Oxfordshire Strategic Economic Plan

Method Note and Supporting Information

October 2023

Oxfordshire Strategic Economic Plan

1. Context

glass.ai were commissioned, alongside SQW and Oxford Brookes Business School, to develop a refreshed Strategic Economic Plan (SEP) for Oxfordshire, on behalf of Oxfordshire Local Enterprise Partnership (OxLEP). The SEP will propel the sub-region's long-term prosperity and competitiveness and requires a base of research and evidence that makes sense of complex issues, considers how the LEP can achieve more inclusive growth, best engage with wider geographic constructs, and also champion the places, sectors and assets that make it distinct. In doing so, the SEP will set the scene for long-term policy, investment and partnerships.

glass.ai's inputs into the project focused on the delivery of new and deeper insights, uncovering sectors, businesses and their capabilities that will underpin a new economic narrative, and support strategic direction. Specifically, this required the preparation of bespoke datasets to establish an up-to-date picture of the Oxfordshire economy and key sectors, to challenge convention, further understanding and validate existing but disparate knowledge. This data informed the collation an evidence base, feeding into baseline and thematic deep dive reports.

The focus of glass.ai activity and key outputs was on developing:

- **A rich and up-to-date dataset of the OxLEP economy and its key sectors.**
- **Data that investigates the activities of businesses across three areas:**
    - Growth signals and expansion activity.
    - Net Zero commitment and progression towards a Just Transition.
    - Import and export activity linked to international presence and trading.
- **Accompanying data visualisation and mapping, to support impactful presentation.**

2. glass.ai Methodology

An overview of the approach taken to developing an Oxfordshire dataset is summarised below. At the heart of this is our crawling capability, which harnesses AI and machine learning comprehension and allows us to obtain unique insights, at a scale and depth, not possible before. This has augmented other research methods and data sources, to deliver a rich and place-centric view of Oxfordshire, shaping long-term strategic priorities and an ambitious growth agenda that reflects the sub-region's economic, social and environmental context.

Reading the Open Web

glass.ai's technology makes sense of vast quantities of written language, that is textual data – whether from company websites, news, social media, government or other sources. Web data is unstructured, fast-moving and hard to query at scale. The crawling capability we have developed tracks hundreds of thousands of topics, signals and other indicators of interest across billions of web pages, watching over more than 40 million organisations globally. To obtain rich textual information on companies, sectors, people and their interactions, the glass.ai crawler collates and analyses data only from the open web – **key sources are illustrated below**.

The Open Web
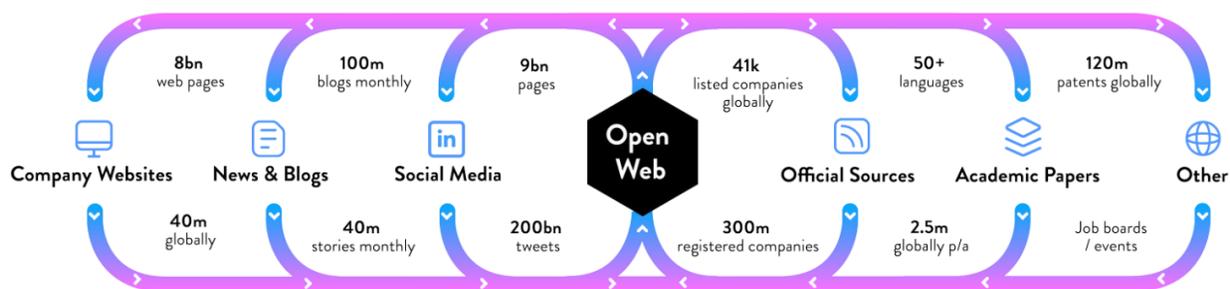
Crawling Characteristics

glass.ai has an ongoing discovery process that reads websites and classifies the site as a company website if it detects certain criteria around content and will predict the sector and geography of the business. Currently glass.ai reads over 2 million UK based company websites. **This represents a significant proportion of UK businesses with a web presence, and those making an important economic contribution**. As well as discovering new content, this ongoing process keeps the existing view of the businesses that it knows about, up to date.

To make this possible, glass.ai has invented AI technology that applies techniques at the frontiers of machine language understanding to deliver innovative research and go deeper in the characteristics of companies, sectors and clusters. These steps are, broadly, described below:

Key Crawling Steps

Semantic analysis

glass.ai detects entities and classifies content from text (e.g. companies, people, products, news) with state-of-the-art precision. So when our AI goes to a website or reads a web source, it makes its own decisions on what is being talked about.

Resource crawling

glass.ai is an intelligent crawler, with smart filtering and crawling that follows links that are likely to discover the data that is most relevant to the results, simulating how a human would efficiently scan a website. We extract large-scale datasets efficiently.

Topics ontology

glass.ai builds language models and large topic maps to help understand the web content and open the data for further investigation. Our current ontology contains around 300k related topics and themes, which are continuously updated.

Entity onboarding

Everything in glass.ai is fully automated. For example, when detecting businesses, glass.ai automatically recognises the type of site, the name of the business, its sector, and then it deeps reads the site to understand the activities of the firm.

