

The invisible majority: linking a national researcher registry to OpenAlex reveals Peru's uncertified artificial intelligence workforce

Milton Vladimir Mamani Calisaya

mmamanic@unap.edu.pe

Universidad Nacional del Altiplano

Charles Ignacio Mendoza Mollocondo

Universidad Nacional del Altiplano

Erardo Espinoza Coaquira

Universidad Nacional del Altiplano

Jesus Pari Flores

Universidad Nacional del Altiplano

Sandra Ines Ponce Umiña

Universidad Nacional del Altiplano

Research Article

Keywords: artificial intelligence, research workforce, national researcher registries, OpenAlex, record linkage, Peru

Posted Date: September 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10797556/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

National researcher registries are the main instrument through which Latin American governments certify their scientific workforce, yet how completely they capture emerging fields is rarely tested. We examine artificial intelligence (AI) in Peru by linking the national registry RENACYT (11,757 certified researchers; May 2024) to OpenAlex. We harvested the 2,052 works in the Artificial Intelligence and Computer Vision subfields published in 2015–2026 with Peru-affiliated authorship (3,546 distinct authors) and matched authors to the registry via three-tier name linkage. Only 24.1% of Peru-affiliated AI authors hold RENACYT certification; the 856 certified AI researchers are 7.3% of the registry. Peruvian AI output grew fivefold between 2015 and 2025. Lima concentrates 47% of certified AI researchers, but per-capita density is highest in southern regions - Arequipa, Tacna and Puno all exceed Lima - and the Universidad Nacional del Altiplano (Puno) ranks third nationally. Certification level correlates only weakly with AI output (Spearman $\rho = 0.09$), and productivity does not differ between Lima and the provinces. Women fall from 32.7% of the registry to 19.5% of the AI subset. The results expose a threefold-larger uncertified AI workforce and provide a replicable, open method for auditing the field-level coverage of any national researcher registry.

1 Introduction

Governments across Latin America operate national researcher registries that certify who counts as a researcher: Mexico's Sistema Nacional de Investigadores (SNI), Brazil's Lattes Platform, and Peru's Registro Nacional Científico, Tecnológico y de Innovación Tecnológica (RENACYT), among others. These instruments allocate reputation, and in several countries monetary incentives, on the basis of certification; they also supply the official statistics through which ministries plan capacity-building investments (González-Brambila & Veloso, 2007; Mena-Chalco et al., 2014). Their usefulness for planning, however, depends on an assumption that is rarely tested: that the registry actually covers the community of people producing research in a given field.

Artificial intelligence is a demanding test of that assumption. The field grows quickly, recruits heavily from recent graduates and industry practitioners, publishes in venues (conferences, preprint servers) that traditional certification criteria may weight poorly, and is precisely the field for which governments most urgently want workforce intelligence (OECD, 2019; UNESCO, 2021). If a national registry systematically misses the youngest and most applied segment of a fast-moving field, policies calibrated on registry statistics will understate national capacity and misallocate training investments.

Peru is an instructive case. It is a middle-income country with a small but rapidly formalizing science system: RENACYT, created in 2019 as the successor to the earlier REGINA registry, certified 11,757 researchers by May 2024, and its registry is published as open data. Peru also exhibits one of the region's starkest patterns of scientific centralization, with the capital historically concentrating most researchers and resources (Ciocca & Delgado, 2017). Whether emerging fields such as AI reproduce that centralization — or open opportunities for peripheral regions — is an open empirical question with direct policy relevance.

This article links the complete public RENACYT registry to OpenAlex (Priem et al., 2022), harvesting every indexed work in the *Artificial Intelligence* and *Computer Vision and Pattern Recognition* subfields with Peruvian authorship from 2015 onward, and matching the resulting author set to the registry by a documented three-tier name-linkage procedure. The linked dataset supports four research questions:

- **RQ1 (Coverage).** What fraction of Peru-affiliated authors publishing in AI holds RENACYT certification, and how has the certified and uncertified population evolved over time?

- **RQ2 (Geography).** How are certified AI researchers distributed across Peru's regions and institutions, in absolute and per-capita terms?
- **RQ3 (Certification and productivity).** How strongly does RENACYT certification level associate with AI-specific output and impact?
- **RQ4 (Gender).** Does the gender composition of the AI subset differ from that of the registry at large?

The study makes three contributions. Substantively, it provides the first field-level census of Peru's certified AI workforce and quantifies, at 3:1, the ratio of uncertified to certified AI authors. Methodologically, it demonstrates that a national registry can be audited for field-level coverage using only open data — the registry itself and OpenAlex — through a replicable linkage pipeline released with this article, applicable to any Latin American registry. For policy, it documents three patterns that certification statistics alone would conceal: an uncertified majority, a per-capita geography that inverts the capital-centric map, and a gender gap that widens precisely where the field is most selective.

1.1 Researcher certification systems in Latin America

Latin American researcher registries share a family resemblance: individual researchers apply for evaluation against publication, degree, and project criteria, and receive a public classification, sometimes tiered. Mexico's SNI, the oldest (1984), ties tiers to salary supplements and has been extensively studied as an incentive system (González-Brambila & Veloso, 2007). Brazil's Lattes Platform takes a different approach — a comprehensive CV infrastructure rather than a selective certification — and has become a standard data source for national-scale scientometrics (Mena-Chalco et al., 2014). Peru's RENACYT, administered by CONCYTEC, is selective and tiered. Under its 2021 regulation, researchers are classified into levels I (highest) through VII plus a "Distinguished Researcher" rank; researchers certified under the preceding 2018 regulation carried classifications in two groups (Carlos Monge Medrano, levels I–IV, and María Rostworowski, levels I–III) that map onto the 2021 scale. Certification requires, among other criteria, indexed publications and confers eligibility for public research funding. Existing studies of RENACYT are few and mostly institution-specific (e.g., Canales Sermeño et al., 2025); to our knowledge no study has measured the registry's coverage of a specific research field.

1.2 AI research in Latin America and its measurement

Bibliometric studies document rapid growth of Latin American AI output, led in volume by Brazil and Mexico (Aviles-Valenzuela et al., 2025). Peru is a smaller producer, and country-level studies of its AI capacity are scarce. The workforce dimension is even less studied than output: analyses of who produces AI research — their certification status, location, gender, and career stage — require linking bibliographic data to person-level administrative records, which is rarely done in the region. Gender imbalance is a documented global feature of the AI field: Stathoulopoulos and Mateos-Garcia (2019) estimate that only 13.8% of AI authors on arXiv are women, against a backdrop of broader gender disparities in science (Larivière et al., 2013). Whether national certification systems attenuate or amplify that imbalance is unknown.

1.3 OpenAlex as an open bibliographic source

OpenAlex is a fully open scholarly knowledge graph covering works, authors, institutions, and a hierarchical topic classification (Priem et al., 2022). Recent evaluations find its coverage and reference completeness broadly comparable to Web of Science and Scopus for recent publication years (Culbert et al., 2025), with the advantages of unrestricted access and richer ORCID capture — properties that matter for replicability in low-resource settings,

where commercial database licenses are a genuine barrier. OpenAlex assigns each work a primary topic nested in a subfield/field/domain hierarchy, which we use to delimit AI.

2 Data and methods

2.1 Sources

RENACYT registry. We use the complete public registry published on Peru's National Open Data Platform (cut-off 30 May 2024): 11,757 unique researchers with certification code, full name, gender, regulation (2021: 8,804; 2018: 2,953), certification status, level, certification date, age bracket, and district-level residence. The registry contains no institutional affiliation and no disciplinary classification — the gap this study's linkage fills.

OpenAlex. Using the public API, we harvested all works satisfying: (a) primary topic in subfield 1702 (*Artificial Intelligence*) or 1707 (*Computer Vision and Pattern Recognition*); (b) publication year ≥ 2015 ; (c) at least one authorship with a Peruvian institutional affiliation. The harvest (August 2026) returned 2,052 unique works carrying 5,234 Peru-affiliated authorships by 3,546 distinct OpenAlex author profiles. For each author we retained works, citations, first and last AI publication year, dominant topics, and modal Peruvian affiliation.

2.2 Record linkage

RENACYT names follow the format SURNAME, GIVEN NAMES; OpenAlex names are free-form. After normalization (case, accents, hyphens, particles), we matched on token sets in three tiers: **exact** (author's full token set identical to a registry entry; $n = 341$), **high** (≥ 3 tokens, all contained in a unique registry candidate including at least one surname and one given name; $n = 227$), and **medium** (2 tokens, unique candidate registry-wide; $n = 288$). Ambiguous cases (> 1 candidate; $n = 121$) and short names ($n = 74$) were left unmatched. Multiple OpenAlex profiles resolving to one registry entry were merged (outputs summed). The result is 856 unique certified researchers with AI output — 24.1% of Peru-affiliated AI authors and 7.3% of the registry. All analyses were replicated on the exact+high subset ($n = 568$) as a robustness check. Substantive conclusions are unchanged: the level–output correlation persists ($p = 0.089$, $p = 0.033$), the female share is similar (21.5%), Puno remains the third region by headcount, the southern Andean share rises slightly (13.2%), and the Lima–provinces productivity comparison remains non-significant ($p = 0.086$). One result is tier-sensitive and flagged where reported: the ordering among the top five institutions (Section 3.4).

Unmatched authors ($n = 2,582$) comprise genuinely uncertified authors — students, recent graduates, practitioners, and foreign collaborators holding temporary Peruvian affiliations — plus a residual of false negatives from unresolved name variants. To bound that residual, we drew a random sample of 100 unmatched authors (fixed seed), retrieved each author's best fuzzy candidate in the registry (token Jaccard combined with sequence similarity), and manually adjudicated every pair with similarity ≥ 0.60 . None was a clear false negative — the closest candidates differed in at least one surname — and one two-token name was an unresolvable ambiguity. By the rule of three, this places the false-negative rate among unmatched authors below approximately 3% (95% upper bound).

2.3 Derived variables

Harmonized level. 2018-regulation levels were mapped onto the 2021 scale following the official transition equivalence (Monge I–IV $\rightarrow$ I–IV; Rostworowski I–III $\rightarrow$ V–VII), yielding a single ordinal scale from VII (entry) to I and Distinguished.

Specialty. Each matched researcher was assigned one of ten specialty clusters by keyword classification of their dominant OpenAlex topic (applied AI, computer vision, AI in health, general machine learning, NLP, neural

networks/deep learning, robotics and control, AI in education, optimization/metaheuristics, semantic web).

Institution and sector. The researcher's modal Peruvian affiliation in OpenAlex, normalized to Spanish institutional names and classified public/private.

Regional density. Certified AI researchers per 100,000 inhabitants by department of residence, using the official INEI population estimates at 30 June 2024 (series 2000–2026).

2.4 Statistical analysis

We report distributions with percentages, Spearman and Pearson correlations between the harmonized level (as an ordinal hierarchy score) and AI output/citations, and a Mann-Whitney U test comparing per-researcher output between Lima–Callao and the rest of the country. Given the skew typical of production data (median 1 work; mean 2.1), rank-based statistics are primary. Analyses used Python (pandas/SciPy); the full pipeline and data are openly released (Section: Data availability).

2.5 Scope conditions

Three boundaries should be borne in mind. First, "AI researcher" is operationalized as *primary-topic dominance*: researchers for whom AI is a secondary line within another dominant field are not captured. Second, RENACYT certification dates begin in 2019 (the registry's creation), so cohort analyses cover 2019–2024 while production series cover 2015–2026 (2026 partial, harvested in August). Third, OpenAlex affiliations reflect a researcher's most recent institutional signal, which may differ from their affiliation at certification.

3 Results

3.1 The certified AI workforce (RQ1a)

Of 11,757 RENACYT researchers, 856 (7.3%) have AI-dominant production, accounting for 1,812 authorships and 13,371 accumulated citations over 2015–2026 (Table 1). The subset's certification structure is pyramidal and nearly identical in shape to the registry at large: 53.0% sit in the entry levels VI–VII and only 10.2% in the top tiers (Distinguished, I, II) (Fig. 1). The subset is young — 56.5% are aged 30–49 — and 83.9% hold active certification status.

Table 1

Descriptive statistics: full RENACYT registry vs. the AI subset. Levels on the 2021 scale (I = highest); 2018 regulation harmonized

Indicator	Full registry	AI subset
N researchers	11,757	856 (7.3% of registry)
Women	32.7%	19.5%
High tiers (Distinguished-IV)	23.7%	26.8%
Resident in Lima and Callao	53.1%	47.1%
Southern Andes (Puno, Cusco, Apurímac, Ayacucho, Huancavelica)	8.0%	10.9%
Aged 30–49	52.7%	56.5%
Active certification	74.8%	83.9%
AI works (2015–2026)	-	1,812
Median AI works per researcher	-	1
Accumulated citations of AI works	-	13,371

3.2 Growth and the uncertified majority (RQ1b)

Peru-affiliated AI output rose from 63 works in 2015 to 340 in 2025, a fivefold increase (CAGR 2015–2024: 18.3%), with sustained acceleration from 2018 (Fig. 2a). The author population grew faster still: new Peru-affiliated authors entering AI each year rose from about 100 in 2015 to more than 600 in 2025. Yet the certified fraction of each entering cohort remains low and roughly flat (Fig. 2b): across the whole window, **only one in four Peru-affiliated AI authors (24.1%) holds RENACYT certification**. The uncertified majority – 2,582 authors – is dominated by thesis students, recent graduates, and practitioners publishing without registry standing, alongside foreign collaborators with temporary Peruvian affiliations. The national AI pipeline is thus roughly three times larger than the certified stock that official statistics report.

