

Harnessing web data for next-generation skills intelligence in the UK

Dr Jeisson Arley Cárdenas Rubio
Cambridge Econometrics

Introduction

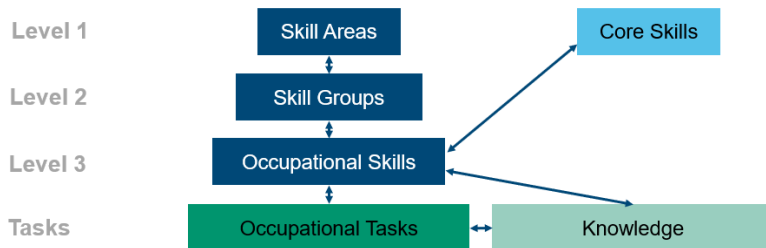
- ▶ Why do skills matter in the labour market?
 - ▶ They enhance employability and support adaptation to technological change. They drive productivity, innovation, and economic growth (Green, 2011).
- ▶ Main challenges in analysing skills data:
 - ▶ Diversity of data and lack of a common language: Skills are described in many ways across different sources (CVs, surveys, etc.).
 - ▶ Rapid changes: The skills in demand evolve quickly due to technology and economic shifts.
 - ▶ Localised information: Needs can vary significantly by region or sector.
 - ▶ Limitations of traditional sources: Occupational surveys or censuses have low update frequency and low granularity.

- ▶ Relevance of a skills taxonomy:
 - ▶ It provides a common language for describing and comparing skills across sectors, employers, and education providers.
 - ▶ It enables better tracking of emerging skills, declining skills, and changing demand in the workforce.
- ▶ Objective:
 - ▶ Develop a UK Standard Skills Classification (SSC): a national framework that standardises how skills are defined and organised (Commissioned by Skills England).
- ▶ How:
 - ▶ The UK skills taxonomy was developed through stakeholder engagement, expert interviews, and the application of AI, specifically Large Language Models (LLMs).

UK Skills Classification

- ▶ A skill is defined as a capability enabling the competent performance of an occupational task (e.g., driving a car, monitoring financial data, cleaning windows).

Hierarchy:



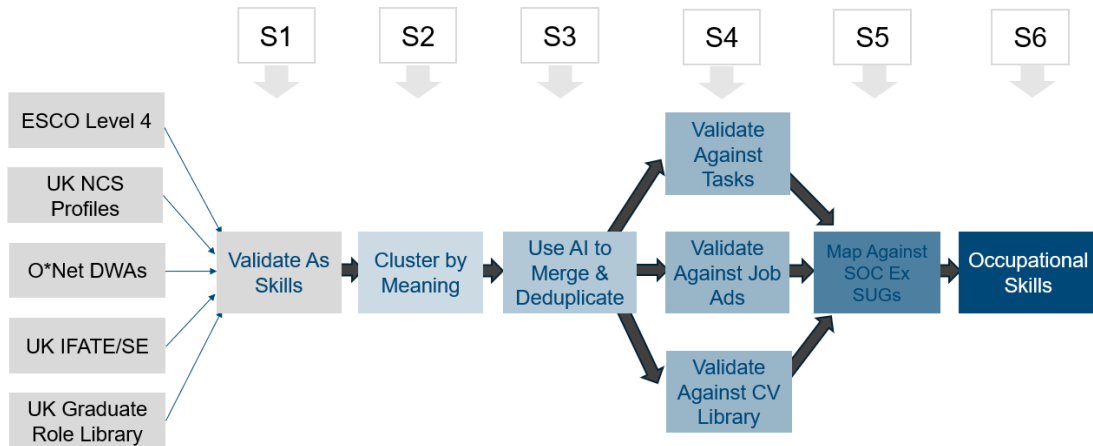
LLMs and Standard Skills Classification

- ▶ Use of LLMs in the SSC
 - ▶ Deduplication of skills: LLMs were used to identify and reconcile duplicate or overlapping skill descriptions across different systems in the UK, improving consistency and integration.
 - ▶ Skills analysis: LLMs were applied to identify and classify skills described in job postings, existing occupational and skills descriptions, and other labour market texts.
 - ▶ Clustering and grouping: LLMs supported the grouping of related skills, helping to build the hierarchical structure of the taxonomy.
 - ▶ Labelling of skill groups: LLMs were used to generate consistent and interpretable labels for clusters of related skills.
 - ▶ Validation support: AI-assisted comparisons between occupations and sectors provided additional checks on the coverage and coherence of the taxonomy.

Text comparison

Baseline Text	Comparison Text	Similarity Score
Validate data at all stages of ETL	Validate data at all stages of Extract, Transform, Load	0.91
Analyse communication using NLP	Analyse communication using Natural Language Processing	0.9
Analyse communication using NLP	Analyse communication using Neuro-Linguistic Programming	0.84
Maintain storage conditions	Maintain storage conditions such as correct temperature, lighting, ventilation, and packaging	0.81
Cut and prepare vegetables	Cook vegetables	0.67
Manage construction projects	Manage projects	0.71
Develop applications in Python	Develop applications in Django	0.79
Develop applications in Python	Develop applications in Laravel	0.58

Processing Steps. Occupational skills - 2024



Prompt Design: example – skill extraction from job postings

You are a labour market researcher extracting occupational skills from job descriptions.