Limitations

The open web provides a rich context around the activities of businesses. **However, not all UK businesses have a website, so our approach is limited to those that have an independent web presence**. Whilst the crawler's capabilities are constantly being developed and optimised, reasons for some websites being unreadable include inactive sites/domains, the nature and detail of information disclosure and format.

When performing targeted research, a challenge when using specific keywords to identify activities is that if a relevant business doesn't list those words on their website. This can be mitigated by supplying a broad range of topics associated with the types of companies that need to be discovered and using semantic analysis to discover topics and content related to the supplied list. **This approach has been followed here, through the development of curated sector taxonomies and drawing on local research inputs.**

Quality Assurance

To quality-assure the data, glass.ai runs both manual and automatic quality assurance processes across the data fields. The glass.ai team checks samples of the data to identify errors and isolate classes of issues. In addition, automated aggregate and statistical checks are run to ensure confidence in the delivered set. All these checks and balances have been used to provide assurance that the data outputs are robust and reliable.
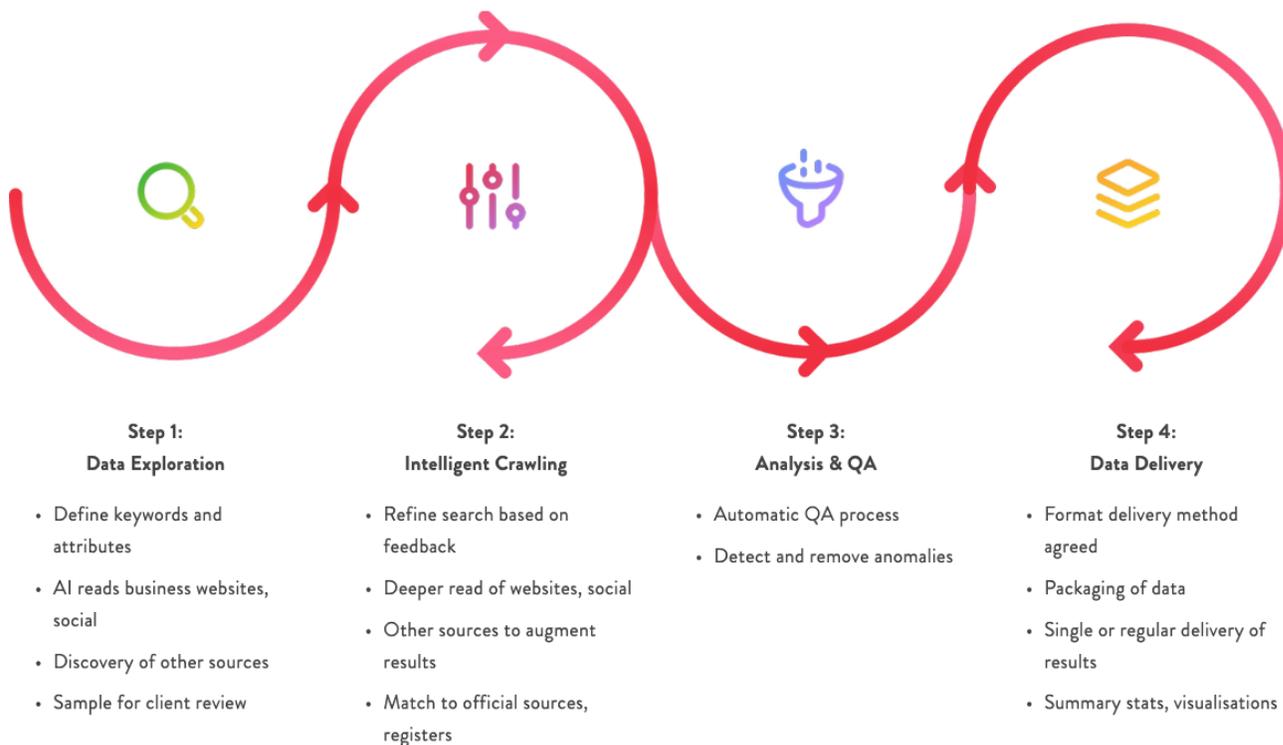
Building an Oxfordshire Dataset: Key Crawling Steps

To deliver a rich and comprehensive Oxfordshire dataset, we developed and deployed a staged approach to delivery:

- **Stage 1: Economic Activity and Sector Crawling** – applying AI technology and machine learning to crawl web sources and build a comprehensive database of all businesses across Oxfordshire. This also identified businesses across key sectors of interest to OxLEP; namely Advanced Engineering, AI, Connected Autonomous Vehicles, Cleantech, Creative Digital, Cryogenics, Fusion, Life Sciences, Quantum and Space. The sector level data was also contextualised as a result of benchmarking sub-regional presence versus comparator areas (England, Greater Cambridge, Oxford 2 Cambridge Arc).

