# Tracking Upskilling and Reskilling via Skill Evolution in GitHub Profile READMEs


**Maria Papoutsoglou**[1] **Emmanouil Rigas**[2] **Nikolaos Gkaripis**[1] **Georgios Meditskos**[1] **Georgia Kapitsaki**[3]

**Efstratios Kontopoulos**[4] **Georgios Koutalieris**[5] **Nick Bassiliades**[1] **Christoph Treude**[6]

[1] *School of Informatics, Aristotle Univ. of Thessaloniki* [2] *School of Medicine, Aristotle Univ. of Thessaloniki* [3] *Computer Science Dept., University of Cyprus*

[4] *Choreograph* [5] *Enora Innovations* [6] *Singapore Management University*

mpapo@csd.auth.gr


## 1 Introduction

The rapid diffusion of **generative AI** is reshaping skill requirements across occupations. New practices, such as prompt engineering, AI-assisted coding and hybrid human–AI workflows, are emerging faster than traditional labour-market monitoring mechanisms can capture.

Online Job Advertisement (OJA) analysis has enabled demand-side monitoring of labour-market trends [1][2]. However, supply-side skill evolution remains harder to observe, because access to CVs and individual professional profiles is limited.

Open software engineering communities, such as GitHub, provide a valuable complementary source [3]. GitHub profile READMEs function as **public, self-curated CV-like documents** where developers describe their skills, interests and professional identity.

Because these profiles can be updated continuously, they offer a dynamic window into emerging skills, upskilling and reskilling processes. This is particularly important for AI-related skills, which may appear in practice before they are formally captured in official classifications or labour-market statistics.

In this work, we show how raw community content from GitHub can detect both ESCO-aligned and non-ESCO skills, revealing that the emerging non-ESCO skills are predominantly AI-related competences evolving rapidly in developer profiles — providing valuable signals for updating official EU labour-market ontologies such as ESCO [5].

## 2 Related Work & Theoretical Framework

Web data sources have transformed labour-market analytics by enabling near real-time observation of occupational trends. Previous work has shown that Online Job Advertisements can analyse skills demand in domains such as the green economy and electric vehicles [1] and regional software labour markets [2].

However, the emergence of AI-related skills creates a new challenge. Skills such as **prompt engineering, vibe coding, LLM-assisted development, bias detection** and AI-enabled problem solving evolve rapidly and affect both hard and soft skills.

**Prompt engineering** is particularly relevant because it is no longer restricted to software developers. Efficient prompting is becoming a cross-sector competence required in education, health, cybersecurity, business and research. It includes technical skills (AI-assisted coding) but also cognitive skills (problem formulation, critical evaluation of LLM outputs, hallucination reduction, bias awareness).

The strong growth of prompt-engineering-related publications in Scopus [4] indicates this is not a marginal trend — it reflects the rapid consolidation of prompting as an emerging competence area.

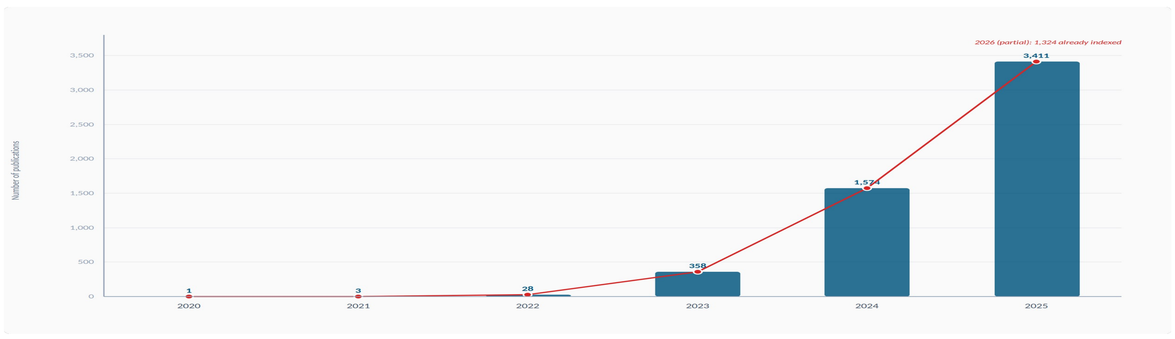



*Figure 1. Yearly evolution of Scopus-indexed publications related to "prompt engineering" after 2020. The trend shows a rapid increase from 1 record in 2020 to 3 in 2021, 28 in 2022, 358 in 2023, 1,574 in 2024 and 3,411 in 2025. The year 2026 is excluded from the plot because it is incomplete; however, 1,324 records had already been indexed, suggesting continued growth.*

***The steep increase after 2022 signals how quickly prompt engineering became a recognised research and professional topic — AI-related skills evolve faster than taxonomies and official statistics can be updated.***

