



Department  
for Education

# **AI analysis of local skills improvement plans**

**Using large language model (LLM)  
technologies to analyse local skills  
improvement plans (LSIPs)**

**November 2023**

# Contents

List of figures	4
List of tables	5
Acknowledgements	6
Executive summary	7
1 Introduction	9
1.1 Local skills improvement plans	9
1.2 Analysis	10
2 Methodology	11
2.1 Structure of the LSIP reports	11
2.2 Individual summaries	11
2.2.1 Summarisation pipeline	11
2.2.2 Refinement of LLM prompts	12
2.2.3 Validation	12
2.2.4 Response from employer representative bodies	13
2.3 Topic identification analysis	13
2.3.1 Predefined topic analysis	13
2.3.2 Self-supervised analysis	14
3 Sectors, transferable skills and cross-cutting themes important to LSIP areas	16
3.1 Sector and thematic view	17
3.1.1 Manufacturing and Engineering	18
3.1.2 Construction	20
3.1.3 Tourism and Hospitality	22
3.2 Growth priorities spanning multiple sectors	24
3.2.1 Green Skills and Net Zero	24
3.2.2 AI and Automation	26
3.3 Transferable skills	28
3.3.1 ICT and digital literacy skills	28
3.3.2 Leadership skills	31
4 Further validation work	33
4.1 Extensions and automated validation checks	33
4.1.1 To identify anything missing	33

4.1.2	To ensure that no key ideas have been hallucinated	33
4.1.3	Semantic scores	34
Annex A	– LSIP areas	35
Annex B	– Example of a relatively poor individual summary	38
Annex C	– Example of a relatively accurate individual summary	39

## List of figures

Figure 1: Strength of discussion on manufacturing and engineering across LSIP areas .	19
Figure 2: Word cloud of the 100 most common words across the reports relating to manufacturing and engineering .....	19
Figure 3: Strength of discussion on construction across LSIP areas .....	21
Figure 4: Word cloud of the 100 most common words across the reports relating to construction .....	21
Figure 5: Strength of discussion on tourism and hospitality across LSIP areas.....	23
Figure 6: Word cloud of the 100 most common words across the reports relating to tourism and hospitality .....	23
Figure 7: Strength of discussion on green skills and net zero across LSIP areas.....	25
Figure 8: Word cloud of the 100 most common words across the reports relating to green skills and net zero .....	26
Figure 9: Strength of discussion on AI and automation across LSIP areas .....	27
Figure 10: Word cloud of the 100 most common words across the reports relating to AI and automation .....	28
Figure 11: Strength of discussion on ICT and digital literacy skills across LSIP areas ....	30
Figure 12: Word cloud of the 100 most common words across the reports relating to ICT and digital literacy skills .....	30
Figure 13: Strength of discussion on leadership skills across LSIP areas .....	32
Figure 14: Word cloud of the 100 most common words across the reports relating to leadership skills .....	32

## List of tables

Table 1: Top 10 clusters identified and named using self-supervised AI techniques .....	14
Table 2: Top 10 sectors covered by LSIPs .....	16
Table 3: Top 10 of sectors extracted across LSIP reports (not ordered) .....	17
Table 4: Top 5 LSIP areas that discuss manufacturing and engineering, average and range across all areas highlighting as priority .....	18
Table 5: Top 5 LSIP areas that discuss construction, average and range across all areas highlighting as priority .....	20
Table 6: All LSIP areas that discuss tourism and hospitality, average and range across all areas highlighting as priority .....	22
Table 7: Top 5 cross-cutting issues covered in LSIPs .....	24
Table 8: Top 5 LSIP areas that discuss green skills and net zero, average and range across all areas.....	24
Table 9: Top 5 LSIP areas that discuss AI and automation, average and range across all areas.....	26
Table 10: Top 5 transferable skills covered in LSIPs .....	28
Table 11: Top 5 LSIP areas that discuss ICT and digital literacy skills, average and range across all areas.....	29
Table 12: Top 5 LSIP areas that discuss leaderships skills, average and range across all areas.....	31
Table 13: List of LSIP areas.....	35

## **Acknowledgements**

The authors would like to thank Jordan Juritz and the members of the Faculty Fellowship programme at Faculty for their expertise developing and implementing the cutting-edge AI approaches presented in this report. They would also like to thank all those in the employer representative bodies from each of the Local Skills Improvement Plan areas who supported the analysis and provided valuable feedback to help validate the findings.

## Executive summary

The recent advances in artificial intelligence (AI), and in particular large language models (LLMs), bring the potential to gather new and rich information from textual documents that would be otherwise too time-consuming or unstructured to analyse. This research trials the use of AI techniques on a series of reports about local skills needs across different areas of England to synthesise and summarise the main themes across these reports.

The AI approaches in this report were developed and implemented as part of the Faculty Fellowship<sup>1</sup>, and the subsequent report and findings have been taken from the database produced as part of that project.

### Results should be interpreted with caution

This analysis is the output of a pilot project applying experimental AI extraction and generation techniques to a set of written reports which constitute a large volume of unstructured data.

Although there is considerable opportunity to use these or similar techniques in a wide range of applications, consideration should be given to the limitations of this emerging technique.

The Local Skills Improvement Plan reports that are summarised were drafted with different formats and approaches, e.g. some areas took a sector-based approach, some focussed on cross cutting skills and did not identify any specific sectors, others specifically excluded some sectors as they are discussed elsewhere. The absence of focus on a particular topic does not mean this is not seen as important, merely it did not feature extensively in the report.

The main findings of the pilot show

- There is good potential for the use of AI to draw insights from a series of unstructured textual reports, but this must be accompanied by appropriate expert validation to ensure an accurate interpretation of results.
- Combining two different approaches to extraction improved the quality of the resulting database. This mixed approach - one applying a pre-defined list of topics of interest and another AI-led self-supervised method to extract novel topics - ensured coverage of key topics as well as more novel topics.
- The use of AI to produce short, one-page summaries of the Local Skills Improvement Plan (LSIP) reports showed mixed results, manual review indicated

---

<sup>1</sup> <https://faculty.ai>

that 75% of summaries were accurate, but some lacked key details, and there was some evidence of hallucinations<sup>2</sup> across at least 3 of the 38 the summaries.