- **Stage 2: Thematic Evidence Gathering** – with a stable dataset created as part of stage 1, we undertook deeper crawling to identify evidence of company growth, Net Zero progression and import/export activity, to further characterise the dynamics of activity across all sectors. This targeted signals at an organisational level across the whole economy, based on select indicators. The dataset was expanded to capture additional fields and associated evidence.

Below is the broad process we followed for the purposes of crawling and data collection:

A Staged Crawling Approach

Step 1: **Data Exploration**	**Step 2:** **Intelligent Crawling**	**Step 3:** **Analysis & QA**	**Step 4:** **Data Delivery**
• Define keywords and attributes • AI reads business websites, social • Discovery of other sources • Sample for client review	• Refine search based on feedback • Deeper read of websites, social • Other sources to augment results • Match to official sources, registers	• Automatic QA process • Detect and remove anomalies	• Format delivery method agreed • Packaging of data • Single or regular delivery of results • Summary stats, visualisations

Stage 1: Economic Activity and Sector Crawling

Key activities within this stage included:

- Established a bespoke research approach, with specific reference to the sectors of interest and how these were defined. Agreed taxonomies of keywords/phrases and other search criterion for the ten sectors, which built on existing sector descriptions, published research or example organisations.

- Discovered all organisations and companies present and active across Oxfordshire, entailed distilling information across a number of open web data points. To ensure more complete coverage (and support data validation), these were augmented with other data like news, sector and regional specific intelligence and official sources.

- Identified any text that may suggest businesses were part of the LEP's sectors of interest, driven by focused language modelling and keyword taxonomies. Sector data was also generated across comparator geographies, to contextualise the results, including benchmarking based on estimated sector size (count, employment and density).

- Validated the final full dataset through automated and manual quality assurance processes. Early-stage samples also supported QA and the generation of data focused on local understanding of sectors.

Stage 2: Thematic Evidence Gathering

Key activities within this stage included:

- Undertook deeper crawling (additional sources, more detailed open web content) to extract information relative to thematic areas of interest. This provided further evidence to support targeted deep dives, with the data being augmented with other qualitative sources.

- Gathered and distilled data, alongside supporting evidence, across the following indicators:

  - **Growth signals** – new office openings, funding and investments, commercial collaborations, new product launches and hiring signals.
  - **Net Zero progression** – Net Zero company accreditations and related job profiles.
  - **Import and Export activities** – evidence of activity domestic and international activity based on products and services.

- Final QA, including refinements made to the format and presentation of the dataset. Select visualisations and mapping was generated to support onward analysis.

Outputs

The main output was a full database with the following data fields:

ID Organisation	**glass.ai organisation identifier (for web sourced content only)**
Name	Name of organisation sourced from the open web
Website	Website of the organisation
glass.ai Sector	Predicted sector of organisation based on web description and other categorisation criteria
glass.ai Sector Group	Higher level grouping of sector based on web description and other categorisation criteria
Description	Description of organisation's main activity based on web content
Priority Sector Term Match	Terms from the priority sector taxonomies matched to crawled content, in a relevant context
Priority Sector Categorisation	Based on language modelling and matching of company activity and focus with keyword taxonomies for LEP sectors of interest
OxLEP Location	Main trading location of organisation read from the web
OxLEP Postcode	Postcode associated with main trading address
OxLEP Local Authority	Location of trading address based on OxLEP Local Authority boundary
HQ in OxLEP	Organisation has evidence that their headquarters is in Oxfordshire
LinkedIn URL	LinkedIn account of organisation, when available
Employee Range	Range of number of employees read from multiple web sources
Registered Number	Registration number of the organisation
Registered Name	Registered name of the organisation
Registered Address	Registered address of the organisation
SIC Code Text	Primary SIC code of the organisation with text description
Last Accounts Date	Latest returned organisation accounts date
Revenue GBP x1000	Latest returned organisation revenue
Year Incorporated	Year that organisation was founded
Organisation Email	General contact email for the organisation, when available

Growth Signal	Binary marker and supporting information if evidence has been discovered relative to growth signals
Net Zero Signal	Binary marker and supporting information if evidence has been discovered relative to Net Zero signals
Import/Export Signal	Binary marker and supporting information if evidence has been discovered relative to Import/Export signals
External Source	Confirmed matching to external sources, where relevant

Additional information provided to support the main dataset included:

- **Comparator results – OxLEP sectors of interest.**
- **Comparator results – glass.ai taxonomy Sector Groups**.

3. Interpretation of Results

The data generated as a result of a targeted and automated crawling process reflects an innovative but established research approach, deployed on a range of sector and placed based assignments within the UK. In doing so, it is possible to overcome the limitations of off-the-shelf company databases or using official sector taxonomies (such as SICs) to map emerging and developing sectors. The overarching emphasis is on the value of the data in terms of new discovery, pinpointing clusters of important activity and showcasing new elements of capability.

There are some considerations that should be observed when interpreting and using the data:

- **The presentation of results within the glass.ai dataset will deviate from those captured in official statistics or surveys, for good reason**. This is linked to a number of factors, including the known limitations of ONS datasets, which offer restricted disclosure of economic activity (based on business size, reporting thresholds, focusing on main site operation and an inability to differentiate between headquarters and local trading locations). The dataset seeks to push beyond these barriers and present a more representative view of economic activity.