An occupational skill begins with an action-based verb followed by a specific noun.

An occupational skill describes a capability and not an attitude or outcome and is developed and not innate. An occupational skill is specific and unambiguous and no more than nine words long.

Please analyse the text between <> and identify occupational skills.

1. Read the entire text and determine if it contains occupational skills.
2. If occupational skills are found, extract and enumerate these phrases, ensuring they are direct excerpts from the text.
3. If no occupational skills are found, simply answer with "NO occupational skills".
4. End your response after listing the relevant phrases or stating the absence thereof. Do not include any other text or introduce the list.

Example response format:

- If occupational skills are present:

"

1. Extracted occupational skill
2. Extracted occupational skill
3. Extracted occupational skill

..."

- If no occupational skills are found:

"NO occupational skills."

Text:<{description}>

Sample – skill extraction from online job postings

1. Ensure financial sustainability
2. Work towards reaching its vision
3. Lead the charity in strategic development
4. Maintain commitment of staff
5. Manage organisational operations
6. Line manage staff
7. Build effective teams
8. Manage complex situations
9. Develop strategy and communicate it clearly

ceo we have an exciting opportunity for a female ceo to represent an organisation externally ensure financial sustainability and work towards reaching its vision for a world in which survivors are free from the fear and experience of sexual violence and supported when it does occur this post is restricted to women applicants only and is exempt under schedule 9 part 1 of the equality act 2010 location oxford with some flexibility to work from homesalary 50k 55kcontract permanenthours full time 37 hours per weekbenefits 28 days of annual leave up to 33 with long service we offer a 5 pension contribution enhanced sickness and parental leave pay regular clinical supervision employee assistance programme and a generous training budget to support your learning and development please note that interviews will be taking place on a rolling basis so apply early to avoid disappointmentabout the role join a team of dynamic caring and committed women working together to make a difference to survivors of sexual violence in oxfordshire you will lead the charity in the strategic development of its vision of a world in which survivors are free from the fear and experience of sexual violence and the effective delivery of its services maintaining the commitment of staff and the mission and ethos of the organisation duties and key responsibilities include vision and leadershipgovernancerelationship managementorganisational managementline management of staff about you the ideal candidate will have experience of managing people will be a strategic thinker and have an understanding of the impacts of sexual violence on survivors and the wider community knowledge of potential funding sources and the current funding environment for charitable organisations including commissioning and sector wide opportunities is essential you will have relevant experience in the not for profit sectormanaging service delivery within a voluntary or community or statutory settingleading organisations with multiple stakeholders and relationshipsleadership and line management of staff building effective teams and managing complex situations developing strategy and communicating it clearly by setting a clear directionstrategic and business development planning and implementing organisational changeoversight of financial management systems and budgets other roles you may have experience of could include ceo chief executive chief executive officer interim ceo interim chief executive interim chief executive officer coo managing director head of director assistant ceo assistant chief executive assistant chief executive officer

Summary of the classification

► Classification of 3,343 skill concepts and 22 skill domains

ID	Domain	Skills
SD.01	Researching & Analysing	184
SD.02	Designing & Creating	243
SD.03	Planning & Resourcing	169
SD.04	Programming & Implementing (Digital Tools & Systems)	170
SD.05	Constructing & Installing	186
SD.06	Manufacturing & Processing	62
SD.07	Operating & Monitoring	332
SD.08	Planting & Growing	43
SD.09	Breeding & Nurturing (Animals)	50
SD.10	Diagnosing & Treating (Healthcare)	238
SD.11	Serving & Caring	96
SD.12	Advising & Supporting	203
SD.13	Educating & Training	200
SD.14	Evaluating & Inspecting	351
SD.15	Accounting & Financing	84
SD.16	Managing & Directing	187
SD.17	Marketing & Selling	50
SD.18	Communicating & Performing	130
SD.19	Recording & Documenting	90
SD.20	Handling & Transporting	96
SD.21	Maintaining & Repairing (Physical Objects)	144
SD.22	Cleaning & Restoring	35

Some preliminary results

Most frequently mentioned skills
Build relationships with clients
Maintain operational records
Respond to customer enquiries
Maintain client records
Manage stakeholder relationships
Identify and implement organisational process improvements
Manage staff recruitment
Conduct quality audits
Work within multidisciplinary healthcare teams
Support students with special educational needs
Resolve customer complaints

- Strengths and weaknesses: LLMs make it possible to efficiently extract and organise skills at scale and across languages, though some inconsistencies and biases still require manual checks.