- There was good alignment between the priority sectors identified by the AI-led approach, manual interrogation of the reports and those covered by national skills priorities, though not complete alignment.
- Prompts to extract and summarise the LSIP reports were adjusted through the pilot project based on feedback from policy officials in conversation with the researcher and Unit for Future Skills analysts. Indicating that the success of generative AI is directly related to the quality of prompts provided.

As a foundation for further research and collaboration, the current analysis and resulting structured database allows quick comparison on the focus of specific LSIP areas on topics of interest. This localised information can be used to cross-check local interventions and identify further local needs. This project and its use of AI techniques is just one way the Department for Education is analysing the content of LSIPs; the department continues to engage with employer representative bodies.

---

<sup>2</sup> AI hallucinations are a phenomenon where an artificial intelligence (AI) model produces a confident response that is not based on real data or events.



# 1 Introduction

Advances in artificial intelligence (AI), and in particular large language models (LLMs) bring the potential to gather new and rich information from textual documents that would be otherwise too time-consuming or unstructured to analyse. This research trials the use of AI techniques on a series of reports about local skills needs across different areas of England to synthesise and summarise the main themes across these reports.

The AI approaches in this report were developed and implemented as part of the Faculty Fellowship<sup>3</sup>, and the subsequent report and findings have been taken from the database produced as part of that project.

The findings in this analysis should be taken with caution and should be treated in the context that they were generated, i.e. in testing the application of AI-led approaches to summarise information within and across related documents. These results should be taken as a guide, and any conclusions checked by referencing the original LSIPs before being used to inform decision-making.

## 1.1 Local skills improvement plans

In autumn 2022, employer representative bodies (ERBs) were designated to lead the development of local skills improvement plans (LSIPs) across 38 areas of England<sup>4</sup>. These plans were published in August 2023 and can be found on the designated ERB websites<sup>5</sup>.

LSIPs contain a wealth of intelligence regarding skills needs (described in a variety of ways including sectors, occupations or cross-cutting/transferable skills), proposed local changes to help address the skills needs, and the operation of the Department for Education's (DfE) skills policies and programmes.

While all LSIP areas followed the process set out in the statutory guidance, the contents of the LSIPs reflect the flexibility and autonomy for ERBs to identify and define skills priorities for their locality so vary in structure and style. A range of data sources were used to underpin local decisions on priorities, including:

- All LSIPs referenced local labour market data on employment and businesses operating in their area.
- 4 out of every 5 LSIPs (79%) presented findings from bespoke surveys of employers in their local area.

---

<sup>3</sup> <https://faculty.ai>

<sup>4</sup> See Annex A for a list of all LSIPs

<sup>5</sup> <https://www.gov.uk/government/publications/designated-employer-representative-bodies/notice-of-designated-employer-representative-bodies>

- Three-quarters (75%) of LSIPs used data supplied by the Unit for Future Skills<sup>6</sup>, with specific reference to the local skills dashboard<sup>7</sup> and local level employment projections.
- Two thirds (68%) of LSIPs specifically referenced analysis from commercial organisations, e.g. Lightcast, Adzuna.

Although the focus of LSIPs is to provide local intelligence to support providers, the intelligence in the reports is also relevant to policy teams across government, alongside other LSIP areas facing shared challenges.

## 1.2 Analysis

The aims of the analysis were to:

1. Synthesise and summarise the main insights and themes from the LSIP reports on the skills needs identified by employers and the priority action areas identified across the country.
2. Test the operation and value of AI and LLMs in analysing unstructured text and producing useful insights across a series of related documents.

The 8-week pilot study produced outputs including **AI-generated individual summaries** and **topic identification analysis** for individual LSIP reports and across all reports. These are presented in the following sections.

The underlying data that was generated as part of this analysis has been released alongside this report.

---

<sup>6</sup> <https://www.gov.uk/government/groups/unit-for-future-skills>

<sup>7</sup> <https://department-for-education.shinyapps.io/local-skills-dashboard>

## 2 Methodology

The analysis produced two main outputs which are discussed in this report:

1. Individual summaries - AI-generated texts that summarise the discussion contained in individual LSIP reports.
2. Topic identification analysis - AI-generated data that can be used to examine where and how a topic was discussed and compare the discussion of particular topics across the full collection of LSIP reports.

The outputs were generated using a combination of natural language processing (NLP) techniques, such as embeddings, and large language models (LLMs) to analyse the collection of LSIPs.

Recently, tools such as langchain<sup>8</sup> (a python<sup>9</sup> module for chaining together LLM calls) and llama index<sup>10</sup> (a python module for integrating text from documents into LLM calls) have emerged that reduce the barrier to entry to using LLMs and linking them with documents.

### 2.1 Structure of the LSIP reports

Each of the LSIP areas produced a summary document in pdf format, approximately 30 pages in length, that describe the skills priorities for their area. The documents do not follow a set format; they vary in structure and style, include complex, sometimes technical language, and contain lots of information across text, tables, charts, and other graphics as well as references to other reports.

In addition to the main summary 30-page report, there were additional annexes often included as separate documents. Annexes frequently included further detail on the intelligence gathered to inform the report, and as such included a lot of content structured as tables and charts. In total the reports covered around 1,300 pages of text, increasing to around 4,000 including all the additional annexes.

### 2.2 Individual summaries

#### 2.2.1 Summarisation pipeline

The initial stages of the project focussed on developing a prototype document summarisation pipeline using langchain and llama index. This pipeline, known as a MapReduce summary:

---

<sup>8</sup> <https://pypi.org/project/langchain/>

<sup>9</sup> <https://www.python.org/>

<sup>10</sup> <https://pypi.org/project/llama-index/>

- Imported each of the documents and applied cleaning functions
- Summarised small sections of each document, using first OpenAI's GPT-3.5 Model and then OpenAI's GPT-4 Model for better results
- Created a concise and information rich database of 'chunks' of information (the 'database')
- Used prompts to generate a library of short (one-page) summaries for each LSIP by summarising the 'chunks'. These summarisations were reviewed internally by policy officials and subsequent modifications were made to the LLM prompts (discussed below).

## 2.2.2 Refinement of LLM prompts

Without commenting on the truthfulness of the comparisons, the initial summaries for each LSIP varied in length and ended with a general statement about the purpose and background to LSIPs. For example, the Cheshire and Warrington LSIP summary included *"The LSIP is a three-year plan with longer-term initiatives to ensure employers are at the heart of the local skills system, and it commits to transparency and showcasing achievements"*.