- **The nature of material on the open web is highly dynamic and constantly changing**. While the crawl represents a real-time view of the sub-regional business base and associated activities, it is a snapshot in time, which will be superseded by change. Examples include the expansion of businesses, relocation, closure, consolidation and diversification of operations.

- **The matching of results to a local authority is based on postcodes, to support spatial analysis**. In some cases, the crawl may have collected partial addresses for the business only, so these have not been matched to a local authority, although the location is confirmed as being within the sub-region. There are some other cases where the business has been shared via a known source and are considered to be a substantiated OxLEP business.

- **Priority sector tagging is linked to full web results only, as a result of how the language models work and the need for substantive textual evidence**. As such, some sub-regional companies may be missed due to the way they describe themselves or due to a lack of a web

presence entirely. There is likely to be variation across these too, based on our experience of delivering focused sector studies.

- **Comparator data is included to add context and to support benchmarking, especially in terms of sector scale and specialisation**. Any limitations in terms of business/sector discovery in OxLEP, will apply equivalently in comparator locations.

- **Revenue data is limited to what companies legally or willingly disclose and is a proportion of all actively trading companies**. There are also challenges associated with assigning these values to activity within a given location, where a business is large, multifaceted and operating across sites, but remains a useful proxy for business scale.

- **Employment numbers are estimates of total UK jobs based on the employees in the organisations, using LinkedIn, and provide an alternate view of business headcount using real-time information**. They can be used to augment other datasets or in isolation and could be deconstructed further (i.e. employment directly attributable to localised geographies).

- **Thematic signals are based on contextualised evidence and reinforced using more than one indicator but may not provide a full picture of related activity**. It may be possible to observe other indicators to widen this view, and some signals will not have a web presence at all. That said, the data helps to showcase a triangulated position and avoids some of the limitations associated with primary research methods (such as surveys).

4. Data Results

The dataset and analysis of key data points within provide a differentiated and deeper view of economic activity and sector of interest presence across Oxfordshire. Some of the key messages emerging from the data are summarised below and will help to shape the direction of the SEP, as well as being a potential trigger for follow-on research.

Key Data Messages

- The most prominent sector groups across the OxLEP area (based on business count) are Real Estate and Construction, Professional Services, Consumer Goods and Services, Leisure and Hospitality and Technology. The Non-Profit and Education, Media and Arts, Industrial and Agriculture, Healthcare and Scientific and Supply Chain and Transport sectors account for over 10,000 businesses.

- Looking at glass.ai sector groupings (as a reflection of sub sectors), those with the greatest presence across Oxfordshire include Real Estate and Property Management, Construction, Consulting, Retail, Hospitality and Restaurants, Information Technology and Services, Hospitals and Medical Practices and Automotive.

- Despite the presence of large multinationals and anchor employers, the OxLEP economy is characterised by small businesses, with by far the greatest majority of these employing no more than 10 people. Small businesses are therefore the backbone of the sub-regional economy and an important contributor to Oxfordshire relative productivity.

- Whilst Oxford is the sub-region's primary population and enterprise centre, with a global identity and world-class assets, other local authorities are important economic contributors, as seen in the size of the business base. As evidence of this, more than 40% of all businesses in Oxfordshire are located in South Oxfordshire and Cherwell.

- Looking at OxLEP's sectors of interest, the data suggests that some have a strong existing presence, whilst, whilst emerging, are building from a relatively small base of businesses and supporting assets. The largest sectors, based on business count, are Life Sciences, Creative Digital, Cleantech and Space. In contrast, Fusion, Cryogenics, Connected Autonomous Vehicles and Quantum are the smallest sectors.

- Observing the spatial profile of sectors of interest, certain local authorities are more diversified and see a stronger presence than others. As an example, Oxford is characterised by a focus on Life Sciences, Creative Digital and Cleantech sectors, whereas the Vale of White Horse and South Oxfordshire exhibit a more even sector spread. Stronger sector presence across local authorities includes Life Sciences and Space in the Vale of White Horse, Creative Digital and Cleantech in South Oxfordshire and Creative Digital in West Oxfordshire and Cherwell.

- Regarding the deep dive signals and how these relate to the LEP's sectors of interest – evidence of growth signals and import/export activity predominate. Growth signal evidence is most significant when looking at the Life Sciences, Cleantech and Creative Digital sectors, whilst import/export activities are most commonly seen across the same sectors. Whilst Net Zero evidence, suggesting investment in a just transition to Net Zero is the smallest in scale overall, there is a clear focus on this within the Cleantech sector.

- Across all other businesses (with sectors outside those of significant interest to OxLEP), similar patterns of evidence prevail. Specifically, a greater emphasis on growth signals and import/export activity, relative to Net Zero progression, albeit there are signs of progress.