3.3 Geography: concentration in volume, not in density (RQ2)

Lima and Callao host 47.1% of certified AI researchers (403 of 856) – slightly *below* their 53.1% share of the registry overall. Per-capita ranking inverts the map (Fig. 3; Fig. 4; Table 2): Arequipa leads with 6.9 certified AI researchers per 100,000 inhabitants, followed by Tacna (6.1) and Puno (4.7), all above Lima (3.4). The southern Andean corridor (Puno, Cusco, Apurímac, Ayacucho, Huancavelica) contributes 10.9% of the subset. Productivity per researcher does not differ significantly between Lima–Callao and the provinces (Mann-Whitney U = 83,344, p = 0.575; both medians = 1): **the territorial gap is one of headcount, not of individual performance**.

Table 2
Regional distribution and density per 100,000 inhabitants. Population: official INEI estimates at 30 June 2024

Region	RENACYT researchers	AI subset	RENACYT /100k	AI /100k
Arequipa	679	110	42.3	6.85
Tacna	186	24	47.0	6.06
Puno	404	57	33.4	4.72
Moquegua	56	7	27.8	3.48
Lima	6,056	389	53.6	3.44
Apurimac	126	14	29.5	3.28
Pasco	44	7	16.6	2.64
Amazonas	155	10	36.0	2.32
La Libertad	683	41	32.1	1.92
Junin	349	24	25.3	1.74
Lambayeque	360	22	26.4	1.62
Loreto	177	15	16.7	1.42
Ancash	184	16	15.3	1.33
Callao	185	14	15.3	1.16
Cusco	243	15	17.4	1.07
Piura	276	23	12.8	1.07
San Martin	101	9	10.7	0.95
Huanuco	211	7	28.3	0.94
Ayacucho	90	6	13.4	0.90
Madre De Dios	65	1	33.0	0.51
Cajamarca	128	7	8.8	0.48
Tumbes	51	1	19.1	0.38
Ucayali	82	2	12.8	0.31
Huancavelica	79	1	23.6	0.30
Ica	142	3	13.4	0.28

3.4 Institutions (RQ2)

The top three institutions by certified AI researchers are the Pontificia Universidad Católica del Perú (78; 34.6% in high tiers), the Universidad Nacional de San Agustín de Arequipa (60), and – notably for a registry-based ranking – the **Universidad Nacional del Altiplano in Puno (48)**, ahead of the Universidad Nacional Mayor de San Marcos (40) and the Universidad Nacional de Ingeniería (37) (Table 3; Fig. 5). Public institutions employ 47.9% of the 743

researchers with an identifiable affiliation. PUCP stands out on seniority composition; the large public regional universities stand out on volume. One caveat: restricting to exact+high linkage tiers reorders the top five – PUCP's count relies disproportionately on medium-confidence matches (its authors more often publish under short, two-token name forms), whereas UNSA's and UNA-Puno's counts are tier-robust; the presence and approximate size of all five communities is stable across tiers.

Table 3
Top 10 institutions by RENACYT-certified researchers with AI output. Main affiliation per OpenAlex; ordering among the top five is sensitive to linkage tier (see Section 3.4)

#	Institution	Sector	AI researchers	AI works	Works/res.	% high tiers
1	Pontificia Universidad Católica del Perú	Private	78	199	2.5	35%
2	Universidad Nacional de San Agustín de Arequipa	Public	60	142	2.4	23%
3	Universidad Nacional del Altiplano (Puno)	Public	48	87	1.8	23%
4	Universidad Nacional Mayor de San Marcos	Public	40	108	2.7	30%
5	Universidad Nacional de Ingeniería	Public	37	75	2.0	22%
6	Universidad César Vallejo	Private	35	46	1.3	20%
7	Universidad Católica San Pablo	Private	33	137	4.2	15%
8	Universidad Peruana de Ciencias Aplicadas	Private	31	112	3.6	23%
9	Universidad Privada del Norte	Private	22	58	2.6	23%
10	Universidad Tecnológica del Perú	Private	22	46	2.1	23%

3.5 Specialization (RQ2)

Applied AI dominates: 41.0% of researchers have a dominant topic in sectoral application domains (solar energy, geosciences, agriculture, services), followed by computer vision (16.8%) and AI in health (11.1%) (Fig. 6; Table 4). Fundamental work – researchers whose dominant topic is neural network methods themselves – is thin (4.9%). NLP (6.5%) concentrates in Lima, Arequipa, and Puno. The regional specialty matrix (Table 4) shows Lima leading every cluster in volume but no cluster in exclusivity.

Table 4

Region x specialty matrix (8 regions with most certified AI researchers). Applied = applied AI (sectoral domains); CV = computer vision; NN/DL = neural networks / deep learning; Sem.W. = semantic web and ontologies

Region	ML	Applied	Educ.	Health	Optim.	NLP	NN/DL	Robot.	CV	Sem.W.	Total
Lima	36	153	10	53	8	22	13	16	70	8	389
Arequipa	6	40	3	1	3	10	8	10	26	3	110
Puno	2	30	2	2	0	7	2	5	3	4	57
La Libertad	2	20	1	5	0	1	2	4	6	0	41
Tacna	4	15	0	1	0	1	0	0	3	0	24
Junin	4	15	0	1	0	0	0	1	3	0	24
Piura	1	11	1	0	0	0	7	0	3	0	23
Lambayeque	1	12	2	3	1	0	1	0	2	0	22

3.6 Certification level and AI productivity (RQ3)

The association between harmonized certification level and AI output is positive but weak: Spearman $\rho = 0.09$ ($p = 0.010$); Pearson $r = 0.14$ ($p < 0.001$); for citations, $\rho = 0.13$ ($p < 0.001$) (Fig. 7; Table 5). Mean output rises from 1.8 works at level VII to 3.3 at level I, but dispersion within levels dwarfs the gradient. RENACYT level, built on whole-career criteria, is at best a coarse proxy for AI-specific capacity – consistent with certification measuring general trajectory rather than field specialization.

Table 5

Associations between RENACYT hierarchy, AI productivity and territory (N = 853 researchers with a current level)

Test	Variables	Statistic	p-value	Interpretation
Spearman	Level hierarchy x AI works	$\rho = 0.088$	0.00975	Weak but significant positive association
Pearson	Level hierarchy x AI works	$r = 0.138$	5.56e-05	Consistent with Spearman
Spearman	Level hierarchy x AI citations	$\rho = 0.13$	0.000149	Weak positive association
Mann-Whitney U	AI works: Lima-Callao vs provinces	$U = 83344$	0.575	No significant difference (medians 1 vs 1)

3.7 Gender (RQ4)

Women are 32.7% of the registry but 19.5% of the AI subset; within the subset's top tiers (Distinguished-II) female share falls to 16.1% (vs. 21.2% in the registry's top tiers) (Fig. 8). The AI field thus amplifies the registry's baseline gender gap at every level, most severely at the top – a pattern consistent with, and in Peru's certified segment slightly less extreme than, the global arXiv-based estimate of 13.8% female AI authorship (Stathoulopoulos & Mateos-Garcia, 2019).

4 Discussion

4.1 An invisible majority, and what it means for registry-based statistics

The central finding is a coverage figure: 24.1%. Any statistic about Peruvian AI capacity computed from RENACYT alone — headcounts, growth rates, regional distributions — describes one quarter of the population that publishes. This is not primarily a failure of the matching procedure (false negatives are bounded at a few percentage points) but a structural feature of how certification interacts with a young, fast-growing field: entry cohorts are dominated by students and early-career authors who have not yet sought or obtained certification. Registries are lagging indicators of emerging-field capacity. For planning purposes — training pipelines, graduate program investment, industry linkage — the uncertified 76% is arguably the more policy-relevant population, and it is invisible in official statistics.

4.2 The per-capita inversion of Peru's scientific map

Peru's scientific centralization is real in absolute terms, but the AI subset shows the periphery is not uniformly peripheral. Arequipa, Tacna, and Puno exceed Lima in certified AI researchers per capita, and the Universidad Nacional del Altiplano's third-place national ranking shows that a high-altitude public regional university can sustain a nationally significant AI community. Combined with the null result on individual productivity (Section 3.3), the evidence contradicts a deficit narrative for provincial AI research: what the provinces lack is volume of people, not performance per person. For decentralization policy, this favors consolidating existing regional poles — which already demonstrate competitive productivity — over greenfield capacity creation.

4.3 Certification level as a weak field signal

The weak level–output correlation ($\rho \approx 0.1$) cautions against using RENACYT tiers as a proxy for field expertise in, for example, panel composition, grant gatekeeping, or national AI strategy staffing. Whole-career certification and field-specific capacity are different constructs; conflating them risks both excluding productive early-career specialists (concentrated in levels VI–VII) and over-crediting senior generalists.

4.4 A gender gap that compounds

The 13-point drop in female share from registry to AI subset, and the further drop within the subset's top tiers, indicates two compounding filters: selection into the field and progression within it. Because the pattern matches global AI-specific evidence rather than Peru-specific certification rules, generic gender policies in science are unlikely to close it; field-targeted instruments (dedicated fellowships, retention tracking at the field level) are indicated.

4.5 A replicable audit method for the region

Every ingredient of this study is open: the registry (national open-data portal), the bibliography (OpenAlex), and the pipeline (released code). The same audit — harvest a field's works, link authors to the registry, measure coverage, geography, and gender — can be run against Mexico's SNI, Colombia's ScienTI, or any registry publishing person-level open data, at zero database cost. We propose registry field-coverage auditing as a routine scientometric exercise for Latin American science policy.

4.6 Limitations

(1) The registry lacks affiliation and discipline, so both are inherited from OpenAlex and reflect current rather than at-certification signals. (2) Name-based linkage over Spanish naming conventions leaves a residual false-negative rate bounded below $\sim 3\%$ by manual adjudication of a random sample of 100 unmatched authors (Section 2.2); all conclusions survive restriction to the exact+high tiers. (3) Primary-topic delimitation misses researchers for whom AI is a secondary method; our counts are therefore conservative for AI-adjacent capacity. (4) OpenAlex coverage, while

comparable to commercial databases for recent years (Culbert et al., 2025), indexes Spanish-language regional venues imperfectly, which may understate applied and provincial output. (5) Certification dates begin in 2019, truncating cohort analysis.

4.7 Future work

Linking this dataset to thesis repositories (e.g., RENATI and university repositories) would trace the formation-to-certification pipeline directly – including whether the uncertified majority eventually certifies, emigrates, or exits research – and is the subject of ongoing work by the authors, focused on the Altiplano region.

5 Conclusion

Linking Peru's complete researcher registry to OpenAlex shows that the country's certified AI workforce – 856 researchers, 7.3% of RENACYT – sits inside a publishing population roughly four times its size, growing fivefold in output over a decade. The certified map is capital-heavy in volume but not in density: southern regions lead per capita, one Altiplano university ranks third nationally, and individual productivity shows no capital advantage. Certification level barely predicts AI output, and the field amplifies the science system's gender gap at every tier. National registries remain indispensable policy instruments, but for emerging fields they should be read as what they are – certified cores, not censuses – and audited against open bibliographic data, for which this study provides a fully replicable template.

Declarations

Funding:

This research received no external funding. **Competing interests:** the authors declare no competing interests. **Author contributions (CRediT):** M.V.M.C.: software, formal analysis, writing – original draft. C.I.M.M.: methodology, writing – review & editing. E.E.C.: methodology, writing – review & editing. J.P.F.: software, writing – review & editing. S.I.P.U.: conceptualization. **Ethics:** the study uses exclusively public administrative and bibliographic data; no personal data beyond the public registry was processed.

Author Contribution

M.V.M.C. developed the software, performed the formal analysis, and wrote the main manuscript text. C.I.M.M. and E.E.C. designed the methodology. J.P.F. contributed to the software. S.I.P.U. conceived the study. All authors reviewed and approved the manuscript.

Acknowledgement

The authors used Claude (Anthropic) as an assistive tool for data-pipeline development and manuscript drafting. All analyses, results, and text were verified and approved by the authors, who take full responsibility for the content of this article.

Data Availability

The datasets generated and analysed during the current study, together with the full analysis code, are openly available at Zenodo: <https://doi.org/10.5281/zenodo.22070643> (mirrored at <https://github.com/vl4dimr/renacyt-openalex-ai-peru>). Source data: RENACYT registry (Plataforma Nacional de Datos Abiertos, open licence) and OpenAlex (CC0).

References

1. Aviles-Valenzuela, A., Acosta-Barreno, K., Espinel-Obregoso, F. P., & Espinoza Carrasco, A. S. (2025). Trends and analysis of artificial intelligence research in Latin America (2013–2023). *Open Information Science*, 9(1), 20250023. <https://doi.org/10.1515/opis-2025-0023>
2. Canales Sermeño, G., Valenzuela Ramos, M. R., Barrios Sayritupac, M. M., & Eras Lévano, C. J. (2025). Estudiantes de ciencias de la salud como investigadores RENACYT de una universidad peruana: un estudio bibliométrico. *Atención Primaria*, 57(5), 103117. <https://doi.org/10.1016/j.aprim.2024.103117>
3. Ciocca, D. R., & Delgado, G. (2017). The reality of scientific research in Latin America; an insider's perspective. *Cell Stress and Chaperones*, 22(6), 847–852. <https://doi.org/10.1007/s12192-017-0815-8>
4. Culbert, J., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*. <https://doi.org/10.1007/s11192-025-05293-3>
5. González-Brambila, C., & Veloso, F. M. (2007). The determinants of research output and impact: A study of Mexican researchers. *Research Policy*, 36(7), 1035–1051. <https://doi.org/10.1016/j.respol.2007.03.005>
6. Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, 504(7479), 211–213. <https://doi.org/10.1038/504211a>
7. Mena-Chalco, J. P., Digiampietri, L. A., Lopes, F. M., & Cesar, R. M., Jr. (2014). Brazilian bibliometric coauthorship networks. *Journal of the Association for Information Science and Technology*, 65(7), 1424–1445. <https://doi.org/10.1002/asi.23010>
8. OECD. (2019). *Artificial intelligence in society*. OECD Publishing. <https://doi.org/10.1787/eedfee77-en>
9. Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*. <https://arxiv.org/abs/2205.01833>
10. Stathoulopoulos, K., & Mateos-Garcia, J. (2019). *Gender diversity in AI research*. Nesta. <https://doi.org/10.2139/ssrn.3428240>
11. UNESCO. (2021). *UNESCO science report: The race against time for smarter development*. UNESCO Publishing.