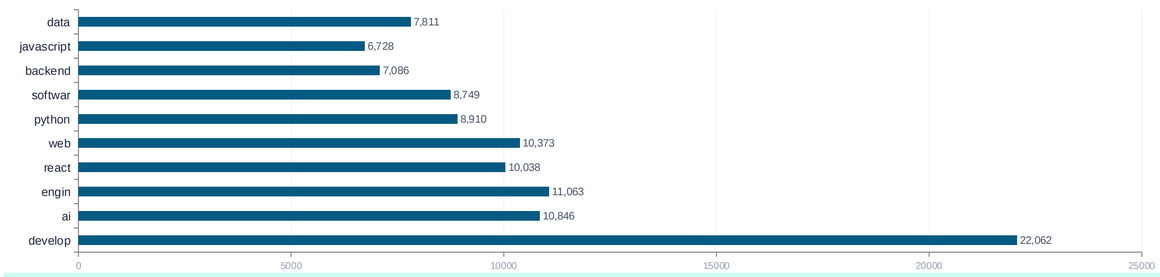
## 3 Data Collection & NLP Pipeline

1. **Profile collection**
   37,898 developer READMEs with revision histories from 195 countries
2. **Commit history extraction**
   GitHub API to retrieve all README revisions with timestamps
3. **Content retrieval**
   README text at each commit SHA for longitudinal tracking
4. **NLP skill extraction**
   TF-IDF (unigrams, bigrams, trigrams) + named entity recognition
5. **ESCO alignment**
   Fuzzy matching against ESCO labels; RDF graph traversal for hierarchy
6. **Skill evolution analysis**
   Temporal tracking: upskilling (within-branch) vs reskilling (cross-branch)

**37,898** READMEs · **195** countries · **Full revision histories**

## 4 Skill Landscape: TF-IDF Analysis

*Top skills by term frequency across 20,059 profiles*

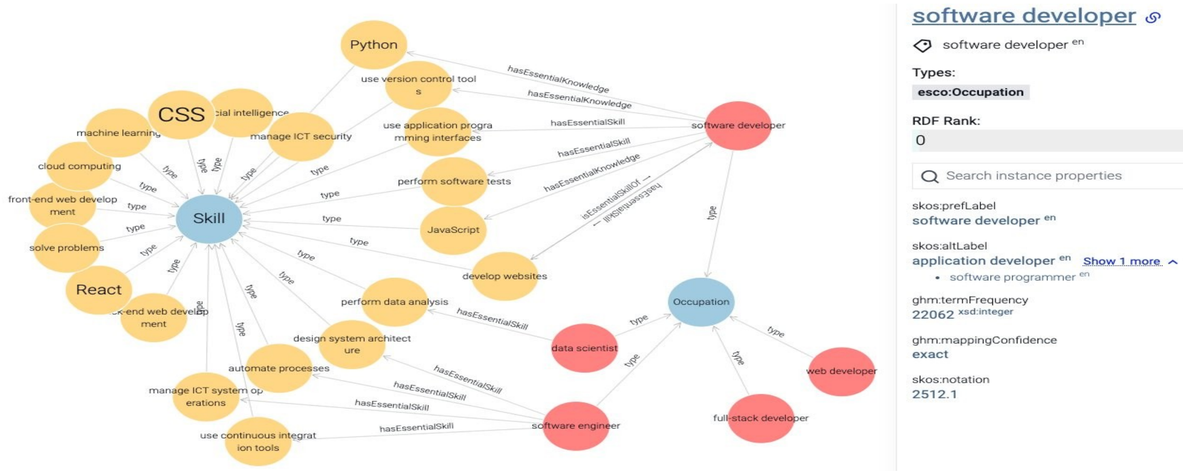



**IDF analysis reveals hidden emerging skills:**

AI agents: 741 profiles · LLMs: 654 · AI-powered dev: 554
RAG: 302 · Prompt engineering: 97 · LangChain: 264
AI engineer (role): 244 · Fine-tuning: 91 · LLMOps: 12

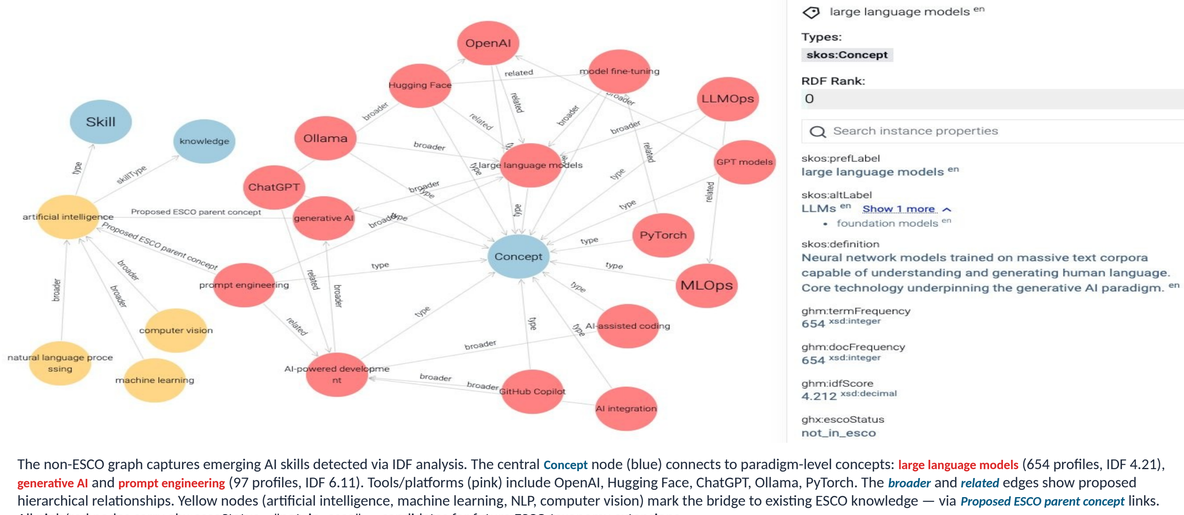
## 5 Skill Mapping: ESCO Alignment & Emerging Skills