Select Data Visualisations

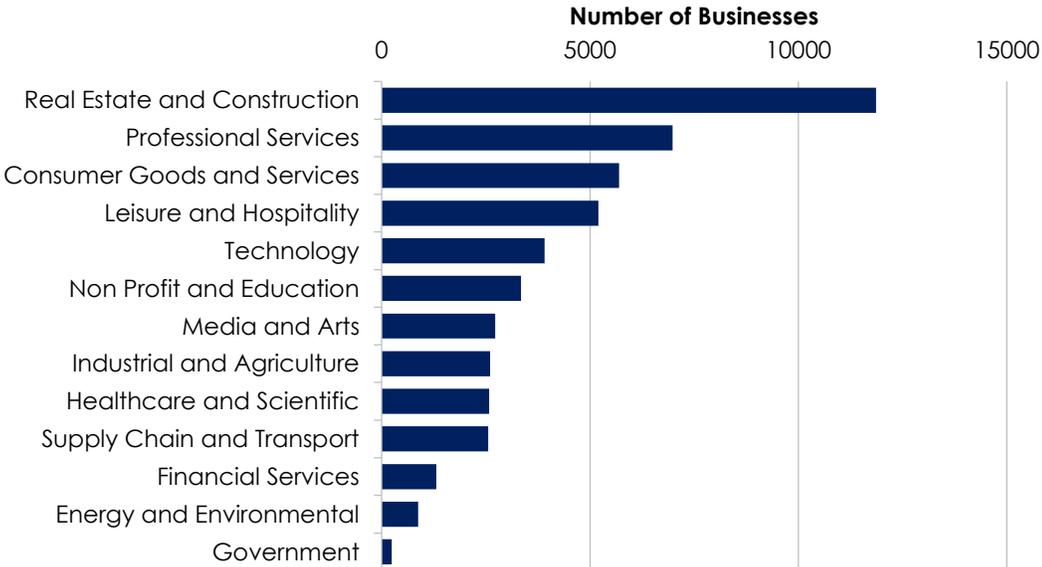**Chart 1: glass.ai Sector Group Analysis (Number of Businesses)**

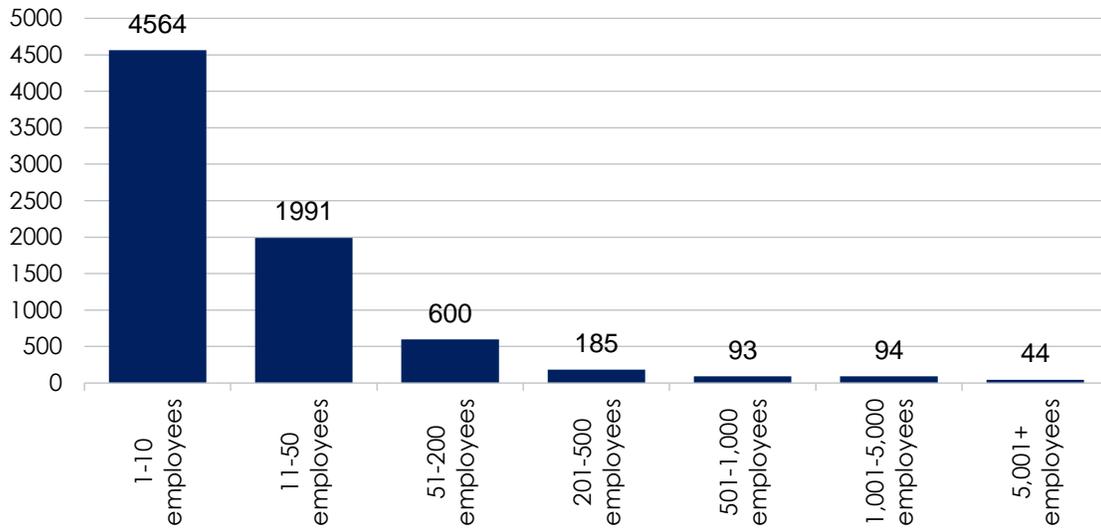**Chart 2: Size of Businesses/Organisations in OxLEP (Number of Employees)**

Category	Value
1-10 employees	4564
11-50 employees	1991
51-200 employees	600
201-500 employees	185
501-1,000 employees	93
1,001-5,000 employees	94
5,001+ employees	44

Chart 3: Number of Businesses identified (Local Authority)