Figures

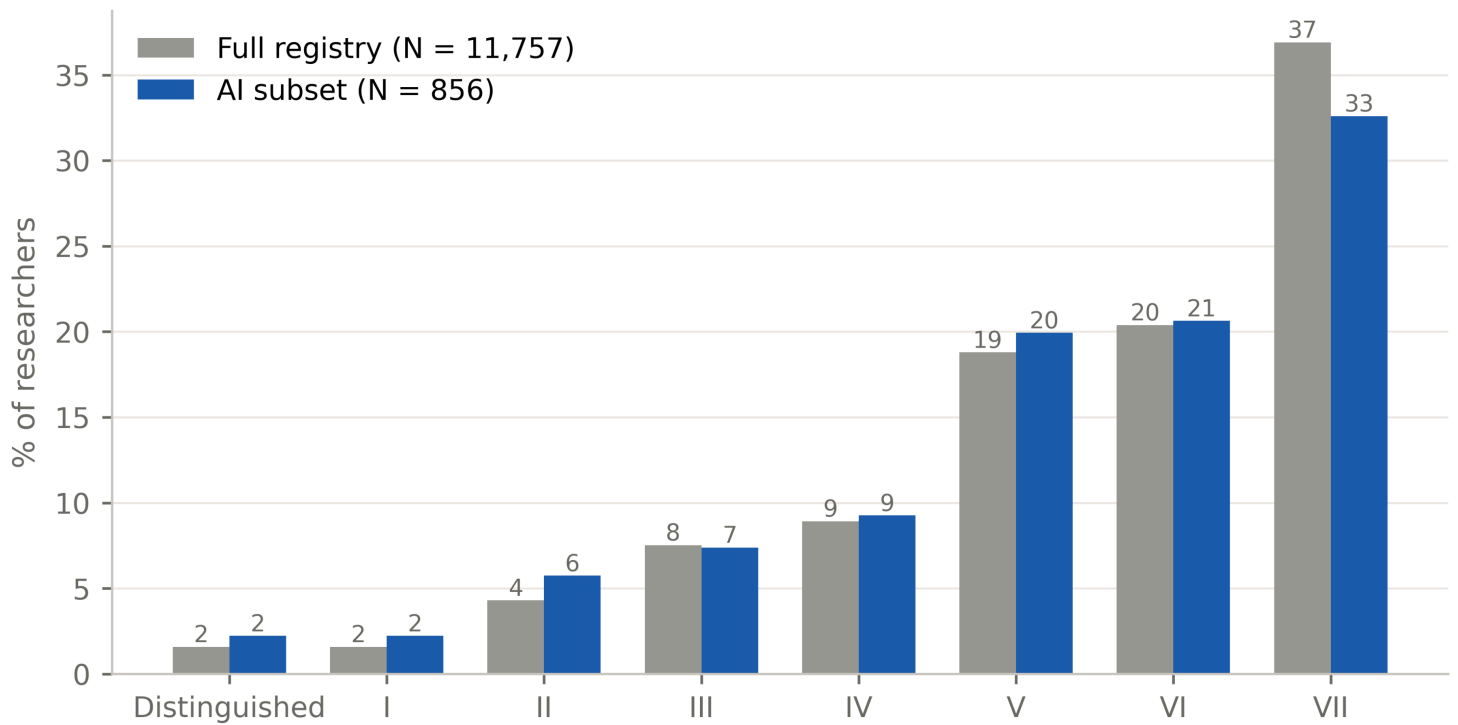


Figure 1

Distribution by RENACYT level (2021 scale, I = highest): full registry vs. AI subset

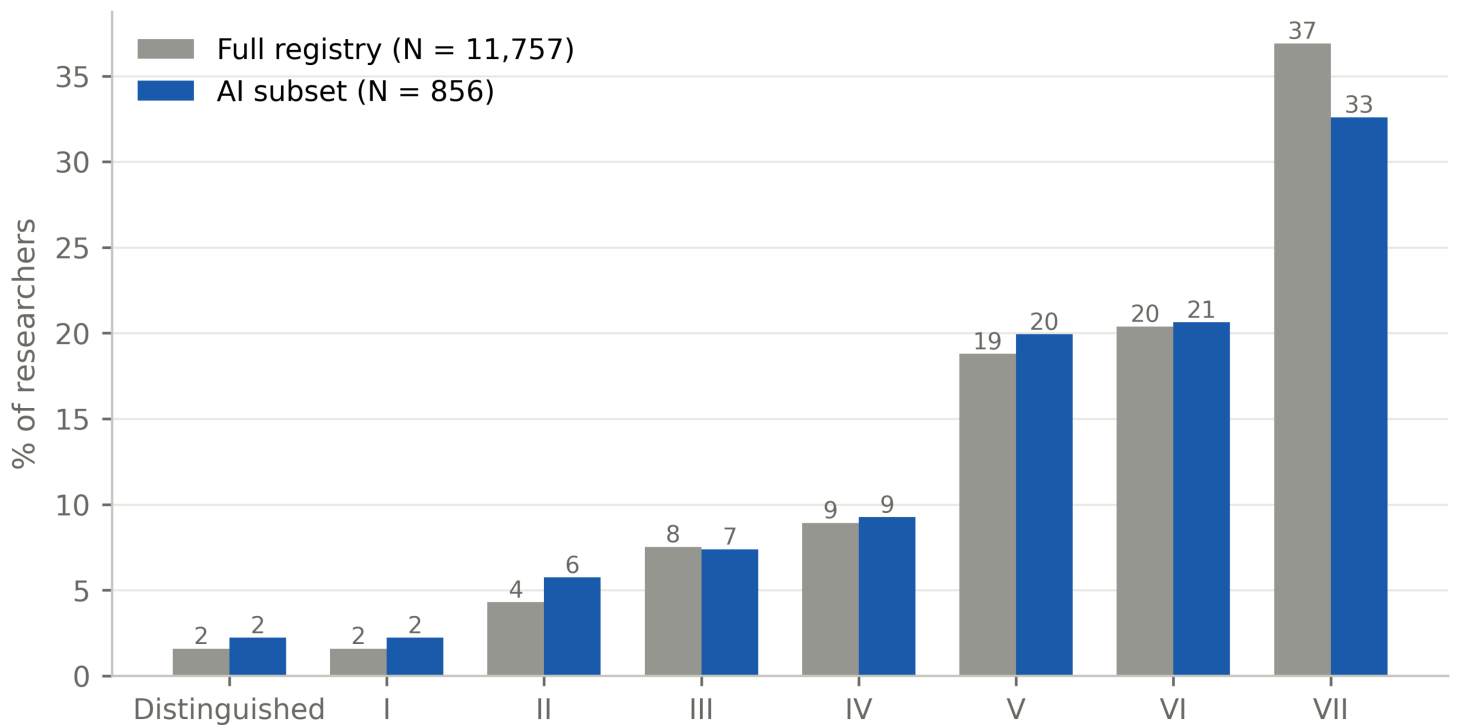


Figure 1

Distribution by RENACYT level (2021 scale, I = highest): full registry vs. AI subset

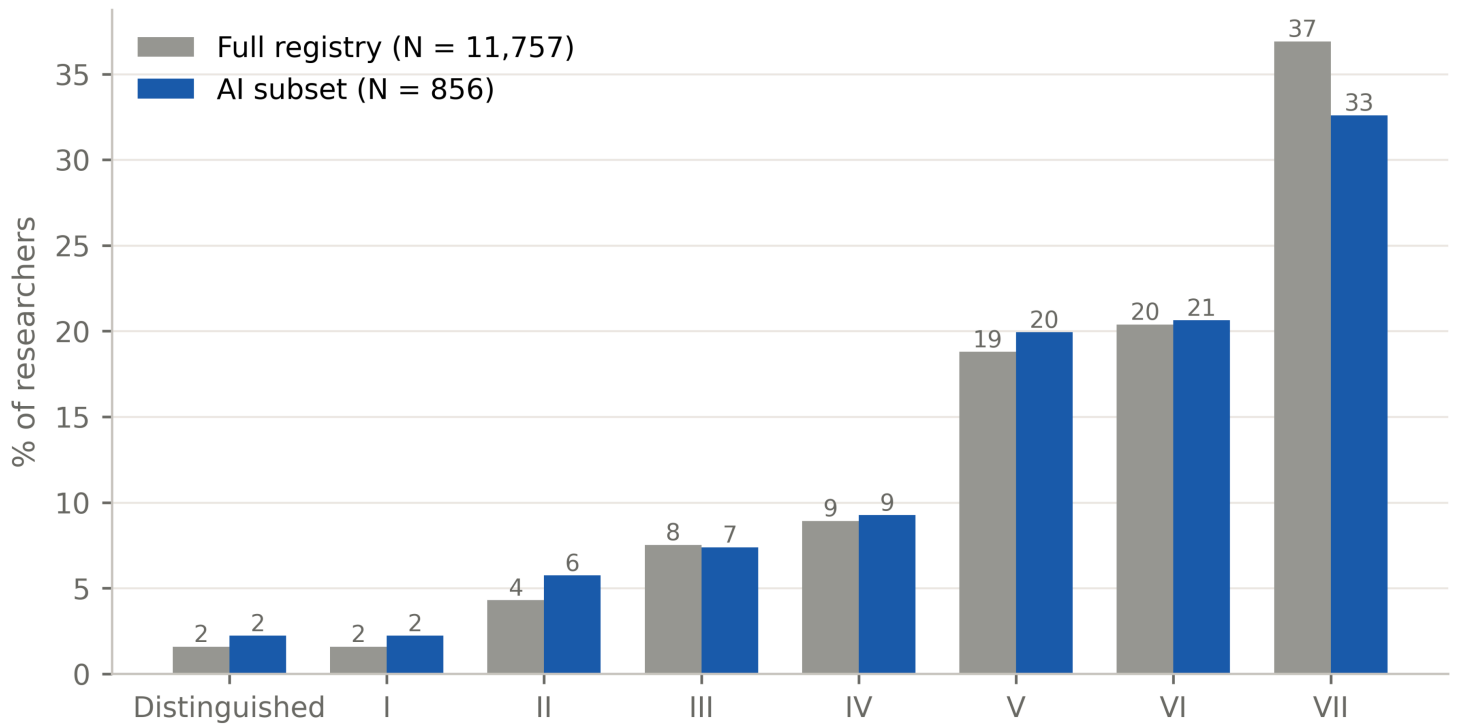


Figure 1

Distribution by RENACYT level (2021 scale, I = highest): full registry vs. AI subset

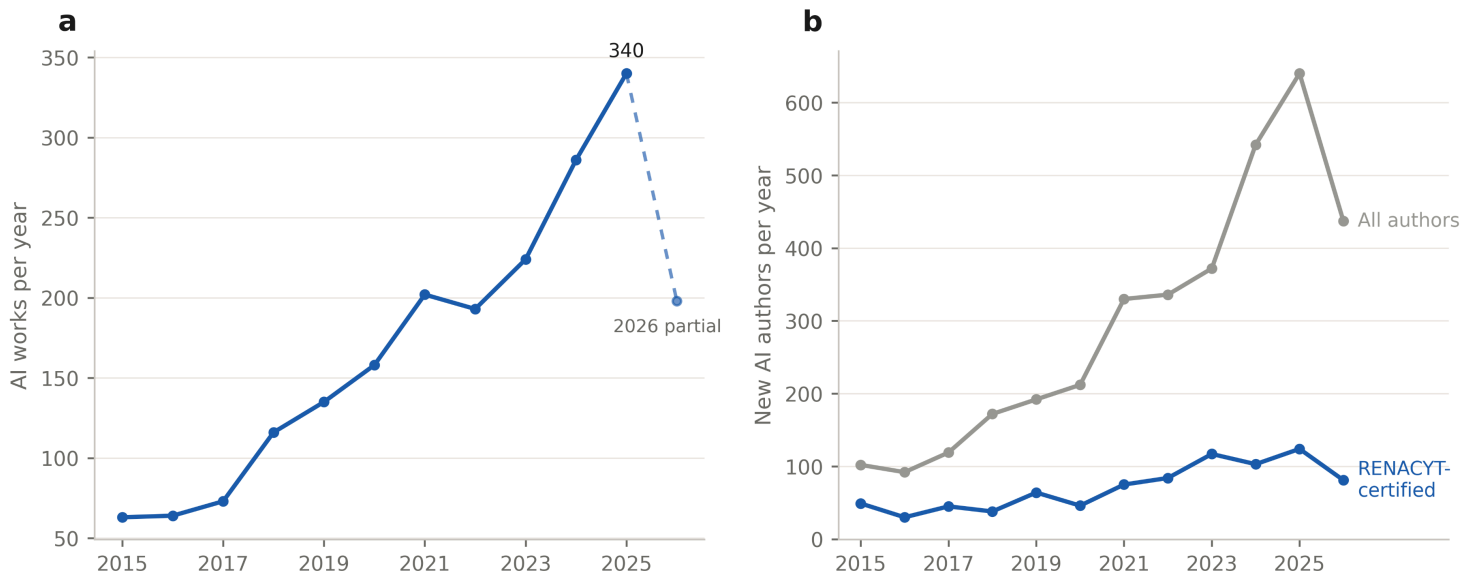


Figure 2

(a) Peru-affiliated AI works per year (2026 partial). (b) New AI authors per year of first AI publication - all authors vs. RENACYT-certified