**ESCO-aligned knowledge graph**




The ESCO-aligned graph maps extracted skills to official ESCO v1.2 instances. Occupations (pink) such as software developer (TF: 22,062, ISCO 2512.1) are linked to essential skills (yellow) via ***hasEssentialSkill*** relations — e.g. develop websites, perform software tests, use APIs. Knowledge concepts (yellow) such as Python, JavaScript, CSS connect via ***hasEssentialKnowledge***. The central Skill node (blue) anchors the SKOS type hierarchy. ~75% of top TF terms map to existing ESCO concepts.

**Emerging skills NOT in ESCO v1.2**




The non-ESCO graph captures emerging AI skills detected via IDF analysis. The central **Concept** node (blue) connects to paradigm-level concepts: **large language models** (654 profiles, IDF 4.21), **generative AI** and **prompt engineering** (97 profiles, IDF 6.11). Tools/platforms (pink) include OpenAI, Hugging Face, ChatGPT, Ollama, PyTorch. The ***broader*** and ***related*** edges show proposed hierarchical relationships. Yellow nodes (artificial intelligence, machine learning, NLP, computer vision) mark the bridge to existing ESCO knowledge — via ***Proposed ESCO parent concept*** links. All pink/red nodes carry ghx:escoStatus = "not_in_esco" — candidates for future ESCO taxonomy extensions.

## 6 Key Findings & Discussion

- GitHub is a promising hub for observing supply-side skill evolution, especially in fast-changing domains such as AI. Developers are highly exposed to technological change and often engage in continuous learning and training.
- The strong increase in prompt-engineering-related publications (Figure 1) reinforces the importance of monitoring emerging AI skills early. New competence areas can become visible and professionally important within very short time windows.
- Open software engineering communities remain valuable for labour-market intelligence. Even as LLMs change development practices, human participation in open knowledge communities continues to generate signals about skills, tools and professional adaptation.
- The distinction between ESCO and non-ESCO skills is especially useful. ESCO-aligned skills support comparability and structure, while non-ESCO skills highlight emerging competences that may require future classification updates.
- AI-related skills may emerge, spread and become professionally relevant before they are fully represented in official labour-market statistics, occupational classifications or skills ontologies — dynamic web sources such as GitHub can provide early supply-side evidence.

## 7 Conclusions & Future Directions

### Conclusions

- This work demonstrates how GitHub profile READMEs can support dynamic skills intelligence by capturing self-reported skill evolution in developer communities.
- The proposed approach contributes to upskilling and reskilling analysis by linking profile changes to ESCO-based competence structures and identifying emerging non-ESCO AI skills.
- The rapid growth of prompt engineering illustrates why dynamic supply-side data sources are needed — AI-related skills may emerge and become professionally relevant before official classifications capture them.

### Future directions

- Construct a historical dataset from GitHub profile README commits to analyse when specific AI skills first appeared and how developer portfolios changed before and after the rise of generative AI.
- Deeper validation of emerging skill detection and enrichment of the non-ESCO skill graph.
- Comparison with demand-side evidence from Online Job Advertisements for supply–demand gap analysis.
- Cross-country analysis of upskilling/reskilling patterns to identify regional digital divides.

## 8 References


[1] Papoutsoglou, M., Rigas, E. S., Kapitsaki, G. M., Angelis, L., & Wachs, J. (2022). Online labour market analytics for the green economy: The case of electric vehicles. Technological Forecasting and Social Change, 177, 121517.
[2] Kapitsaki, G., Chatzivasili, L., Papoutsoglou, M., & Galster, M. (2024). An exploratory study on soft skills present in software positions in Cyprus. Proc. 18th ACM/IEEE ESEM, 200–211.
[3] GitHub Blog. How researchers are using GitHub Innovation Graph data to reveal the digital complexity of nations.
[4] Scopus bibliometric search. Elsevier. https://www.scopus.com
[5] European Commission. ESCO: European Skills, Competences, Qualifications and Occupations. https://esco.ec.europa.eu



**Acknowledgements**

This work is supported by Enora Innovation, an agile SME developing cutting-edge IT products and services. Founded in 2019, Enora focuses on innovative solutions in e-health, agri-food, mobility, transport and environmental sustainability.



Contact: **Maria Papoutsoglou — mpapo@csd.auth.gr**