| Local Authority	Value
West Oxfordshire	9684
Vale of White Horse	10060
Oxford	11414
Cherwell	12434
South Oxfordshire	12545
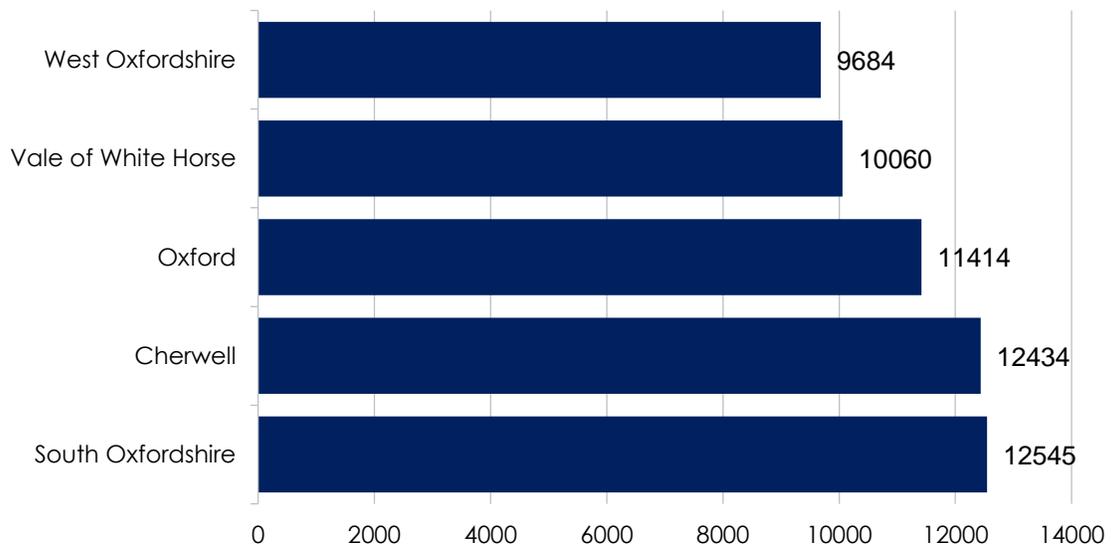
Table 1: glass.ai Sector Group Analysis for OxLEP (LAs) and Comparator Geographies

	Cherwell		Oxford City		South Oxfordshire		Vale of White Horse		West Oxfordshire	
	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ
Consumer Goods and Services	1314	1.1	1030	0.9	1235	1.0	916	0.9	978	1.0
Energy and Environmental	137	0.6	120	0.5	198	0.8	238	1.2	143	0.7
Financial Services	242	0.6	257	0.7	353	0.8	196	0.6	224	0.7
Government	40	0.5	52	0.8	35	0.5	50	0.8	33	0.6
Healthcare and Scientific	443	1.2	802	2.3	394	1.0	522	1.7	275	0.9
Industrial and Agriculture	700	1.2	212	0.4	566	0.9	509	1.1	505	1.1
Leisure and Hospitality	1095	0.7	1114	0.8	1057	0.7	783	0.7	828	0.7
Media and Arts	451	0.6	683	1.0	600	0.8	505	0.9	397	0.7
Non-Profit and Education	488	0.4	1109	1.1	544	0.5	566	0.7	368	0.4
Professional Services	1498	0.9	1227	0.8	1778	1.1	1186	0.9	1138	0.9
Real Estate and Construction	2748	1.9	1928	1.4	2597	1.8	1950	1.6	2454	2.1
Supply Chain and Transport	787	1.6	385	0.9	517	1.0	405	1.0	369	1.0
Technology	762	0.9	808	1.1	835	1.0	754	1.1	601	0.9
TOTAL	**10,705**		**9,727**		**10,709**		**8,580**		**8,313**	

Source: glass.ai data
Location Quotient (LQ) calculated relative to the national proportion of employment in each sector.
Note: Local Authority information unavailable for 1,866 businesses (out of 49,900 records)

Table 2: glass.ai Sector Group Analysis for OxLEP and Comparator Geographies

Glass.ai Sector Taxonomy	OxLEP		Greater Cambridge		O2C Arc		England
	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses
Consumer Goods and Services	1,999	1.0	716	0.7	9,149	1.0	115,063
Energy and Environmental	464	1.2	180	0.9	1,782	1.0	22,618
Financial Services	476	0.7	193	0.6	2,086	0.7	39,491
Government	182	1.5	48	0.8	568	1.1	6,834
Healthcare and Scientific	1,248	2.0	947	3.3	4,148	1.5	34,618
Industrial and Agriculture	1,031	1.0	418	0.9	5,438	1.3	55,255
Leisure and Hospitality	2,368	1.0	1,046	0.9	10,244	1.0	137,038
Media and Arts	1,116	0.9	585	1.0	4,699	0.9	68,259
Non-Profit and Education	2,381	1.3	1,334	1.6	8,653	1.1	99,936
Professional Services	2,411	0.9	1,192	1.0	10,934	0.9	148,451
Real Estate and Construction	2,253	0.9	866	0.8	10,601	1.0	135,950
Supply Chain and Transport	901	1.1	220	0.6	4,461	1.3	45,514
Technology	1,410	1.0	913	1.4	6,110	1.0	7,094
Other Sectors	1,113	0.6	454	0.5	6,035	0.8	99,448
TOTAL	**19,353**		**9,112**		**84,908**		**1,085,569**

Source: glass.ai data
Location Quotient (LQ) calculated relative to the national proportion of employment in each sector.