The prompts used to extract and summarise the LSIP reports were adjusted throughout the pilot project based on feedback from the LSIP policy team in discussion with the researcher and analysts from the Unit for Future Skills. Final summaries of the individual areas were restricted to a maximum length of 5 paragraphs – around one side of A4. The prompts included the instruction to "keep a more balanced and neutral tone and not to repeat the purpose of the general LSIP". This improved both the quality and consistency of the individual summaries, indicating that the success of generate AI is directly related to the quality of prompts provided.

The final summaries that were produced are judged to flow well and do not repeat ideas or phrases. Although at this stage they would still require some review and revision from an expert in the subject area, they offer the potential to be a valuable time and resource saving tool that gets much of the way to a shareable summary in considerably less time.

See [Annex B](#) and [Annex C](#) for example individual summaries, the full set have not been published as part of this report.

## 2.2.3 Validation

Although the vast majority (75%) of individual LSIP summaries produced by the AI algorithm were either wholly or mostly correct, there were some incorrectly summarised LSIP reports. Either where key sectors and skills were not correctly extracted or, in a

minority of cases, where the AI-model incorrectly produced an output not based on the original data (hallucination<sup>11</sup>).

Examples of incorrect or 'hallucinated' parts included in the individual LSIP summaries include:

- **Hull and East Yorkshire LSIP** - In the summary a number of key sectors for growth are extracted - construction, manufacturing, health and social care, digital, agri-skills, and business. This fails to mention sectors in the original LSIP report including energy and the green economy, and misattributes 'business' as a key growth sector.
- **Liverpool City Region LSIP** - From the AI-produced summary "Liverpool City Region identifies key growth sectors including health and life sciences, professional business services, digital and creative technology, and port and logistics." The original LSIP report mentioned many other priority sectors including education and hospitality and tourism, "port and logistics" was covered in an annex to the report.

## 2.2.4 Response from employer representative bodies

Individual summaries were shared back with the author ERBs for review. They were asked set questions relating to the summaries and the extent to which they captured the themes of the original LSIP reports. All 38 areas responded to our short survey.

Overall, the respondents said the summaries shared were accurate in their content but missed out on some specific details or nuances contained in the original reports. When asked whether they or their team would be likely to use the summaries in their work, the majority said unlikely or unsure. The main reason given was that they had produced their own summaries and resources for the purpose of sharing with stakeholders and for communications.

To produce a library of one-page, accurate and consistent summaries to provide a quick and easy overall impression of local needs and provide reference for policy queries, there will need to be further manual work to redraft and bring in ERBs' comments.

## 2.3 Topic identification analysis

### 2.3.1 Predefined topic analysis

The database generated for the summarisation process could also be used to produce a measure of how closely each LSIP focused on pre-determined topics, relative to other LSIPs; effectively creating a structured database on what LSIPs have focused on from

---

<sup>11</sup> AI hallucinations are a phenomenon where an artificial intelligence (AI) model produces a confident response that is not based on real data or events.

their unstructured text. The topics included key sectors, transferable skills and cross-cutting themes. A semantic similarity score was then calculated showing the alignment between the 'chunk' and the topic of interest.

For a given LSIP and a given topic, the majority of chunks were not relevant to the topic but there were a few relevant chunks with associated high semantic similarity scores (HSSS). These were averaged to show the extent to which an LSIP discussed a particular topic, referred to as the 'strength of discussion'.

OpenAI's GPT-3.5 Model was then used to verify whether the semantic similarity scores were a reliable measure of the discussion related to the given topics. It was found that the semantic similarity scores were not particularly precise, for example, a chunk that discussed "digital skills" in general terms would have a high semantic similarity to "digital skills" and "digital skills needed". However, the chunk may not state that digital skills are actually needed.

### 2.3.2 Self-supervised analysis

To address the lack of accuracy in the predefined topic method, a self-supervised method was developed to detect and extract novel themes across LSIPs. This relied on the database created during summarisation, a sentence transformer embedding model, and OpenAI's GPT-4 Model.

The sentence transformer model takes each 'chunk' in turn from the database and maps it into a high-dimensional space whilst preserving the semantic and syntactic 'distance' between items. This high-dimensional mapping process produces a model that clusters semantically similar items nearer to each other than items that are semantically different.

Using a uniform manifold approximation and projection (UMAP) technique, this high-dimensional map was reduced down to 50 dimensions and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBScan) techniques were applied to identify small clusters within this space. OpenAI's GPT-4 Model was used to generate suitable names for each cluster based on the evidence points contained in the cluster compared to others.

**Table 1: Top 10 clusters identified and named using self-supervised AI techniques**

Cluster name
Employer-Engaged Training and Education Strategies
Digital Skills Development and Training in the Workforce
Net Zero and Green Skills Development in the Workforce
Apprenticeship Programmes and Policies
Workforce Diversity and Inclusion Strategies

UK Local Skills Development Strategies and Initiatives
Workforce Upskilling and Reskilling Needs and Strategies
Workplace and Transferable Skills Development
Employer and Education Provider Collaboration for Skills Development
Development and Delivery of Short, Modular Courses for Professional Upskilling and Training

### 3 Sectors, transferable skills and cross-cutting themes important to LSIP areas

Using AI techniques, information from the 38 unstructured LSIP reports has been compiled into a structured dataset. This can be queried to identify, amongst other things, the top sectors, transferable skills and cross-cutting themes mentioned in specific LSIPs. These are ordered according to the ‘strength of discussion’ score which shows the extent to which an LSIP discussed a particular topic, see [section 2.3.1](#) for more information.

It should be noted that not all areas took a sector-based approach in their LSIP and there is no common definition of sectors used across areas.

The top sectors identified across all reports were manufacturing and engineering, health and social care, digital, and creative industries - the top 10 sectors mentioned are listed in Table 2. These were identified through a combination of AI and manual review. Although most of the sectors identified by LSIP areas align with those covered by national skills priorities<sup>12</sup>, the local nature of the LSIPs means that some sectors are only be mentioned by a subset of areas. In general, these are more area-specific sectors for example tourism and hospitality, financial and professional services, and agriculture and land-based sectors.

