The geography of digital transformation: Skill diversity, product innovation, and firm performance in Poland

Extended abstract
Submitted to the 65th ERSA Congress, Sofia, 25–28 August 2026
Special Session: Digitalisation, Skills, and Regional Transformation
Weronika Wilczewska
Lublin University of Technology, POLAND
w.wilczewska@pollub.pl

---

1. Introduction and motivation

The digital transformation of manufacturing is widely recognised as a key driver of competitiveness, yet the mechanisms linking workforce digital competences to digitalized product offerings remain poorly understood at the regional level. European policy frameworks—from the Digital Decade targets to regional smart specialization strategies—implicitly assume that investing in digital skills translates into more innovative, digitally enriched products. However, the empirical evidence supporting this assumption is surprisingly thin, especially regarding the heterogeneity of the process across territories with different knowledge infrastructures, industrial compositions, and absorptive capacities.

This paper addresses two interrelated research gaps. First, prior studies predominantly measure digital skills as a monolithic input (e.g., the share of ICT workers), neglecting the compositional diversity of competences that firms demand. Yet recombinant innovation theory suggests that it is the breadth and variety of knowledge inputs—not merely their volume—that enables novel product configurations. We therefore distinguish between digital skill *intensity* and digital skill *diversity*, arguing that the latter is a more relevant predictor of product digitalization. Second, while a growing body of research connects skills to aggregate firm performance, the intermediate channel through which competences become embedded in products is rarely measured directly. We propose that product digitalization acts as a mediating mechanism, and that this mediation is moderated by the regional context in which the firm operates.

Poland offers a particularly instructive empirical setting. As a major European manufacturing economy undergoing rapid but spatially uneven digital transformation, it exhibits substantial inter-regional variation in both skill supply and industrial structure. The sixteen NUTS-2 voivodeships range from the innovation-dense Mazowieckie region around Warsaw to predominantly agricultural eastern regions with markedly thinner digital ecosystems. Investigating how skill diversity translates into product digitalization across these diverse contexts provides insights with broader relevance for EU cohesion and industrial policy.

2. Theoretical framework and hypotheses

Our conceptual framework draws on three interconnected strands of literature. The first is the resource-based view of the firm and, more specifically, the knowledge-based theory of competitive advantage, which emphasizes the role of heterogeneous knowledge resources in enabling innovation. Digital skills constitute a critical subset of such resources, and their diversity—spanning software development, data analytics, IoT engineering, cybersecurity, UX design, and cloud architecture—provides the combinatorial building blocks necessary for embedding digital features into physical products.

The second strand is recombinant innovation theory, which posits that novel products emerge from the recombination of existing knowledge elements in new configurations. In this view, a workforce with a diverse portfolio of digital competences is better equipped to generate digitalized product innovations than one characterized by deep specialization in a single skill domain. This leads to our first hypothesis: **H1:**

Digital skill diversity, rather than skill intensity alone, positively predicts the degree of product digitalization in manufacturing firms.

The third strand concerns regional innovation systems (RIS) and the role of institutional thickness, knowledge spillovers, and agglomeration externalities in shaping firm-level outcomes. Regions endowed with universities, research centres, technology parks, and dense inter-firm networks offer richer opportunities for knowledge recombination and digital capability development. We therefore posit: **H2:** *The positive effect of skill diversity on product digitalization is stronger in regions with denser innovation ecosystems.* Finally, if product digitalization is indeed a mediating mechanism between skills and financial outcomes, we expect: **H3:** *Product digitalization mediates the relationship between digital skill diversity and firm financial performance, and this mediation effect varies across Polish regions.*

3. Data and measurement strategy

We construct an original firm-level panel dataset for the period 2020–2025 by integrating three complementary data sources. The first source consists of online job postings scraped from major Polish recruitment platforms (Pracuj.pl, NoFluffJobs, JustJoinIT, and OLX Praca). Each posting is mapped to the employer firm and classified according to the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy using a combination of keyword matching and natural language processing (NLP) techniques. From the classified postings, we derive two key firm-level indicators: **digital skill intensity**, measured as the proportion of postings requiring at least one digital competence, and **digital skill diversity**, captured by Shannon entropy and the Herfindahl–Hirschman Index (HHI) computed over the distribution of demanded ESCO digital skill categories. These firm-level indicators are subsequently aggregated to the NUTS-2 regional level.

The second data source involves systematic web content analysis of corporate websites. For each manufacturing firm in our sample, we retrieve product and service descriptions from both current pages and historical snapshots obtained via the Wayback Machine, enabling longitudinal tracking. Rather than relying on general digital sentiment scores, we construct a **multidimensional product digitalization index** by identifying concrete digital features in product descriptions. The index comprises six dimensions: (i) IoT connectivity and sensor integration, (ii) AI and machine learning components, (iii) cloud-based services and platforms, (iv) process automation and robotics references, (v) data analytics and visualization capabilities, and (vi) digital user interfaces (apps, dashboards, APIs). Each dimension is scored on a binary or ordinal scale, and the composite index is constructed using principal component analysis (PCA) to allow data-driven weighting.