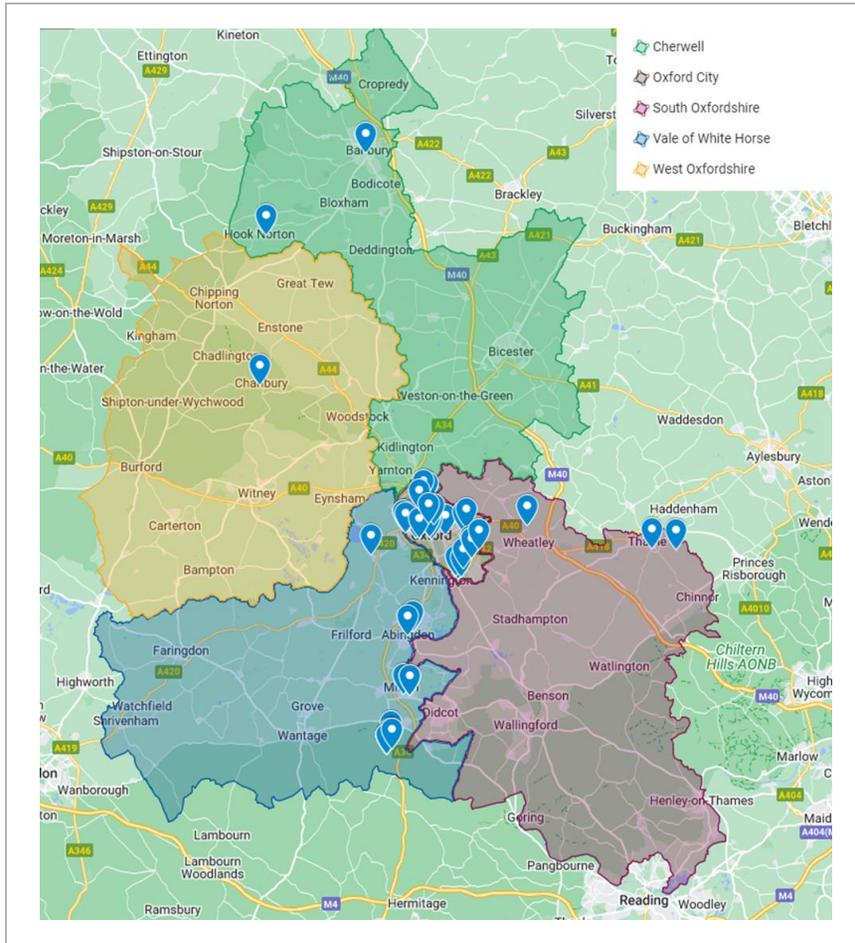**Chart 4: Businesses identified in OxLEP in 'Focus' Sectors (Number of Businesses)**
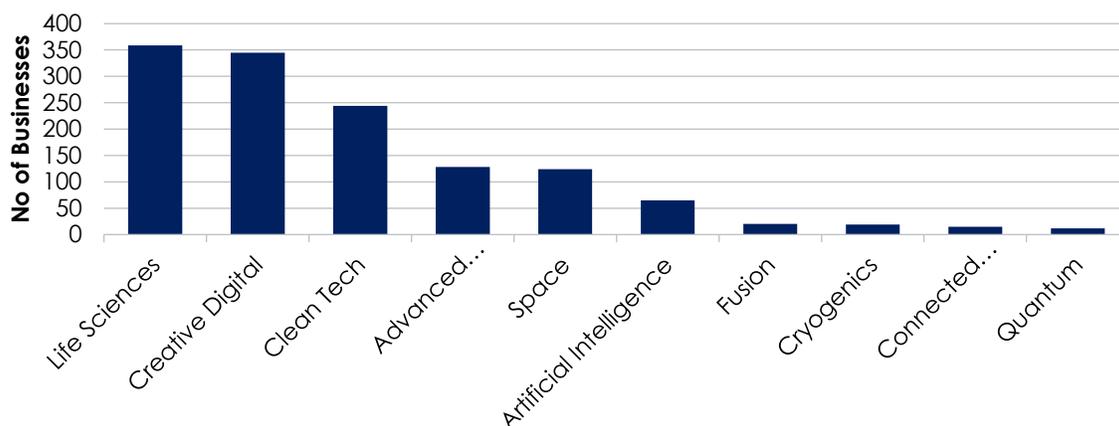
Table 3: Focus Sector Analysis for OxLEP (LAs) and Comparator Geographies

	Cherwell		Oxford City		South Oxfordshire		Vale of White Horse		West Oxfordshire	
	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ
Advanced Engineering	30	0.6	12	0.2	22	0.4	38	0.9	13	0.3
Artificial Intelligence	2	0.1	30	1.2	3	0.1	19	0.9	1	0.0
Clean Tech	34	0.4	62	0.8	47	0.5	44	0.6	28	0.4
Connected Autonomous Vehicles	1	0.2	8	2.0		0.0	5	1.4		0.0
Creative Digital	63	0.3	92	0.4	67	0.3	43	0.2	51	0.3
Cryogenics	3	1.0		0.0	3	1.0	7	2.9	3	1.3
Fusion	2	0.2	1	0.1	4	0.4	8	1.0	1	0.1
Life Sciences	36	0.5	112	1.8	51	0.8	102	1.9	34	0.7
Quantum	2	0.5	3	0.9		0.0	4	1.3	1	0.3
Space	10	0.3	7	0.2	13	0.3	58	1.8	1	0.0