**Table 2: Top 10 sectors covered by LSIPs**

Sectors
Manufacturing and Engineering
Health and Social Care
Digital (IT and Communications)
Creative Industries
Education
Construction (and built environment)
Finance and Professional Services
Tourism and Hospitality
Transport and Logistics
Agriculture (Agri-tech and Land-based)

As part of the validation process the information contained in the structured dataset is compared to two manual reviews of the LSIPs, one by policy officials and another from

---

<sup>12</sup> Including, but not limited to, the key growth sectors of Digital Technology, Green Industries, Life Sciences, Advanced Manufacturing and Creative Industries.



analysts. Both these sources focus on priority or key sectors so that is the focus of this validation exercise.

**Table 3: Top 10 of sectors extracted across LSIP reports (not ordered)**

<b>LLM analysis</b>	<b>Manual review 1</b>	<b>Manual review 2</b>
Manufacturing and Engineering	Manufacturing and Engineering	Manufacturing and Engineering
Health and social Care	Health and Social Care	Health and Social Care
IT and Communications	IT and Communications	Digital <sup>13</sup>
Creative Industries	Creative Industries	Creative Industries
Construction	Construction	Construction
Finance and Professional Services	Professional Services	Business and Professional Services
Tourism and Hospitality	Tourism and Hospitality	Tourism and Hospitality
Education	Education	
	Transport and logistics	Haulage and Logistics
	Agriculture	Agri-Tech and Land-based
Public Sector		
Retail & Leisure		
		Energy and Green Economy

### 3.1 Sector and thematic view

The following sections provide examples of possible outputs that can be generated from the database. The plans can highlight core priorities, but do not provide a comprehensive review of all skills.

The LSIPs do not provide specific or quantitative output plans of the numbers of jobs or skills expected in an area.

---

<sup>13</sup> Digital was often classified as a cross cutting need and not classified as a sector.

### 3.1.1 Manufacturing and Engineering

More than 70% (27 of 38 areas) of LSIPs highlighted manufacturing and engineering and related sectors as a priority in their reports. The strength of discussion relating to manufacturing and engineering had a relatively large variance across areas that highlighted it as a priority.

Topics covered by LSIPs in their discussion of the manufacturing and engineering sector include, for example:

- Addressing misperceptions of the sector and promote the opportunities and benefits within the industry more clearly,
- Attracting new entrants to the sector alongside upskilling the existing workforce, particularly considering the challenge of an aging workforce in this sector, and
- Modernising the curriculum, particularly in the context of emerging roles, to provide training that emphasises areas like automation, digital skills, engineering skills, and intermediate-level provisions.

**Table 4: Top 5 LSIP areas that discuss manufacturing and engineering, average and range across all areas highlighting as priority**

Area	Strength
Kent and Medway	0.60
The Marches	0.60
Oxfordshire	0.55
Cornwall and the Isles of Scilly	0.54
North East	0.54
Average: 0.49 Range: 0.37-0.60	



### 3.1.2 Construction

More than 70% (27 of 38 areas) of LSIPs highlighted construction and related sectors as a priority in their reports.

Discussion relating to the construction sector covered a wide range of topics, for example:

- Tailoring training to meet needs of particular areas of growth, such as off-site construction, sustainable construction and site supervisors,
- Specific occupations or skills that are in high demand, including carpentry and joinery skills, plumbing and retrofitting, and understanding how these might continue to change in the future, and
- Growing demand for skills related to environmental technology, sustainability and achieving net zero emissions.

**Table 5: Top 5 LSIP areas that discuss construction, average and range across all areas highlighting as priority**

Area	Strength
The Marches	0.53
Kent and Medway	0.51
Hull and East Yorkshire	0.49
Greater Lincolnshire	0.48
Oxfordshire	0.48
Average: 0.45	
Range: 0.31-0.53	



### 3.1.3 Tourism and Hospitality

Around 20% (8 out of 38 areas) of LSIPs highlighted the tourism and hospitality sector as a priority in their reports. LSIPs that highlighted the tourism and hospitality sector as a local skills priority were in London and the South East, the North West, and Yorkshire and the Humber.

Tourism and hospitality industries may be important to other local areas but are not highlighted as priority sectors by the LSIPs for a number of reasons.

Topics covered by LSIPs in their discussion of the tourism and hospitality sector include:

- Highlighting the potential transferable skills across various industries and the importance of digital skill enhancement,
- Attracting new entrants to the sector by changing perception of the industry to one that offers meaningful career pathways and progression, and
- Tailoring training to offer short modular courses covering in demand skills including catering and security skills alongside core attributes of customer service and communication.

**Table 6: All LSIP areas that discuss tourism and hospitality, average and range across all areas highlighting as priority**

Area	Strength
Cumbria	0.53
North of Tyne	0.50
Oxfordshire	0.48
Hertfordshire	0.48
Lancashire	0.48
Brighton and Hove, East Sussex, West Sussex	0.46
Greater London	0.43
Enterprise M3 LEP including Surrey	0.43
York and North Yorkshire	0.38
Average: 0.46	
Range: 0.38 – 0.53	



## 3.2 Growth priorities spanning multiple sectors

There is a strong correlation between issues identified by LSIPs and those well known by DfE and across government. As a foundation for further research and collaboration, the current analysis allows quick comparison on the focus of specific LSIP areas on these issues. This localised information can be used to cross-check local interventions and identify further local needs.

**Table 7: Top 5 cross-cutting issues covered in LSIPs**

Cross-cutting themes	Average strength across areas
Careers Education, Information, Advice and Guidance (CEIAG), Careers and Enterprise Company, National Careers Service	0.63
Skills Planning and Flexibilities	0.58
Apprenticeships	0.57
Careers Information advice and guidance	0.56
Green Skills and Net Zero	0.52

### 3.2.1 Green Skills and Net Zero

Green skills and the needs created by net zero targets are the only category of skills that ERBs were required (by legislation and statutory guidance) to consider while developing their LSIPs.