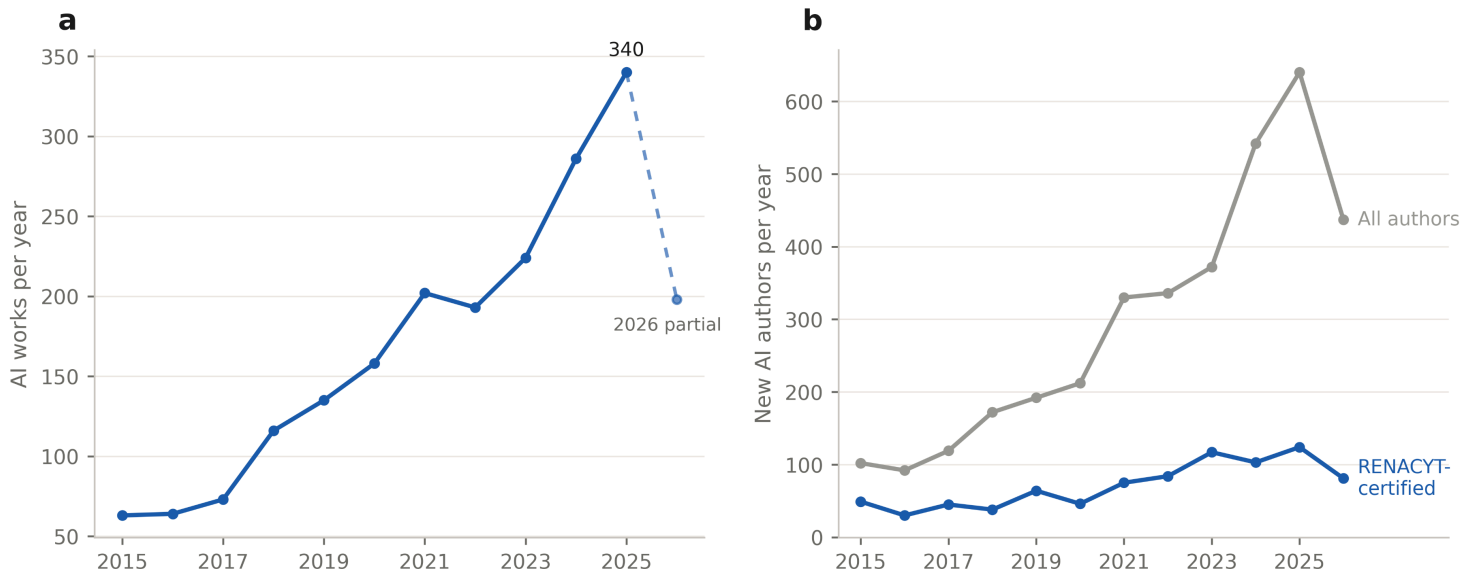


Figure 2

(a) Peru-affiliated AI works per year (2026 partial). (b) New AI authors per year of first AI publication - all authors vs. RENACYT-certified

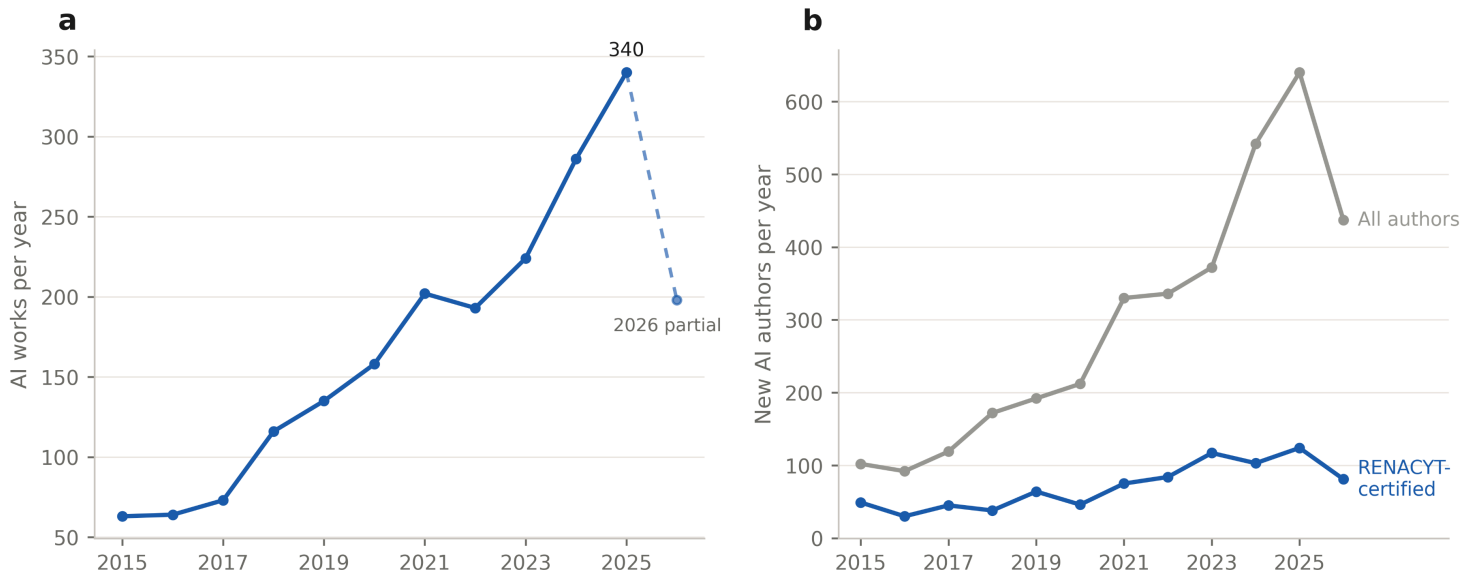
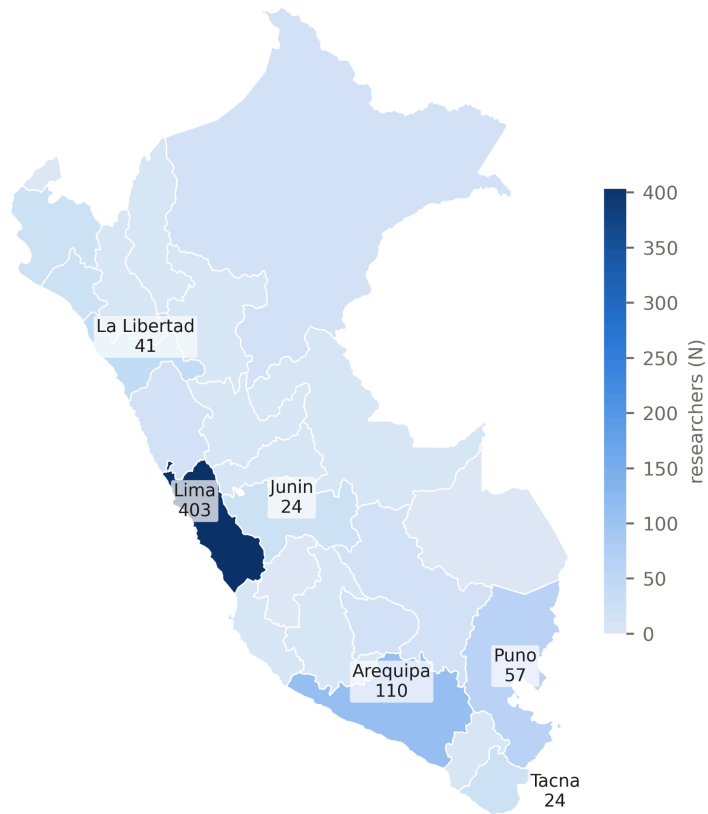
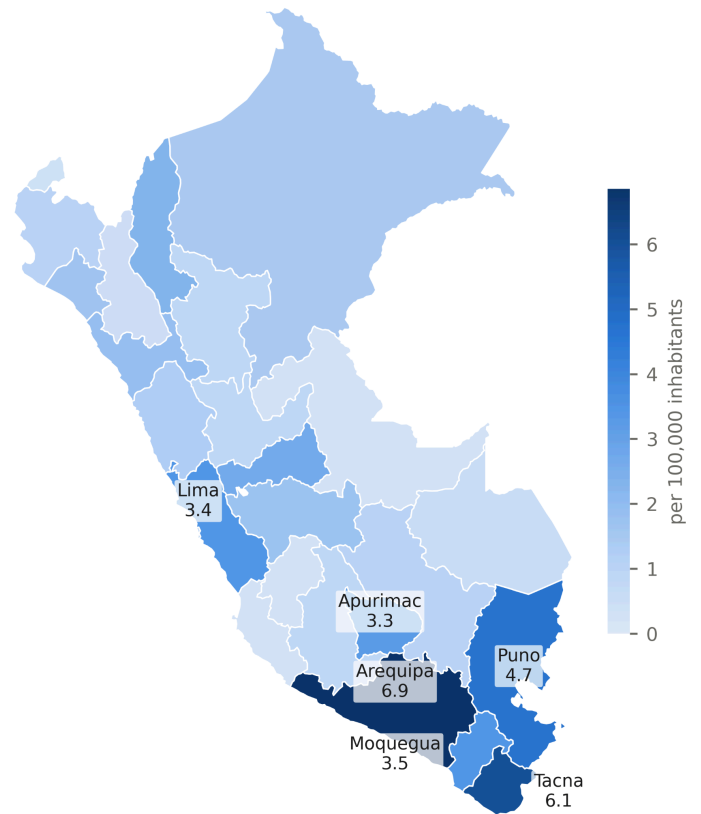


Figure 2

(a) Peru-affiliated AI works per year (2026 partial). (b) New AI authors per year of first AI publication - all authors vs. RENACYT-certified

a**b****Figure 3**

Geographic distribution of RENACYT-certified AI researchers by department: (a) absolute counts; (b) per 100,000 inhabitants. Lima includes Callao. Population: official INEI estimates, 30 June 2024