Skills transferability

11. Corporate Managers & Directors	29	26	14	9	15	23	13	13	16	9	25	27	25	5	5	5	5	9	16	17	20	24	6	7	4	9
12. Other Managers & Proprietors	29	31	12	10	15	20	12	14	14	11	25	29	28	11	7	7	7	12	22	18	23	25	8	10	8	14
21. Science, Research, Engineering & Technology Professionals	15	12	23	8	9	17	22	7	8	9	14	11	13	7	13	8	6	5	7	8	9	14	10	5	6	5
22. Health Professionals	8	9	7	39	14	8	8	27	7	6	9	10	19	4	4	2	2	21	8	10	8	13	3	4	3	5
23. Teaching & Other Educational Professionals	14	13	7	12	63	11	5	22	5	12	13	14	19	3	2	2	2	20	9	10	8	17	2	7	2	6
24. Business, Media & Public Service Professionals	21	16	14	9	11	23	13	13	10	10	21	21	20	5	4	5	5	8	11	13	18	17	4	3	3	6
31. Science, Engineering & Technology Associate Professionals	13	12	19	8	6	14	21	6	9	9	13	11	15	7	16	10	9	6	9	8	9	14	13	5	8	7
32. Health & Social Care Associate Professionals	9	10	5	23	18	10	5	24	7	7	10	12	17	4	3	2	3	19	10	13	9	13	3	4	3	6
33. Protective Service Occupations	14	11	7	7	5	10	7	8	35	3	11	12	14	5	5	5	2	8	14	30	7	16	5	11	4	11
34. Culture, Media & Sports Occupations	7	8	6	5	11	8	6	7	3	17	7	5	7	3	5	3	6	6	7	4	7	6	4	3	2	4
35. Business & Public Service Associate Professionals	22	19	12	9	12	21	12	12	12	8	22	24	25	6	6	5	5	9	15	16	22	22	6	7	5	9
41. Administrative Occupations	20	19	8	7	11	18	9	11	10	5	20	31	34	4	3	3	4	8	16	15	23	23	4	6	4	10
42. Secretarial & Related Occupations	13	15	6	9	9	12	8	10	7	6	14	25	39	5	3	3	4	9	15	11	16	19	4	5	5	9
51. Skilled Agricultural & Related Trades	7	11	7	4	4	6	8	5	6	4	7	7	11	30	7	7	5	9	7	6	6	7	9	9	21	7
52. Skilled Metal, Electrical & Electronic Trades	6	7	11	3	2	5	17	4	6	6	6	6	9	7	34	15	14	4	7	7	7	6	23	9	13	8
53. Skilled Construction & Building Trades	5	6	7	2	2	5	10	2	6	4	5	4	5	6	15	27	11	3	5	5	4	5	17	5	13	6
54. Textiles, Printing & Other Skilled Trades	3	4	4	2	1	3	6	2	2	5	4	4	6	4	10	8	20	2	4	2	6	3	13	3	7	7
61. Caring Personal Service Occupations	6	8	3	16	16	5	4	18	7	5	7	8	14	6	3	2	3	19	8	9	6	10	3	5	4	6
62. Leisure, Travel & Related Personal Service Occupations	11	14	4	7	6	7	6	9	13	6	11	15	19	5	4	4	4	9	20	17	17	20	5	9	4	12
63. Community & Civil Enforcement Occupations	15	15	6	10	8	13	7	18	32	4	14	21	25	5	8	5	3	13	23	44	16	27	6	15	6	16
71. Sales Occupations	15	16	5	6	5	13	7	8	6	7	18	23	23	4	5	2	6	6	17	11	34	21	4	8	4	12
72. Customer Service Occupations	12	12	7	8	9	10	7	9	9	4	12	16	20	3	3	2	2	8	15	14	17	19	3	6	3	7
81. Process, Plant & Machine Operatives	6	7	8	3	2	5	13	3	6	5	6	6	8	7	23	16	17	4	6	6	6	6	24	11	16	12
82. Transport & Mobile Machine Drivers & Operatives	4	6	3	3	8	2	4	4	8	3	6	7	9	5	7	3	3	5	8	10	7	9	7	23	7	12
91. Elementary Trades & Related Occupations	3	6	4	2	2	3	6	3	4	2	4	4	7	16	10	12	8	5	4	5	4	4	13	10	14	9
92. Elementary Administration & Service Occupations	4	7	2	3	3	3	3	4	7	3	5	7	9	3	4	3	5	5	9	8	8	7	6	9	6	11
	11	12	21	22	23	24	31	32	33	34	35	41	42	51	52	53	54	61	62	63	71	72	81	82	91	92

Strengths and weaknesses

Strengths

- ▶ Time efficiency: Producing a skills taxonomy within the required deadlines would be extremely difficult without AI support.
- ▶ Language understanding: “Oversowing, top dressing, aeration and marking lines,” “Reseeding lawn,” “Top dressing lawn,” “Lawn aeration”.
- ▶ Processing, standardising, and classifying large volumes of textual data. 90,000 job descriptions to extract skills. 500,000 skills extracted.
- ▶ Adaptable to diverse country contexts: Tests conducted in Spanish contexts have shown similar success.

Weaknesses

- ▶ May add new meaning when rewriting skill-related phrases; bias towards US English; manual verification is still required (e.g. “keep prisoners records” → “keep client records”)!

Thanks!