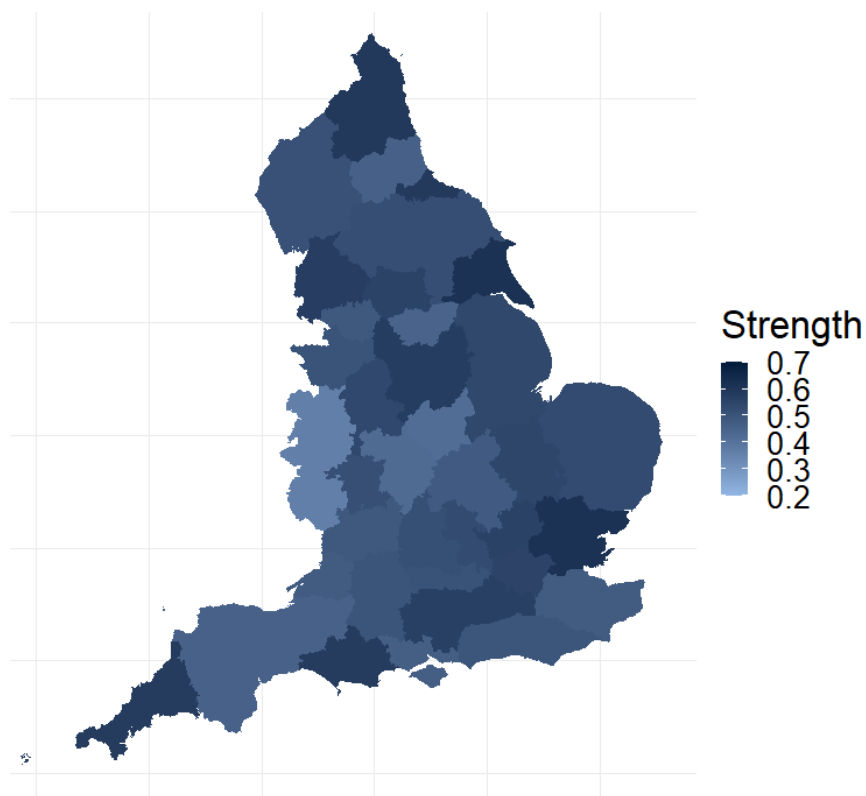
The lens through which LSIPs examine green skills varies. Some ERBs have approached it as a sector and examined the skills needs of businesses in that industry, while others have looked at the green skills needs of employers across a wide range of sectors.

**Table 8: Top 5 LSIP areas that discuss green skills and net zero, average and range across all areas**

Area	Strength
Hull and East Yorkshire	0.62
Essex, Southend-on-Sea and Thurrock	0.62
North of Tyne	0.59
Tees Valley	0.59
Cornwall and the Isles of Scilly	0.58
Average: 0.52 Range: 0.37-0.62	



Figure 7: Strength of discussion on green skills and net zero across LSIP areas<sup>17</sup>



The wide-ranging discussion regarding green skills and net zero across LSIPs covers, for example:

- Developing sector-specific short courses on green technology, clean energy, and low-carbon transition skills, alongside the integration of green skills into all study programmes,
- Addressing language and terminology barriers related to net zero requirements and broader understanding of decarbonization policies, and
- Cross-sector relevance and transferable nature of green skills, particularly in net zero supply chains including energy storage, green finance, and agri-tech.

---

<sup>17</sup> Liverpool City Region LSIP is excluded from this analysis due to extraction issues, much of the content of the LSIP was contained in tables and annex which were not included in the AI extraction.

Figure 8: Word cloud of the 100 most common words across the reports relating to green skills and net zero



### 3.2.2 AI and Automation

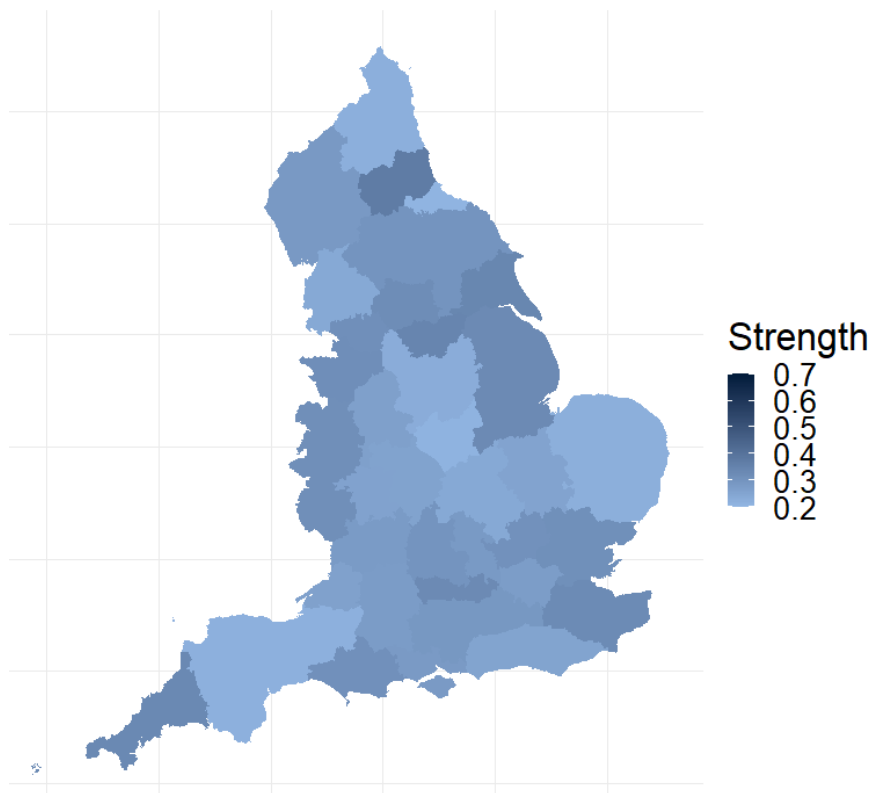
Unlike green skills and net zero, consideration of AI and automation in relation to local skills needs was not a requirement of ERBs while developing their LSIPs. Consequently the average strength of discussion relating to AI and automation was notably lower than for other topics including green skills and net zero.

Despite this, almost all areas covered AI and automation to some degree, recognising that it is an area they will continue to engage with employers on as they keep the LSIP under review for the next round of LSIP development in 2025-26. The lens through which LSIPs examine AI and automation varies. Similar to green skills and net zero, some ERBs have approached it as a sector and examined the skills needs of businesses in that industry, while others have looked at the AI and automation needs of employers across a wide range of sectors.