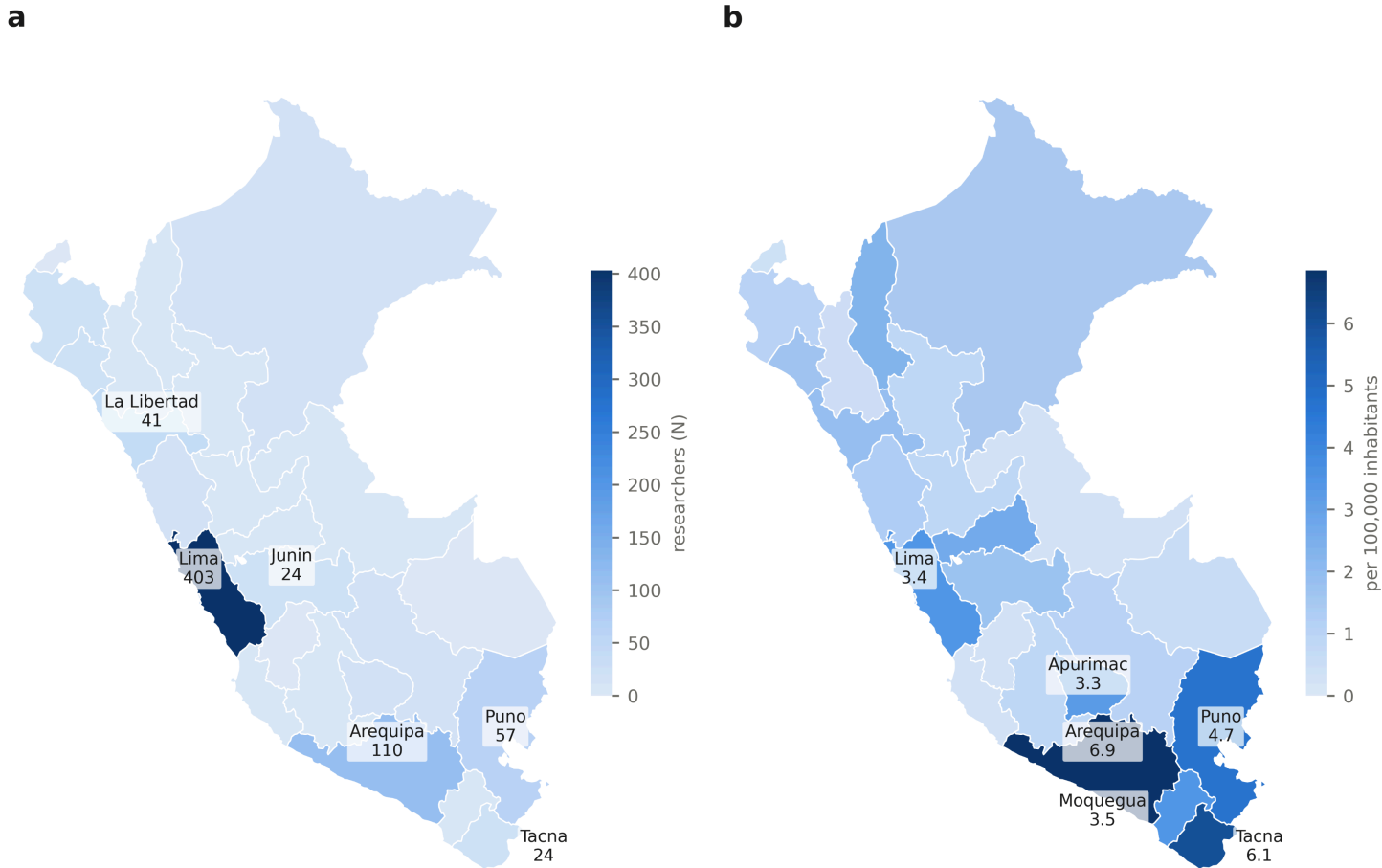
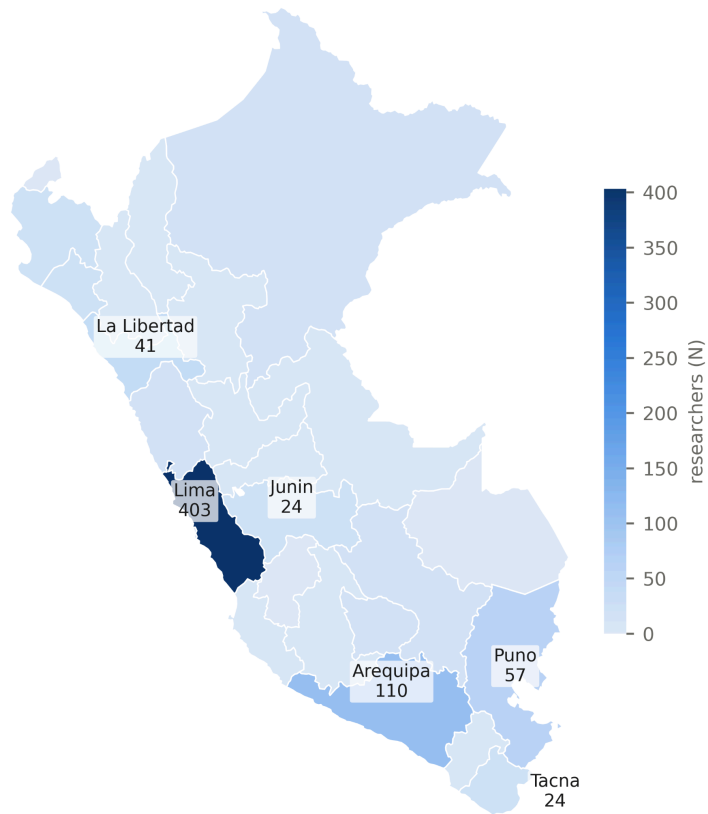
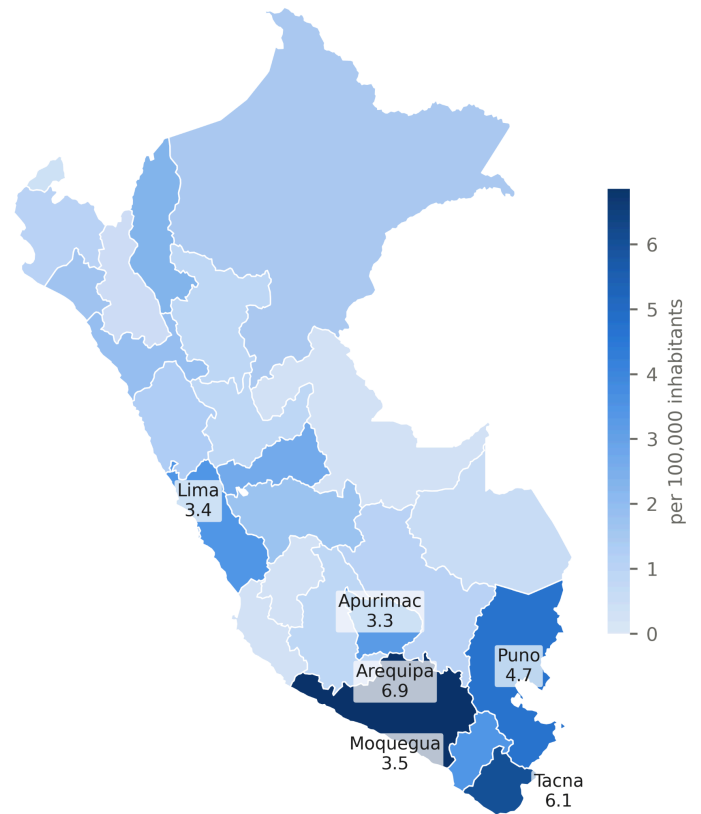


Figure 3

Geographic distribution of RENACYT-certified AI researchers by department: (a) absolute counts; (b) per 100,000 inhabitants. Lima includes Callao. Population: official INEI estimates, 30 June 2024

a**b****Figure 3**

Geographic distribution of RENACYT-certified AI researchers by department: (a) absolute counts; (b) per 100,000 inhabitants. Lima includes Callao. Population: official INEI estimates, 30 June 2024

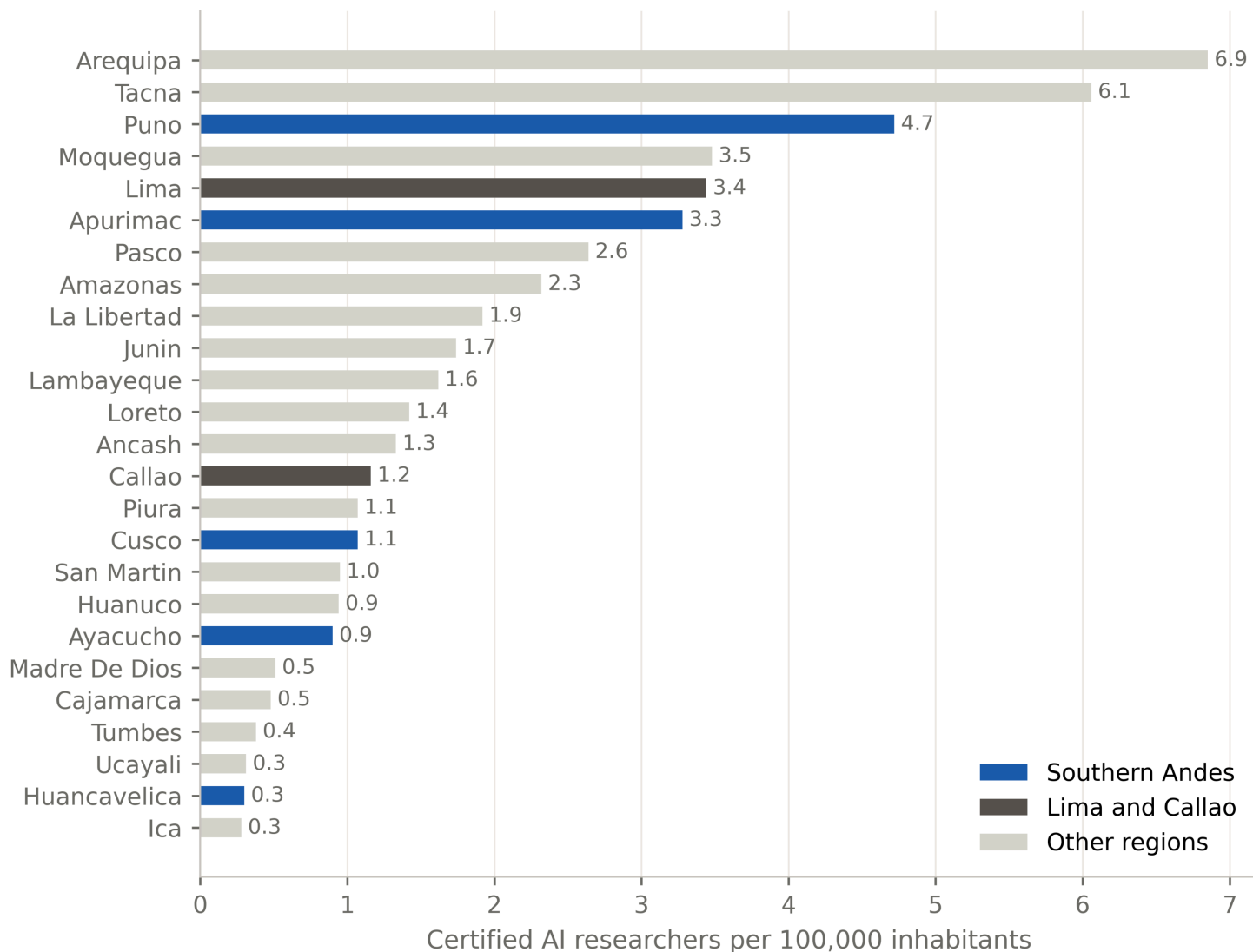


Figure 4

Certified AI researchers per 100,000 inhabitants, highlighting Lima-Callao and the southern Andes

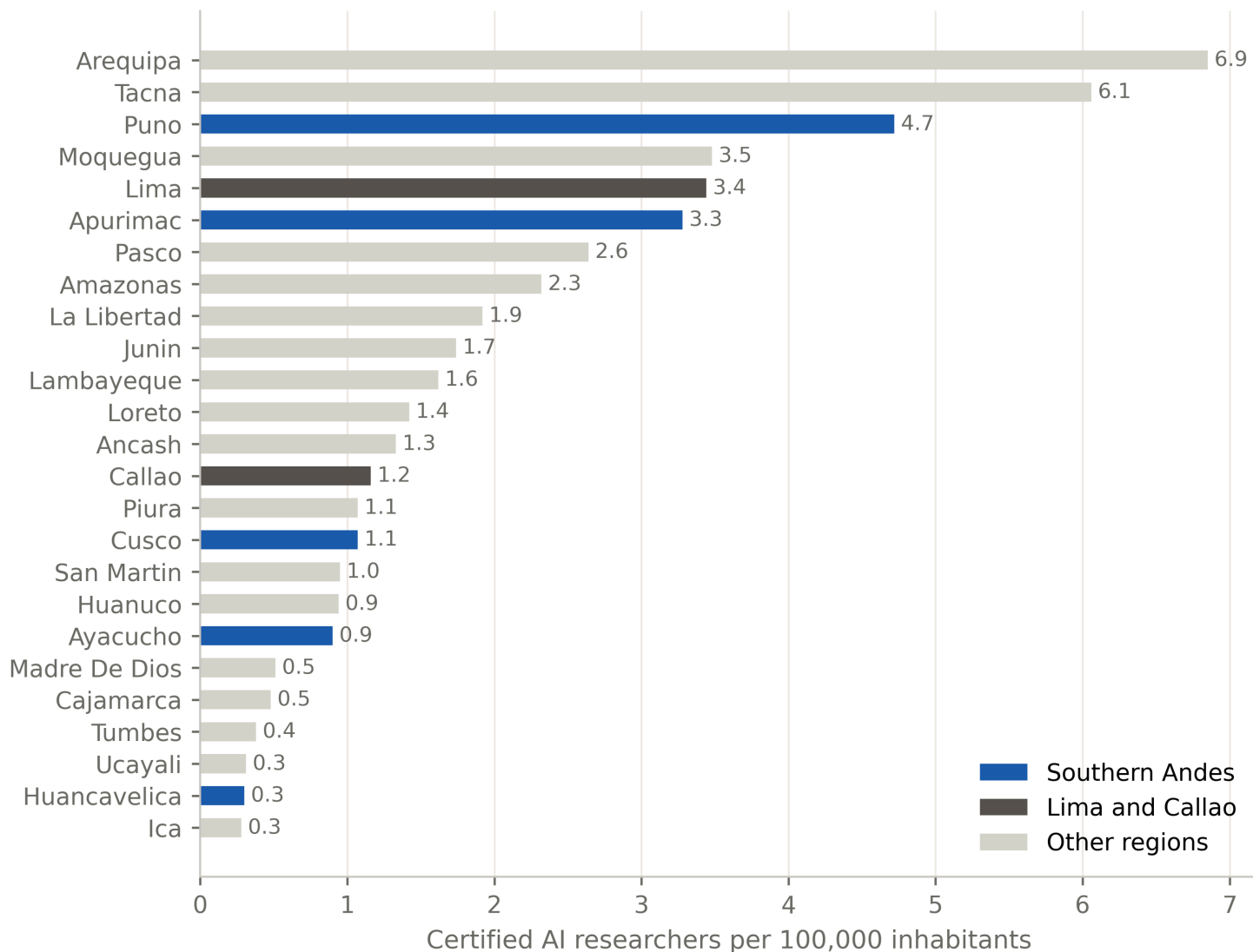


Figure 4

Certified AI researchers per 100,000 inhabitants, highlighting Lima-Callao and the southern Andes