The third source provides firm-level financial data extracted from the Polish National Court Register (Krajowy Rejestr Sądowy, KRS) and the EMIS database. Key performance measures include revenue, value added, return on assets (ROA), and labour productivity. We supplement these with regional-level contextual variables drawn from GUS (Statistics Poland) and Eurostat, including R&D expenditure, patent applications, broadband penetration, tertiary education rates, and the presence of technology parks and clusters, which serve as proxies for regional innovation ecosystem density.

4. Empirical strategy

Our identification strategy combines several approaches to address endogeneity and selection concerns inherent in observational studies of skills and innovation. The baseline specification employs panel regression models with firm fixed effects to control for time-invariant unobserved heterogeneity, and year fixed effects to absorb common macroeconomic shocks. All skill measures enter with a one-year lag to mitigate reverse causality—the concern that firms digitalize products first and then post skill-seeking vacancies.

To further strengthen causal inference, we implement coarsened exact matching (CEM), pairing treated firms (those exhibiting above-median growth in skill diversity) with control firms matched on sector, size, age, region, and baseline performance. Difference-in-differences (DiD) estimates on the matched sample provide a complementary robustness check. For the mediation analysis, we follow a structural equation modelling (SEM) approach, decomposing the total effect of skill diversity on financial performance into a direct path and an indirect path channelled through product digitalization. Regional moderation is tested by interacting skill diversity with regional innovation ecosystem indicators and by estimating subgroup models for regions classified into innovation tiers.

5. Tentative results and expected contributions

Preliminary descriptive analysis reveals pronounced regional heterogeneity in both digital skill portfolios and product digitalization levels across Polish voivodeships. Mazowieckie, Małopolskie, and Dolnośląskie regions exhibit the broadest and deepest digital skill bases, while eastern regions such as Podlaskie, Lubelskie, and Warmińsko-Mazurskie lag substantially behind. Initial regression results support H1: skill diversity measures (Shannon entropy) emerge as stronger and more consistent predictors of product digitalization than skill intensity alone, suggesting that the breadth of the competence portfolio matters more than its aggregate volume.

The interaction terms in our moderation analysis provide tentative support for H2, indicating that the skill diversity–product digitalization link is significantly amplified in regions with above-median R&D intensity and technology park presence. Regarding H3, preliminary SEM estimates indicate a statistically significant indirect effect of skill diversity on financial performance through product digitalization, though the magnitude of this mediation varies considerably across regional groupings. These results—to be refined as the full dataset is assembled—suggest that addressing the digital skills gap requires attention not only to the *quantity* of digital competences but also to their *compositional diversity*, and that place-based policies need to account for the absorptive capacity of regional innovation systems.

6. Policy implications and conclusions

The findings carry several implications for regional development policy. For EU cohesion policy and national smart specialization strategies, the results underscore the importance of fostering diverse digital skill ecosystems rather than concentrating resources on narrow occupational profiles. Regions that succeed in cultivating a broad spectrum of digital competences—from data science and AI to UX design and cybersecurity—appear better positioned to translate human capital investments into digitally enriched manufacturing output. Moreover, the significant moderating role of regional innovation infrastructure suggests that skill investments alone may be insufficient; complementary investments in R&D facilities, technology transfer institutions, and inter-firm networking platforms are needed to fully unlock the potential of a diverse digital workforce. The paper concludes by discussing limitations related to web-scraped data representativeness and avenues for extending the framework to other EU member states undergoing similar digital transitions.

Keywords: *digital skills, product digitalization, manufacturing, regional disparities, ESCO, job postings, web content analysis, recombinant innovation*

JEL Codes: J24, L60, O33, R11, R12

References (selected)

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3–30.

Brynjolfsson, E. & McAfee, A. (2014). *The Second Machine Age*. W. W. Norton & Company.

Capello, R. & Lenzi, C. (2019). Regional innovation patterns from an evolutionary perspective. *Regional Studies*, 53(8), 1117–1132.

Foray, D. (2015). *Smart Specialisation: Opportunities and Challenges for Regional Innovation Policy*. Routledge.

Frenken, K., Van Oort, F. & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685–697.

Hidalgo, C. A. & Hausmann, R. (2009). The building blocks of economic complexity. *PNAS*, 106(26), 10570–10575.

Iacus, S. M., King, G. & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1–24.

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

Rodriguez-Pose, A. & Crescenzi, R. (2008). Research and development, spillovers, innovation systems, and the genesis of regional growth. *Regional Studies*, 42(1), 51–67.

Spiezia, V. (2011). Are ICT users more innovative? An analysis of ICT-enabled innovation in OECD firms. *OECD Journal: Economic Studies*, 2011(1), 99–119.