Table 9: Top 5 LSIP areas that discuss AI and automation, average and range across all areas

Area	Strength
North East	0.38
South Yorkshire	0.35
Hull and East Yorkshire	0.34
Cornwall and the Isles of Scilly	0.34
Greater Lincolnshire	0.33
Average: 0.29	
Range: 0.21-0.38	

Figure 9: Strength of discussion on AI and automation across LSIP areas<sup>18</sup>



Topics covered by LSIPs in their discussion of AI and automation include, for example:

- Addressing evolving skills needs, particularly data and digital skills, in the context of automation and digitalisation in various sectors,
- Considering the impact of technological change and automation on the labour force, specifically in relation to helping employers adapt to technology and integrating these tools into existing practices, and
- Consistent delivery of digital training across programmes in both basic and advanced digital skills, including those related to automation, robotics, virtual reality, software programming and data analytics.

---

<sup>18</sup> Liverpool City Region LSIP is excluded from this analysis due to extraction issues, much of the content of the LSIP was contained in tables and annex which were not included in the AI extraction.

Figure 10: Word cloud of the 100 most common words across the reports relating to AI and automation



### 3.3 Transferable skills

Transferable skills are skills, knowledge or abilities which can be applied to a wide range of different jobs and industries. They can help build a resilient workforce as technological and social changes manifest in different combinations of skills within an occupation, or in new occupations for example in the case of green technologies.

Transferable skills, such as digital literacy and leadership skills, are therefore understandably mentioned across many areas.

Table 10: Top 5 transferable skills covered in LSIPs

Transferable skills	Average strength across areas
ICT and digital literacy skills	0.53
Leadership skills	0.47
Adaptability skills	0.45
Literacy skills	0.45
Collaboration and teamwork skills	0.43

#### 3.3.1 ICT and digital literacy skills

It is widely anticipated that ICT and digital skills will become increasingly important to many jobs across sectors in the future. This extends beyond roles that can be defined as

'digital' to those that do not currently require ICT or digital skills. In addition to this, a small number of occupations are likely to need more advanced digital skills such as those discussed in the AI and automation section above.

Across the LSIPS, various aspects related to ICT and digital skills, digital education and their importance in the modern workforce are discussed, for example:

- Growing demand for digital skills and the challenges faced by employers in finding appropriate training,
- Considering comprehensive digital skills education and inclusivity in the workforce, alongside age, gender and socio-economic disparities in digital education participation, and
- Need for a broad spectrum of digital skills, from basic to advanced, and the importance of tailoring training to meet the needs of employers in various sectors.

**Table 11: Top 5 LSIP areas that discuss ICT and digital literacy skills, average and range across all areas**

<b>Area</b>	<b>Strength</b>
South Yorkshire	0.62
Derbyshire and Nottinghamshire	0.60
North of Tyne	0.59
Essex, Southend-on-Sea and Thurrock	0.59
Tees Valley	0.59
Average: 0.53 Range: 0.43-0.62	

Figure 11: Strength of discussion on ICT and digital literacy skills across LSIP areas<sup>19</sup>

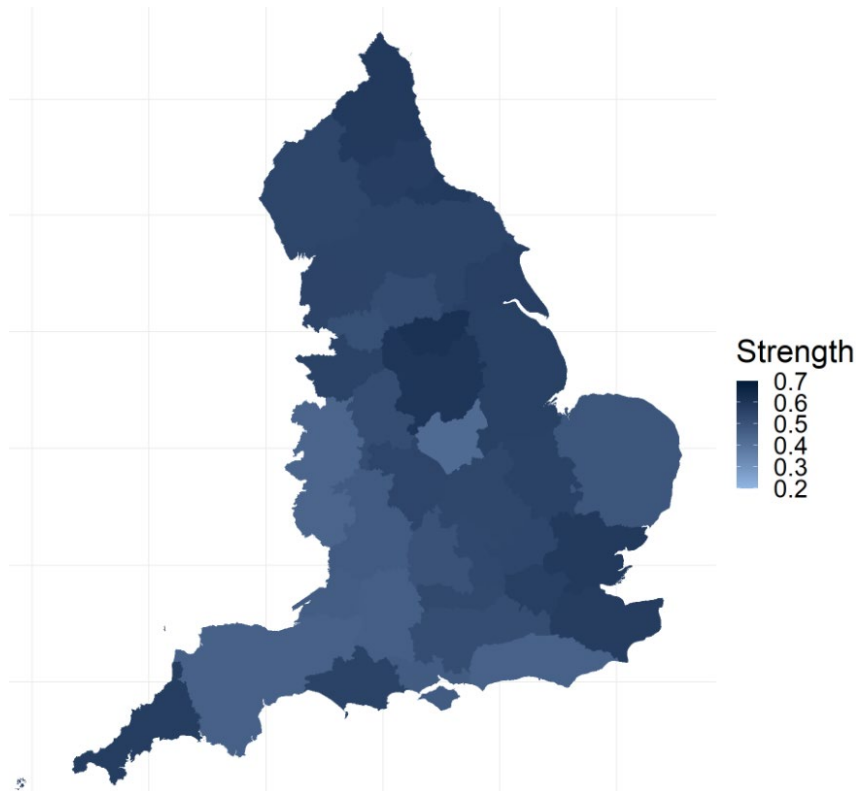


Figure 12: Word cloud of the 100 most common words across the reports relating to ICT and digital literacy skills



<sup>19</sup> Liverpool City Region LSIP is excluded from this analysis due to extraction issues, much of the content of the LSIP was contained in tables and annex which were not included in the AI extraction.

### 3.3.2 Leadership skills

Leadership skills can include skills, knowledge or abilities that aid individuals in guiding and encouraging a group of people toward achieving a common goal. Compared to previously discussed ICT and digital skills, leadership skills are often included in discussion of 'soft' skills and could include abilities such as communication, decision-making and problem-solving, as well as inspiring and empowering teams.

Across LSIP areas discussion relating to these kinds of skills are relatively high and cover issues such as:

- Creating and promoting flexible training programmes and initiatives that enhance leadership skills among current and future leaders and managers,
- Discussion of these types of skills alongside other transferable skills and so-called employability skills including confidence, creativity and knowledge of workplace expectations, and
- Their critical role, alongside digital skills, in enhancing productivity and adaptability in various industries with a particular focus on green sectors.