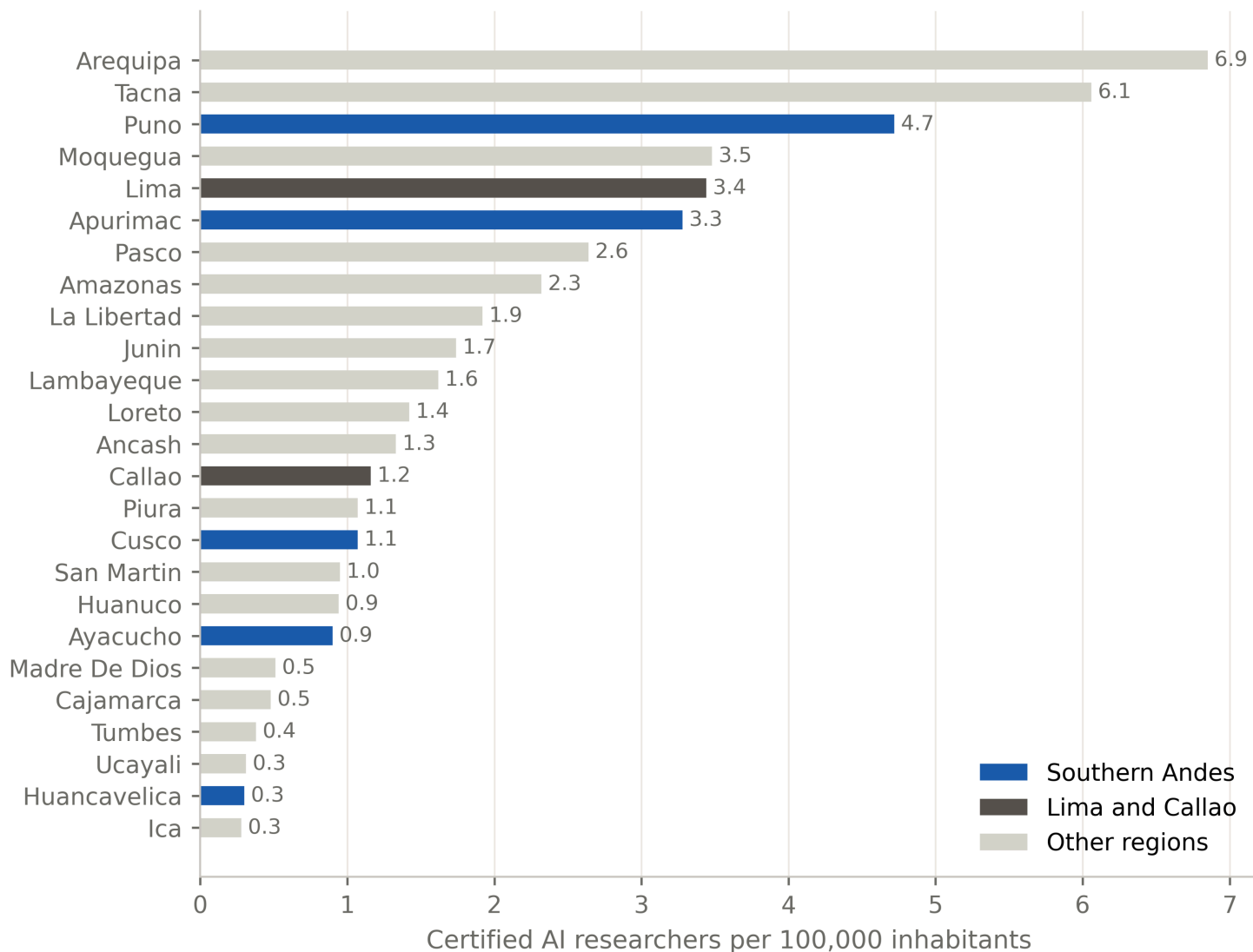


Figure 4

Certified AI researchers per 100,000 inhabitants, highlighting Lima-Callao and the southern Andes

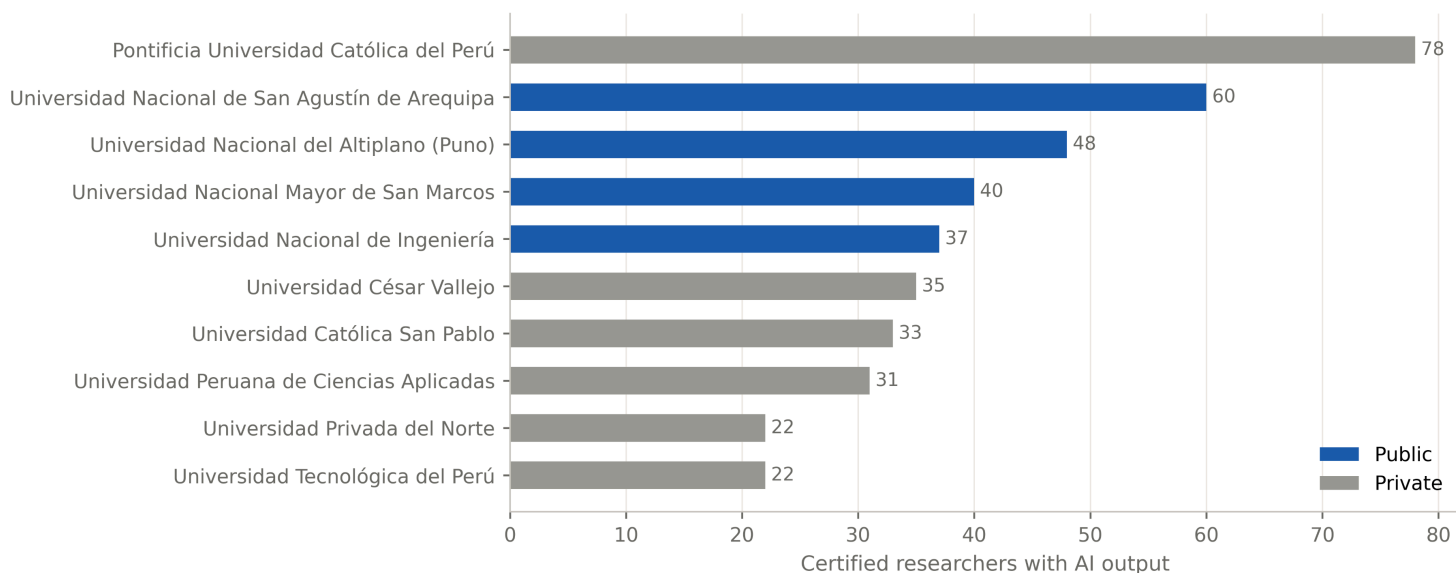


Figure 5

Top 10 institutions by certified AI researchers (blue: public; grey: private)

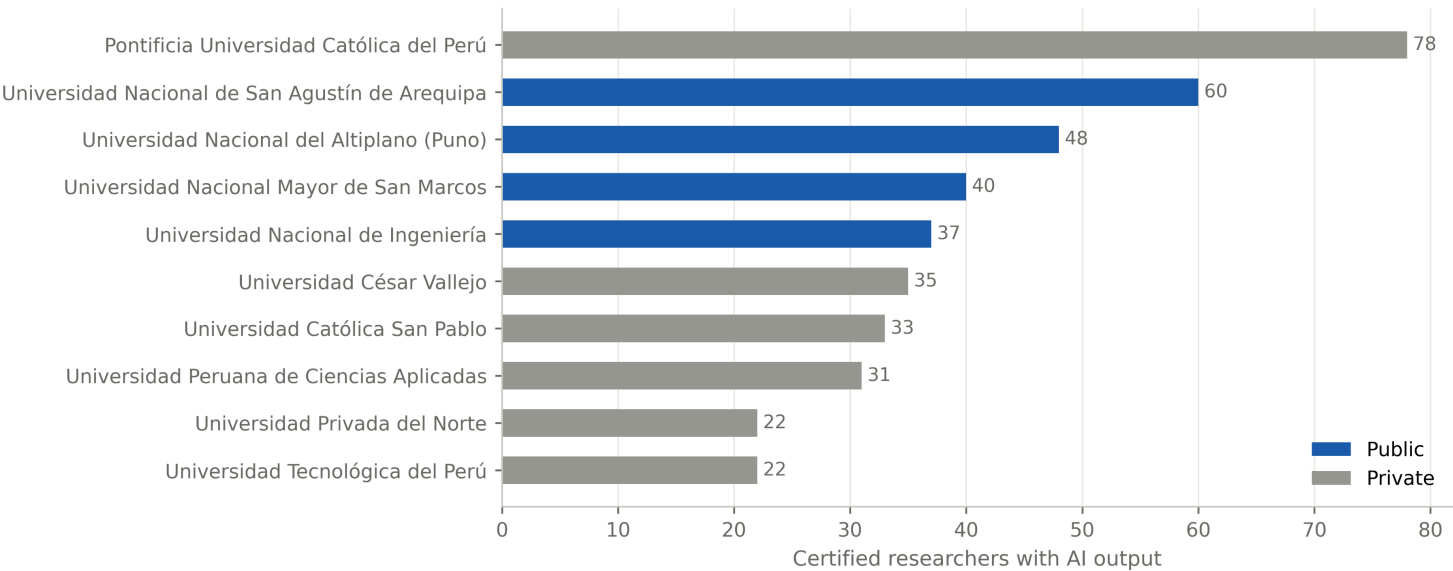


Figure 5

Top 10 institutions by certified AI researchers (blue: public; grey: private)

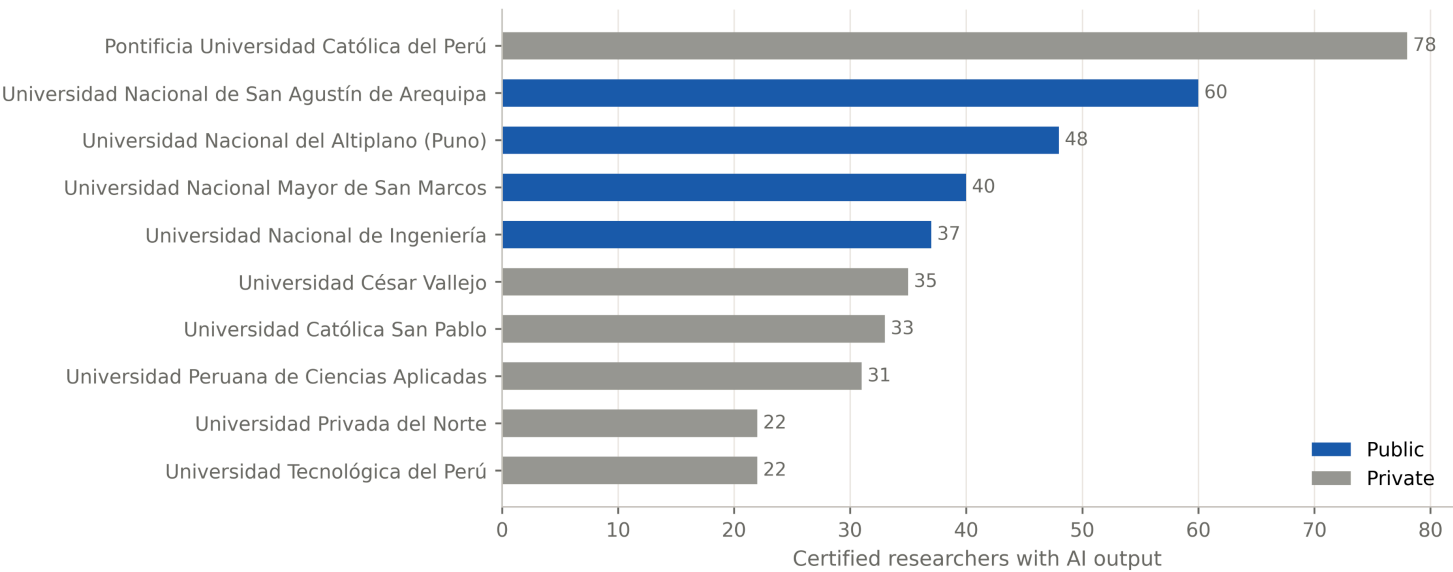


Figure 5

Top 10 institutions by certified AI researchers (blue: public; grey: private)

