 skills measured across 392 regions.

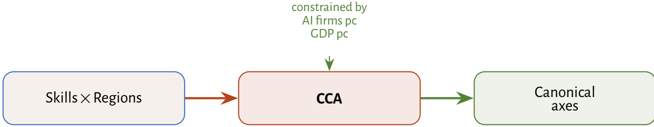
Which skill bundle predicts where AI start-ups locate?

The input: a region × skill matrix

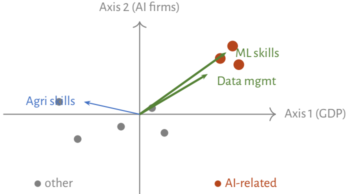


Cell = RCA (Revealed Comparative Advantage) specialisation index

Step 2 — Constrain by the outcome



Step 3 – Canonical axes

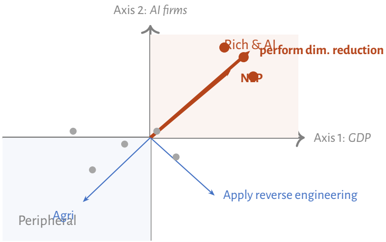


Arrow = skill; **direction** = assoc. with AI firms / GDP pc

Reading CCA Output

Back

The biplot — annotated



dir. = assoc. with axis **length** $\propto$ strength

What CCA tells us

- ✓ **Dimensionality reduction guided by theory:** compresses 13,460 skills into interpretable axes, like PCA but anchored to the outcome.
- ✓ **Simultaneous skill + region map:** region near a skill arrow $\Rightarrow$ skill abundant there *and* co-occurring with AI firms.
- ✓ **Multiple constraints:** each variable (GDP, digital firms, ...) adds an orthogonal canonical axis — models are comparable.

Interpretation

CCA reveals **association**, not causation. Read as: *"skills co-locating with AI start-ups after controlling for regional development."*

Method

Skill Intensity & Specialization

- **Step 1: Occupation-level skill intensity** (following Caravella et al. (2023), Santoalha et al. (2021))

$$X_{so} = \frac{\sum_g A_{og} X_{sg}}{\sum_g \sum_{s'} X_{s'g}}$$

where $X_{sg} = 1$ if occupation g requires skill s; $A_{og} = 1$ if occupation g belongs to occupation group o

- **Step 2: Region–skill intensity (RAIS $\times$ ESCO)**

$$X_{srt} = \sum_o X_{so} L_{ort}$$

where L_{ort} = FTE employment in occupation o, region r, year t

- **Step 3: Aggregation**

$$X_{rt} = \sum_s X_{srt}, \qquad X_{st} = \sum_r X_{srt}, \qquad X_t = \sum_s X_{st}$$

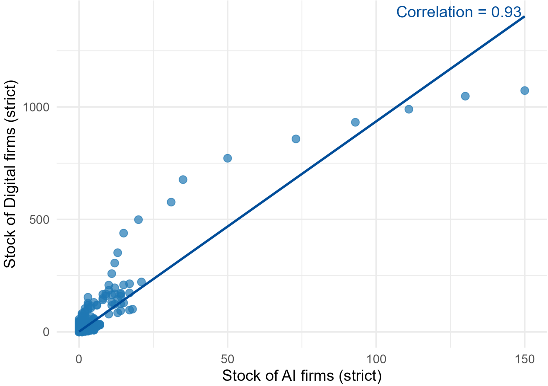
Go back

- **Step 4: Skill specialization (Location Quotient)**

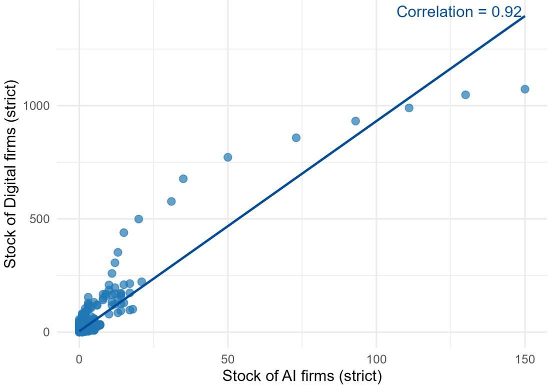
$$LQ_{srt} = \frac{\left(\frac{X_{srt}}{X_{rt}}\right)}{\left(\frac{X_{st}}{X_t}\right)}$$