**Table 12: Top 5 LSIP areas that discuss leaderships skills, average and range across all areas**

Area	Strength
West Midlands and Warwickshire	0.57
North of Tyne	0.55
Cumbria	0.54
Worcestershire	0.54
Essex, Southend-on-Sea and Thurrock	0.53
Average: 0.47	
Range: 0.41-0.57	





## 4 Further validation work

Validation of AI-extraction and generation is a challenge for this emerging field as a whole, aside from manual comparison with original texts and other summaries there are a number of different types of validation exercise that are recommended. As the field matures it is likely there will be a shift more towards automated testing rather than human testing over time.

Projects that summarise large amounts of information across multiple sources are quite hard to assure even when done by humans. One of the benefits of the AI is that it is easy to recreate any of the workings, so it is viable to conduct spot checks. However, the concern with generative AI is that it does not do the same thing every time, so spot checks are less of a guarantee compared to when using more traditional methods of AI.

This section discusses approaches that could be considered to apply further validation to AI generated outputs.

### 4.1 Extensions and automated validation checks

#### 4.1.1 To identify anything missing

Option 1: Repeat the AI-summarisation steps using different parameters (specifically trying a higher temperature parameter for more creative responses). Then add some human intervention to inspect the newly generated themes and see if anything is picked up that seem to have been missed the first time of asking. This would need to be treated with care, as the LLM is more likely to hallucinate on a higher temperature setting, but may also pick out more themes that were harder to spot.

Option 2: Use a more traditional information retrieval method such as term frequency-inverse document frequency (TF-IDF) and ensure that the “high importance” words in each document have been picked up by the summaries.

Option 3: Remove everything already extracted and see what’s left, by:

- Creating vector embeddings of all the documents
- Filter out all of the topics that are already identified
- Pass what is left through some other summarisation
- See if any other key points are picked out.

#### 4.1.2 To ensure that no key ideas have been hallucinated

Option 1: Reverse engineer the clusters. If there is a result which says there are 253 entries relating to “Skills and Training for Net Zero and Green Economy Transition” then it should be possible to search for that topic and find around 253 entries.

Option 2: Human intervention could be added here to spot check the results that were found, or an automated test could be set to fail if there are an insufficient number of topics found.

### **4.1.3 Semantic scores**

Option 1: Rigorous human testing to spot check whether the key themes extracted from individual LSIPs match those highlighted by a human reviewer. This has already been undertaken to an extent by the manual review discussed above.

Option 2: Test region by region and use an AI-approach to rank the order of importance of the following skills: {skills} in the following document {document}. See if the output aligns with the structured dataset produced by the LLM approach.

## Annex A – LSIP areas

Table 13: List of LSIP areas

LSIP specified area	Designated ERB	Locations covered
Brighton and Hove, East Sussex, West Sussex	Sussex Chamber of Commerce	Brighton and Hove, East Sussex, West Sussex
Buckinghamshire	Buckinghamshire Business First	Buckinghamshire
Cambridgeshire and Peterborough	Cambridgeshire Chambers of Commerce	Cambridgeshire and Peterborough Combined Authority
Cheshire and Warrington	South Cheshire Chamber of Commerce and Industry	Cheshire East, Cheshire West and Chester, Warrington
Cornwall and the Isles of Scilly	Federation of Small Businesses	Cornwall, Isles of Scilly
Cumbria	Cumbria Chamber of Commerce	Cumbria
Derbyshire and Nottinghamshire	Federation of Small Businesses	Derby, Derbyshire, Nottingham, Nottinghamshire
Dorset	Dorset Chamber of Commerce and Industry	Bournemouth, Christchurch and Poole, Dorset
Enterprise M3 LEP (including all of Surrey)	Surrey Chambers of Commerce	Hampshire (excluding the districts of Eastleigh, Fareham, Gosport, Havant, New Forest), Surrey
Essex, Southend-on-Sea and Thurrock	Essex Chambers of Commerce	Essex, Southend-on-Sea, Thurrock
Gloucestershire	Business West Chambers of Commerce	Gloucestershire
Greater Lincolnshire	Federation of Small Businesses	Lincolnshire, North Lincolnshire, North East Lincolnshire, Rutland

<b>LSIP specified area</b>	<b>Designated ERB</b>	<b>Locations covered</b>
Greater London	BusinessLDN (previously London First)	Greater London
Greater Manchester	Greater Manchester Chamber of Commerce	Greater Manchester Combined Authority
Heart of the South-West	Devon & Somerset Local Skills Improvement Plan	Devon, Plymouth, Somerset, Torbay
Hertfordshire	Hertfordshire Chamber of Commerce	Hertfordshire
Hull and East Yorkshire	Hull & Humber Chamber of Commerce	East Riding of Yorkshire, Kingston upon Hull
Kent and Medway	Kent Invicta Chamber of Commerce	Kent, Medway
Lancashire	North & Western Lancashire Chamber of Commerce	Blackburn and Darwen, Blackpool, Lancashire
Leicester and Leicestershire	East Midlands Chamber of Commerce	Leicester and Leicestershire
Liverpool City Region	Liverpool City Region Chambers of Commerce	Liverpool City Region
Norfolk and Suffolk	Norfolk Chambers of Commerce	Norfolk, Suffolk
North East	North East Automotive Alliance (NEAA) Limited	Durham, Gateshead, South Tyneside, Sunderland
North of Tyne	North East England Chamber of Commerce	Newcastle upon Tyne, Northumberland, North Tyneside
Oxfordshire	Thames Valley Chamber of Commerce Group	Oxfordshire
Solent	Hampshire Chamber of Commerce	Isle of Wight, Portsmouth, and the following districts in Hampshire: Eastleigh, Fareham, Gosport, Havant, New Forest, Southampton