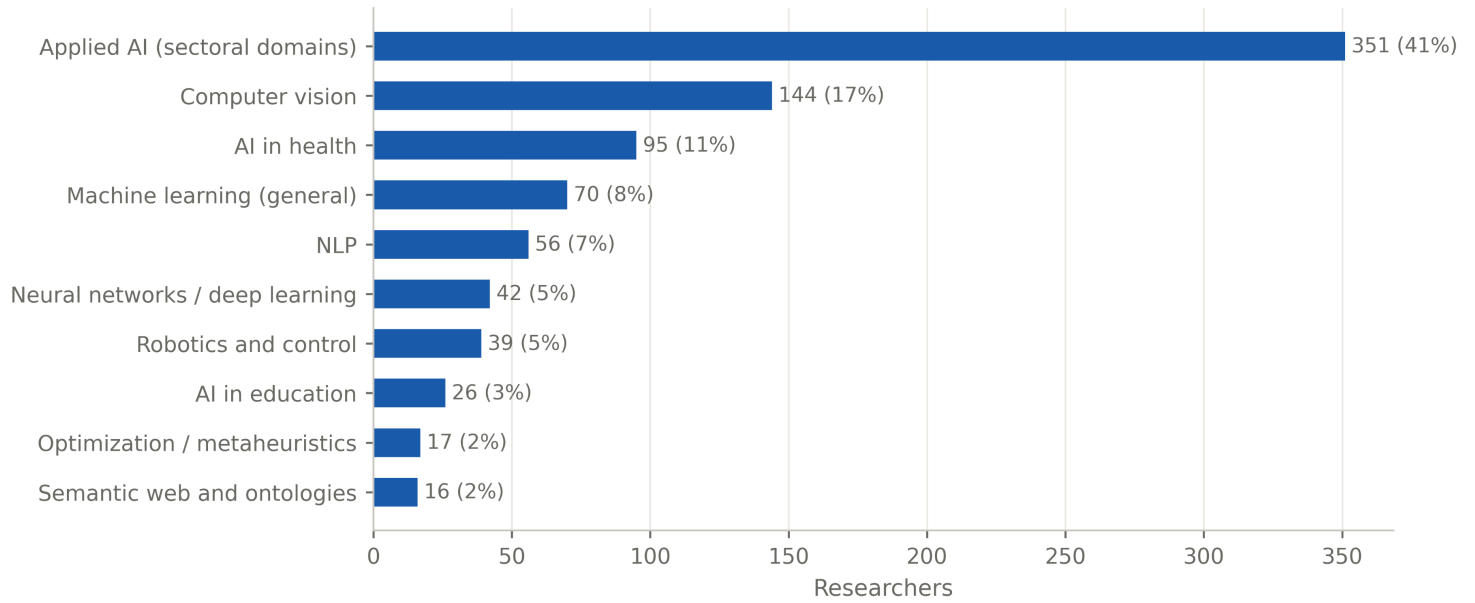


Figure 6

Specialty clusters by dominant OpenAlex topic

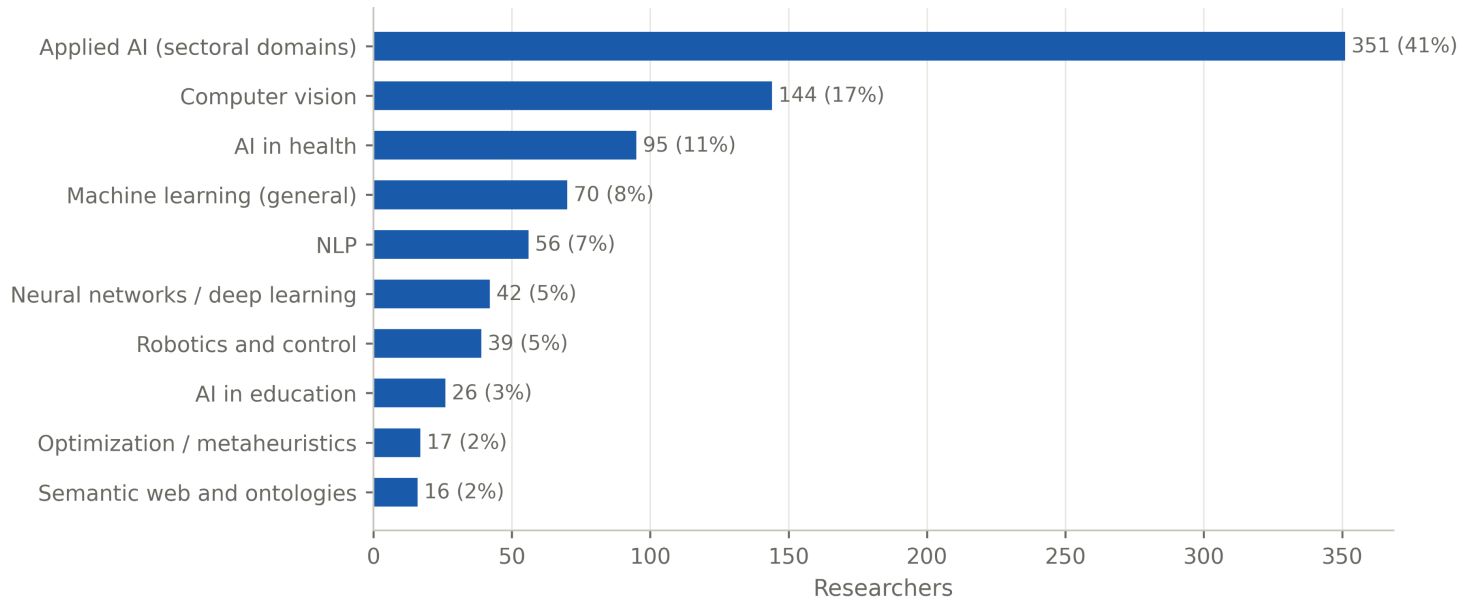


Figure 6

Specialty clusters by dominant OpenAlex topic

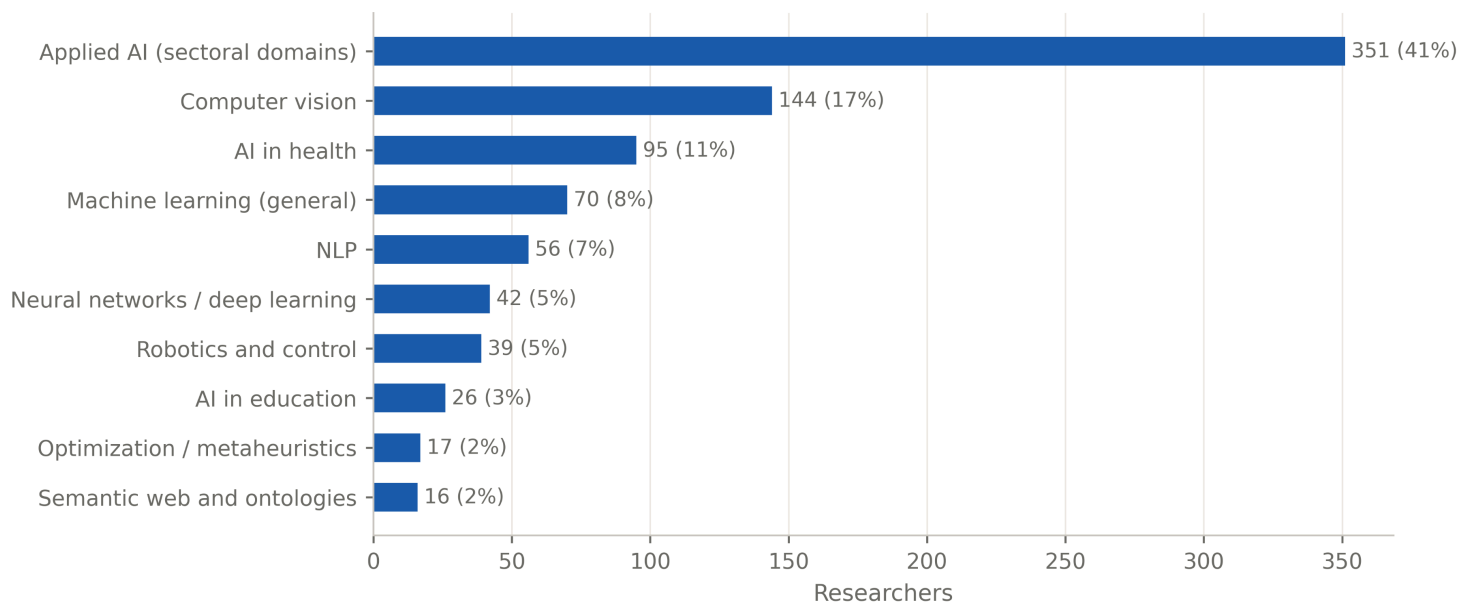


Figure 6

Specialty clusters by dominant OpenAlex topic

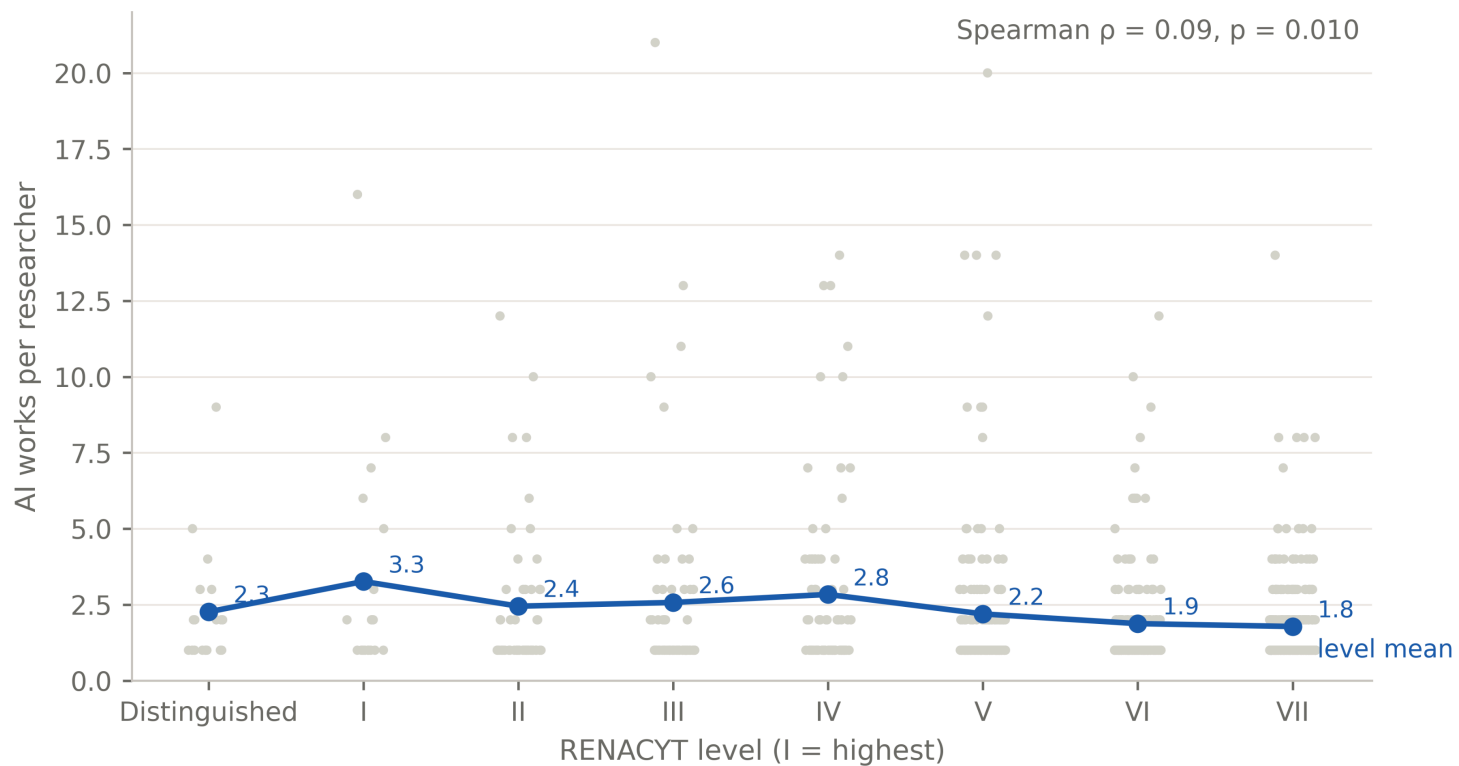


Figure 7

AI works per researcher by certification level (points: researchers; line: level means). Spearman rho = 0.09, p = 0.010

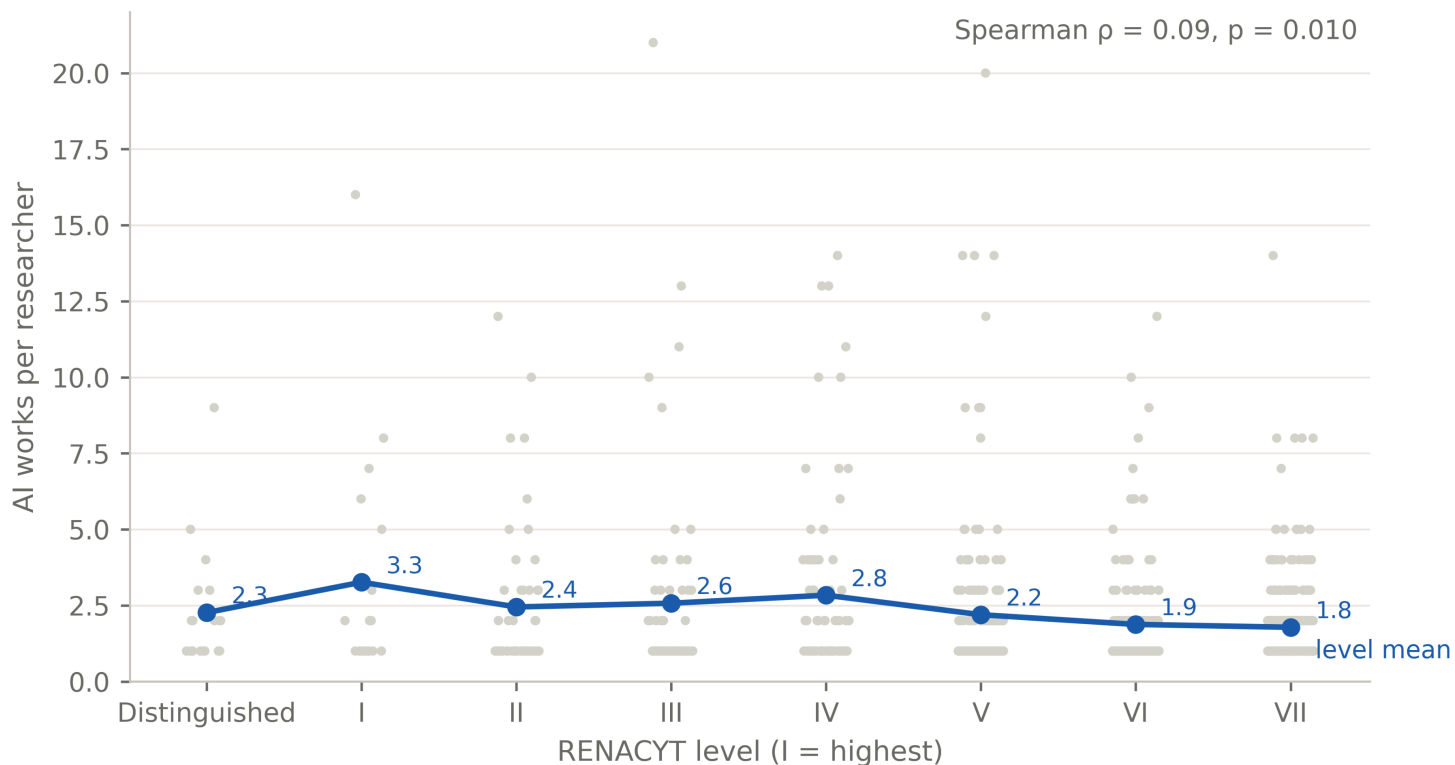


Figure 7

AI works per researcher by certification level (points: researchers; line: level means). Spearman rho = 0.09, p = 0.010