Correlation between stock of digital and AI firms

Go back



Full sample

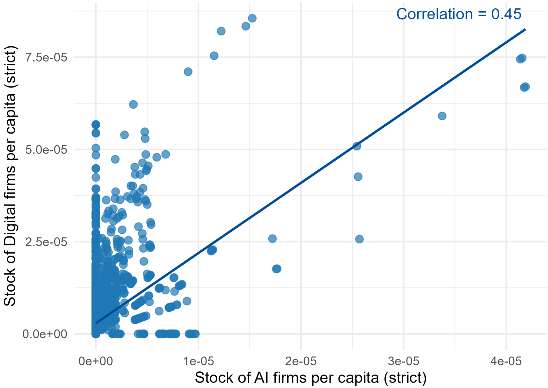


Restricted sample: regions with at least one AI firm

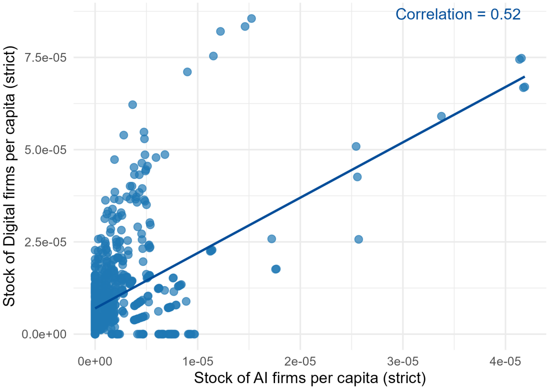
Notes: The left panel shows the correlation between the stock of digital and AI firms (strict definition) for the full sample. The right panel restricts the sample to regions with at least one AI firm between 2006 and 2021.

Correlation between stock of digital and AI firms per capita

Go back



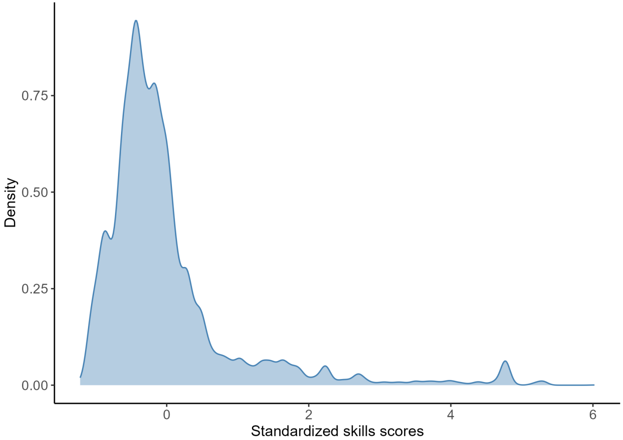
Full sample



Restricted sample: regions with at least one AI firm

Notes: The left panel shows the correlation between the stock of digital and AI firms per capita (strict definition) for the full sample. The right panel restricts the sample to regions with at least one AI firm between 2006 and 2021.

Distribution of Core Skills (CCA2 Scores)



Notes: Kernel density of skill scores along CCA2 of Model 4, after *z*-standardisation.

Go back

Robustness: Alternative Firms Classifications

CCA variable loadings and share of inertia, alternative specifications

Panel A: Excluding RPA-only firms

	M1 (baseline)		**M1 (no RPA)**	
Variable	CCA1	CCA2	CCA1	CCA2
GDP per capita	0.81	−0.05	0.81	−0.05
AI firms per capita	0.28	0.45	0.27	0.45
Share of inertia	0.09	0.02	0.09	0.02

Panel B: Broad definition of digital firms

	M2 (strict digital)		**M2 (broad digital)**	
Variable	CCA1	CCA2	CCA1	CCA2
GDP per capita	0.80	−0.10	0.80	−0.11
AI firms per capita	0.29	0.43	0.29	0.43
Digital firms per capita	0.41	0.36	0.46	0.34
Share of inertia	0.09	0.02	0.09	0.02

Go back

References I

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, 51(3):351–364.

Cai, J. J., Gu, X., Sheng, L., Xia, M., Zhao, L., and Zhu, W. (2025). Ai as "co-founder": Genai for entrepreneurship. Working Paper, arXiv:2512.06506.

Caravella, S., Cirillo, V., Crespi, F., Guarascio, D., and Menghini, M. (2023). The diffusion of digital skills across EU regions: structural drivers and polarisation dynamics. *Regional Studies*, 10(1):820–844.