<b>LSIP specified area</b>	<b>Designated ERB</b>	<b>Locations covered</b>
South Yorkshire	Doncaster Chamber of Commerce	South Yorkshire Combined Authority
South-East Midlands	Northamptonshire Chamber of Commerce (incorporating Milton Keynes Chamber)	Bedford, Central Bedfordshire, Luton, Milton Keynes, North Northamptonshire, West Northamptonshire
Stoke-on-Trent and Staffordshire	Staffordshire Chamber of Commerce & Industry	Staffordshire, Stoke-on-Trent
Swindon and Wiltshire	Business West Chambers of Commerce	Swindon, Wiltshire
Tees Valley	North East England Chamber of Commerce	Darlington, Hartlepool, Middlesbrough, Redcar and Cleveland, Stockton on Tees
Thames Valley Berkshire	Thames Valley Chamber of Commerce Group	Bracknell Forest, Reading, Slough, West Berkshire, Windsor and Maidenhead, Wokingham
The Marches	Shropshire Chamber of Commerce	Herefordshire, Shropshire and Telford and Wrekin
West Midlands and Warwickshire	Coventry & Warwickshire Chamber of Commerce	Warwickshire, West Midlands Combined Authority
West of England and North Somerset	Business West Chambers of Commerce	West of England Combined Authority, North Somerset
West Yorkshire	West & North Yorkshire Chamber of Commerce	West Yorkshire Combined Authority
Worcestershire	Herefordshire & Worcestershire Chamber of Commerce	Worcestershire
York and North Yorkshire	West & North Yorkshire Chamber of Commerce	North Yorkshire, York

<https://www.gov.uk/government/publications/designated-employer-representative-bodies/notice-of-designated-employer-representative-bodies>

## Annex B – Example of a relatively poor individual summary

This summary was judged by DfE officials and the author ERB to be a relatively poor summary of the original LSIP report. This summary was produced in the first to apply the AI algorithm to summarise LSIP reports, prompts were later update to produce a final summary of each LSIP report.

The Local Skills Improvement Plan (LSIP) for [REDACTED] identifies several cross-sector labour market needs, particularly in the construction, manufacturing, health and social care, and Agri-Tech and Food sectors. The plan recognises the impact of automation and digitalisation on these sectors and the need to embed digital skills in the workforce. It also addresses the region's Net Zero and green economy ambitions, with a significant number of employers reporting skills shortages in low carbon/net zero related skills. The LSIP proposes solutions such as identifying key occupations, understanding available provision, improving employer knowledge of technical education provision, and establishing sector-specific working groups.

In the construction and manufacturing sectors, key roles such as construction project managers, site managers, quantity surveyors, Mechanical Engineers and Production and Process Engineers require upskilling through apprenticeships or higher education. The health and social care sector is expected to grow due to an ageing population and pressures from the Covid pandemic. The plan identifies a need to change perceptions of these roles to reduce high staff turnover and increase investment in training. It also highlights the need for increased dementia awareness, mental health knowledge, care for the vulnerable and elderly, and safeguarding knowledge.

The Agri-Tech and Food sector, which accounts for a higher level of employment in the [REDACTED] region compared to the national level, could attract more young people with technology and data skills through automation. The plan also addresses the shortage of educational teaching professionals and the need to address skills gaps. The LSIP emphasises the need for improved careers guidance in [REDACTED], reflecting local employment opportunities and delivered by professionals with knowledge of the local landscape. It proposes a review of current careers guidance activity and the development of a plan for conducting careers guidance.

The LSIP presents a comprehensive strategy to align post-16 technical education with local labour market needs. The plan emphasises the need for a system where key partners can provide support more freely, ensuring the availability of entry-level pathways and promoting pre-employment support for those from disadvantaged backgrounds. The aim is to improve soft skills and employability skills, which are crucial for local economic growth and improving employability and progression for learners. The LSIP also underscores the importance of consulting with local colleges and Independent Training Providers (ITPs) to understand traineeship provision post-August 2023.

Lastly, the LSIP for [REDACTED] recognises the importance of all stakeholders and calls for working together to achieve ambitious priorities. The plan thanks stakeholders for their engagement and looks forward to continuing positive relationships for the benefit of [REDACTED]. It also emphasises the need to maintain productive collaboration with other LSIPs and address common priorities, which is seen as crucial for the successful delivery of the LSIP.

## Annex C – Example of a relatively accurate individual summary

This summary was judged by DfE officials and the author ERB to be a relatively good summary of the original LSIP report. This summary was produced in the final attempt to apply the AI algorithm to summarise LSIP reports, after prompts were updated following feedback.

The Local Skills Improvement Plan (LSIP) for ██████████ is a strategic document that outlines the region's skills needs and priorities, with a particular focus on digital skills. The plan is based on robust evidence gathered through partnerships with Chambers of Commerce, business representative organisations, and other stakeholders. It aims to support local economic growth and boost productivity by addressing the skills requirements identified by employers. The LSIP also serves as a framework for aligning other skills plans and supports net-zero targets and climate change adaptation.

The LSIP identifies a strong demand for digital skills in the region, reflecting the changing economic landscape and the need for workers to adapt to new technologies. Specific digital skills, such as Cyber Security, are required for certain roles or sectors. The plan also highlights the importance of training and education providers in keeping up with emerging technologies and skills beyond mainstream understanding. The LSIP aims to translate employer needs into demands understood by training and education providers, and to address the increasing need for digital skills in the job market.

The LSIP outlines several actions that employers, education or training providers, and other stakeholders should take to deliver the necessary skills. These include investing in workforce development, engaging positively with the skills system, and taking responsibility for addressing digital skills gaps through workplace training. The plan also emphasises the need for close collaboration between employers and providers to ensure curriculum design reflects employer needs. Furthermore, the LSIP recommends the development of bespoke training and upskilling packages for employers, and the establishment of ██████████ as a place for world-class technical education.

The LSIP also identifies several cross-sector and region-wide needs. These include addressing skills deficiencies in social media & digital marketing skills, web content management skills, and data skills, which are critical for the region to keep pace with the evolving digital landscape. The plan also highlights the need for a holistic and joined-up offer of business support and skills support for technology adoption and upskilling employees. Furthermore, the LSIP emphasises the importance of improving the region's digital skills base and simplifying navigation of the education and training system.



Department  
for Education

© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3).

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

About this publication:

enquiries [www.education.gov.uk/contactus](https://www.education.gov.uk/contactus)

download [www.gov.uk/government/publications](https://www.gov.uk/government/publications)



Follow us on Twitter:  
[@educationgovuk](https://twitter.com/educationgovuk)



Like us on Facebook:  
[facebook.com/educationgovuk](https://facebook.com/educationgovuk)