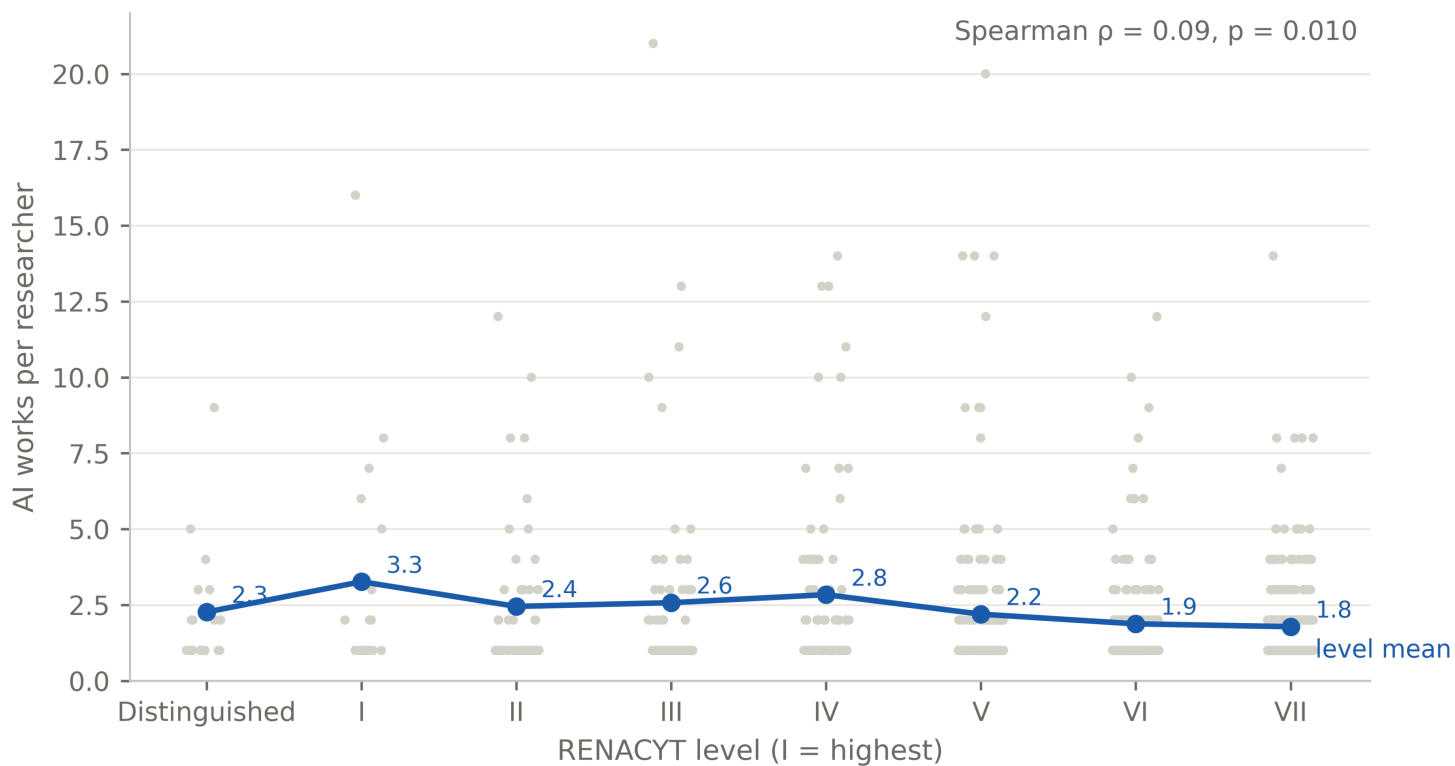


Figure 7

AI works per researcher by certification level (points: researchers; line: level means). Spearman rho = 0.09, p = 0.010

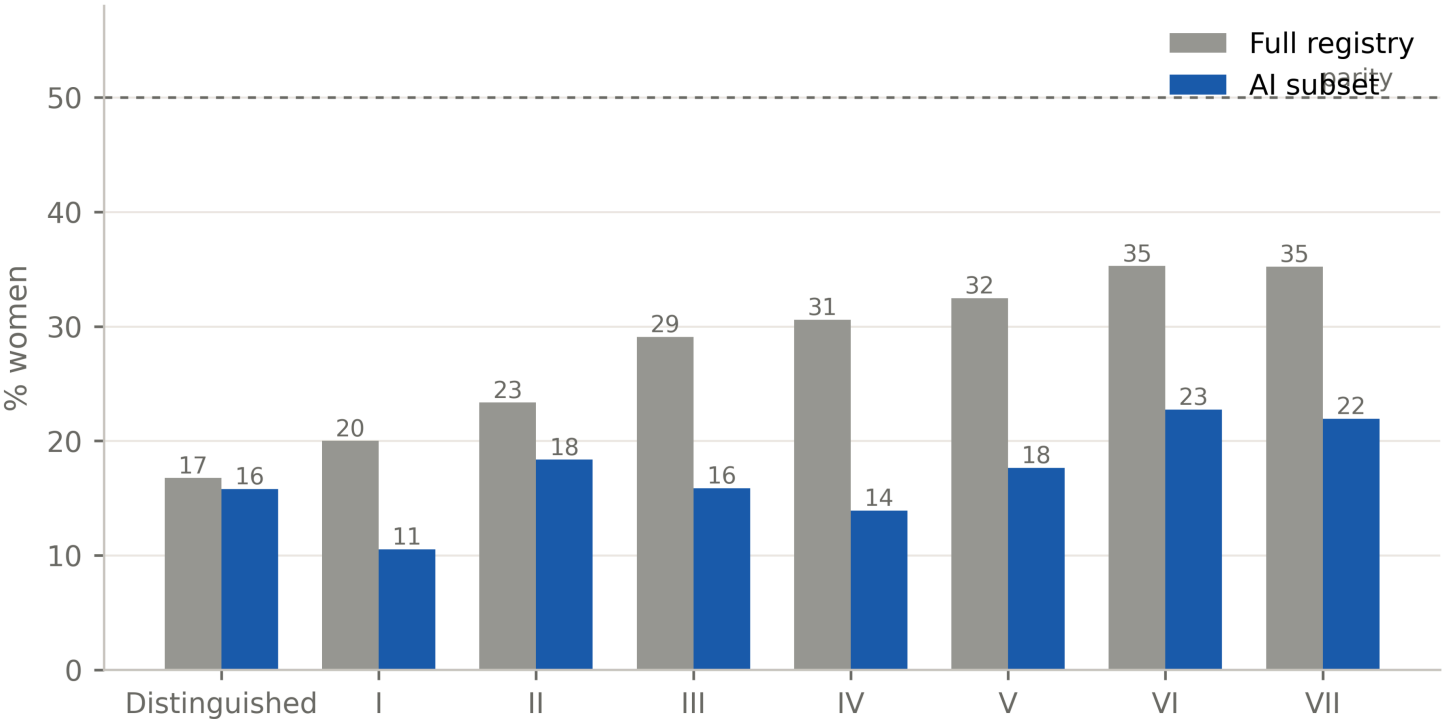


Figure 8

Female share by certification level: full registry vs. AI subset

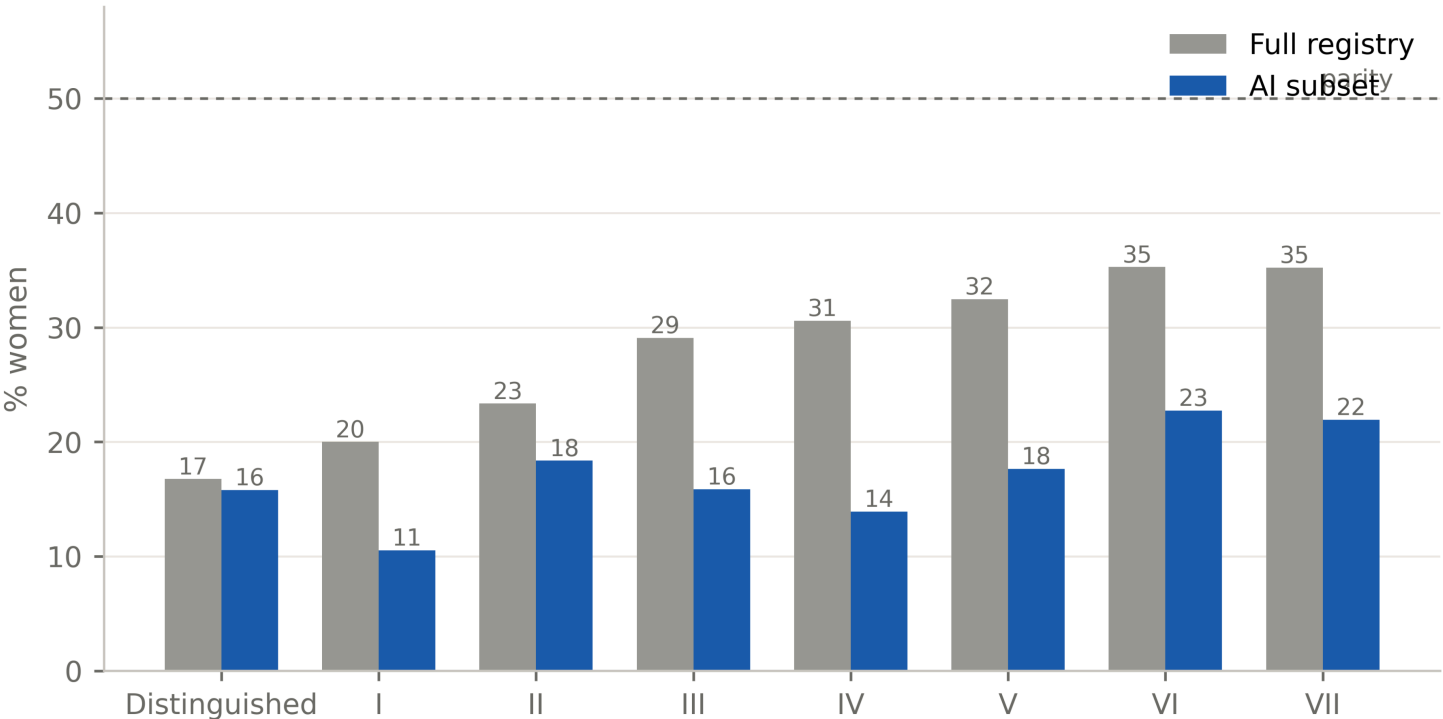


Figure 8

Female share by certification level: full registry vs. AI subset

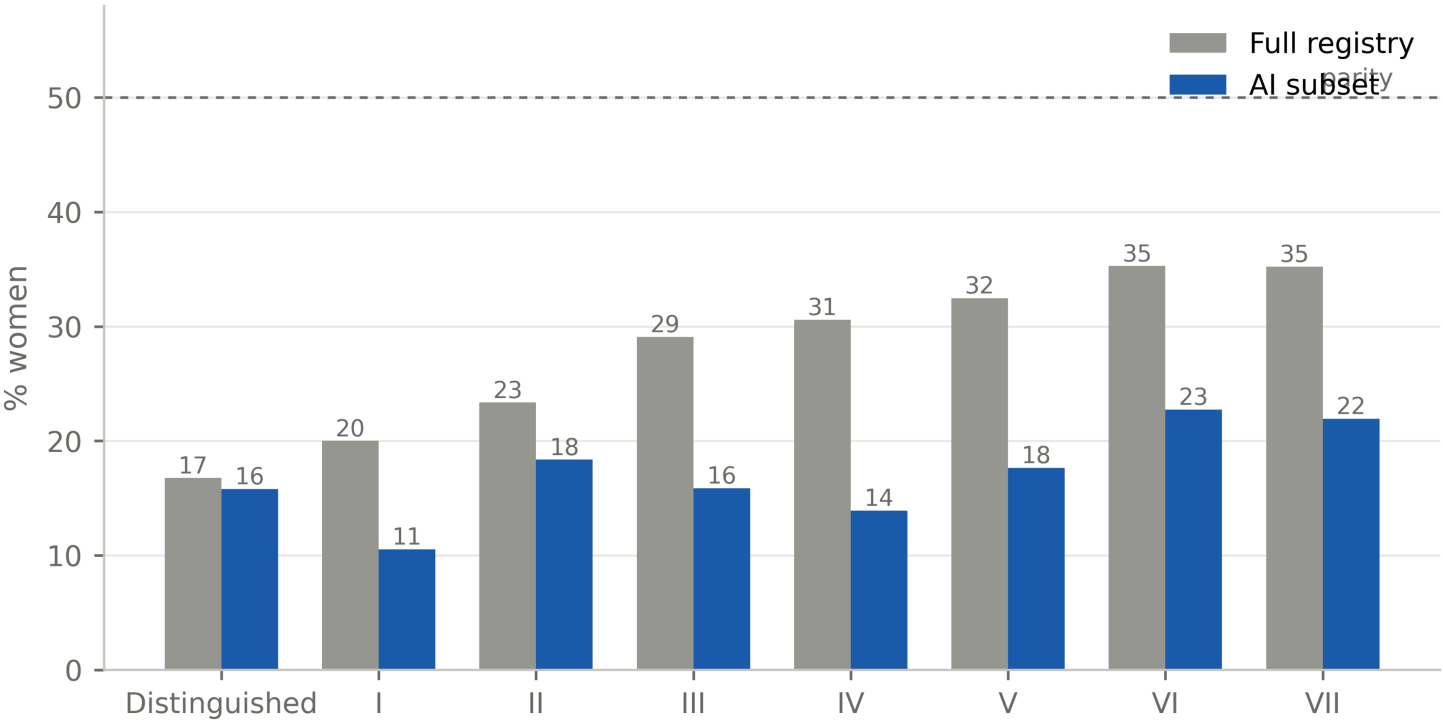


Figure 8

Female share by certification level: full registry vs. AI subset