Ciarli, T., Kenney, M., Massini, S., and Piscitello, L. (2021). Digital technologies, innovation, and skills: Emerging trajectories and challenges. *Research Policy*, 50(7):104289.

Colombelli, A., D'Amico, E., and Paolucci, E. (2023). When computer science is not enough: universities knowledge specializations behind artificial intelligence startups in Italy. *Journal of Technology Transfer*, 48(5):1599–1627.

Colombelli, A., Paolucci, E., Raguseo, E., and Elia, G. (2024). The creation of digital innovative start-ups: the role of digital knowledge spillovers and digital skill endowment. *Small Business Economics*, 62(3):917–937.

D'Alessandro, F., Santarelli, E., and Vivarelli, M. (2025). The KSTE + I approach and the advent of AI technologies: evidence from the European regions. *Journal of Technology Transfer*, 51(1):414–451.

Dechezleprêtre, A., Díaz, L., Fadic, M., and Lalanne, G. (2023). How the green and digital transitions are reshaping the automotive ecosystem. OECD Science, Technology and Industry Policy Papers 133, Organization for Economic Cooperation and Development (OECD).

References II

European Commission (2024). Monitoring the twin transition of industrial ecosystems: Electronics – Analytical report. Technical report, European Commission; Fraunhofer ISI; Technopolis Group.

Françoso, M. S., Boschma, R., and Vonortas, N. (2024). Regional diversification in brazil: the role of relatedness and complexity. *Growth and Change*, 55(1):e12702.

Galetti, J. R. B., Tessarin, M. S., and Morceiro, P. C. (2021). Skill relatedness, structural change and heterogeneous regions: evidence from a developing country. *Papers in Regional Science*, 100(6):1355–1377.

Hosseinioun, M., Neffke, F., Zhang, L., and Youn, H. (2025). Skill dependencies uncover nested human capital. *Nature Human Behaviour*, pages 1–15.

IDB (2022). Fintech in latin america and the caribbean: A consolidated ecosystem for recovery. Technical report, Inter-American Development Bank.

Igna, I. and Venturini, F. (2023). The determinants of AI innovation across European firms. *Research Policy*, 52(2):104661.

Isaksen, A. (2014). Industrial development in thin regions: trapped in path extension? *Journal of economic geography*, 15(3):585–600.

Kolev, J., Haughey, A., Murray, F., and Stern, S. (2022). Of Academics and Creative Destruction: Startup Advantage in the Process of Innovation. National Bureau of Economic Research.

References III

Mageste, S., Plottier, C., Rocha, C., and Saporito, N. (2024). Empresas emergentes (start-ups) en américa latina y el caribe: una primera aproximación a su identificación y características. Documentos de proyectos 68855, Naciones Unidas, Comisión Económica para América Latina y el Caribe (CEPAL).

Neffke, F., Henning, M., and Boschma, R. (2011). How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic geography*, 87(3):237–265.

Neffke, F. M. (2019). The value of complementary co-workers. *Science advances*, 5(12):eaax3370.

Nomaler, M. O. and Verspagen, B. (2024). Reinterpreting economic complexity in multiple dimensions. Technical report, arXiv. arXiv:2409.01830.

Perez, C. (2010). Technological revolutions and techno-economic paradigms. *Cambridge journal of economics*, 34(1):185–202.

Santoalha, A., Consoli, D., and Castellacci, F. (2021). Digital skills, relatedness and green diversification: A study of European regions. *Research Policy*, 50(9):104340.

Schumpeter, J. A. (1934). *The Theory of Economic Development... by Joseph A. Schumpeter, Translated from the German by Redvers Opie...* Harvard University Press.

Szulc-Wódarska, E. and Kowalski, A. M. (2024). Start-ups formation in italy: the role of regional innovation systems. *European Planning Studies*, 32(4):882–903.

References IV

Tanner, A. N. (2016). The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases. *Journal of Economic Geography*, 16(3):611–635.

Tushman, M. L. and Anderson, P. (1986). Technological Discontinuities and Organizational Environments. *Administrative Science Quarterly*, 31(3):439.

WIPO (2022). *World intellectual property report 2022: The direction of innovation*. WORLD INTELLECTUAL PROPERTY ORGANIZATION.

Xiao, J. and Boschma, R. (2023). The emergence of artificial intelligence in European regions: the role of a local ICT base. *Annals of Regional Science*, 71:747–773.