Source: glass.ai data
Location Quotient (LQ) calculated relative to the national proportion of employment in each sector.
Note: Local Authority information unavailable for 1,866 businesses (out of 49,900 records)

Table 4: Focus Sector Analysis for OxLEP and Comparator Geographies

Glass.ai Sector Taxonomy	OxLEP		Greater Cambridge		O2C Arc		England
	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses	LQ	No. of Businesses
Advanced Engineering	128	1.3	88	1.8	587	1.4	5,413
Artificial Intelligence	65	1.3	76	3.1	236	1.1	2,720
Clean Tech	244	1.5	91	1.1	827	1.2	8,898
Connected Autonomous Vehicles	15	1.9	4	1.0	47	1.3	438
Creative Digital	345	0.8	180	0.8	1,635	0.8	25,013
Cryogenics	19	3.5	6	2.2	41	1.7	300
Fusion	20	1.1	12	1.4	101	1.3	972
Life Sciences	359	3.0	347	5.7	1,167	2.2	6,635
Quantum	12	1.7	6	1.7	37	1.2	378
Space	124	1.7	36	1.0	425	1.4	3,921

Source: glass.ai data
Location Quotient (LQ) calculated relative to the national proportion of employment in each sector.

Map 1: Advanced Engineering Businesses in OxLEP
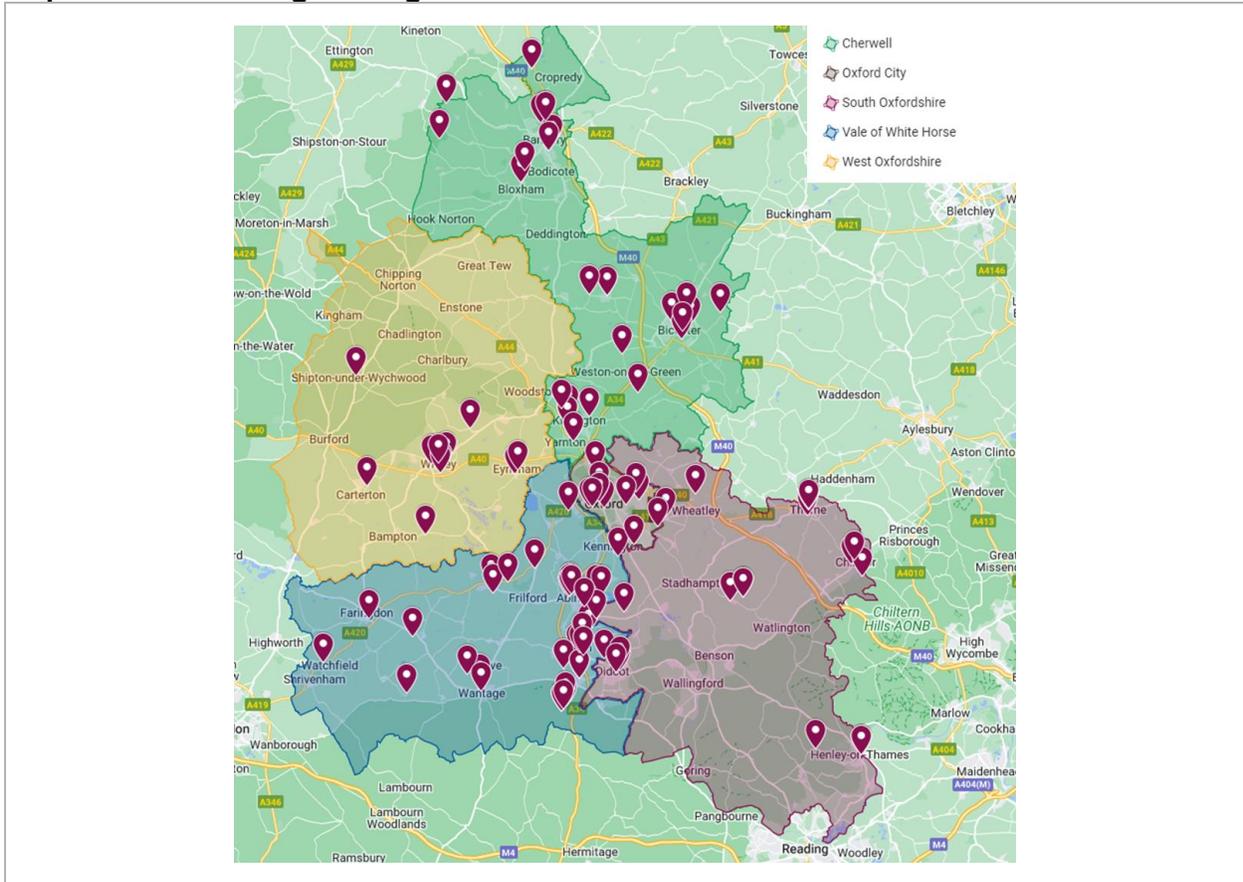
Source: glass.ai
Note: More detailed information on these maps and the specific businesses within each cluster can be found here
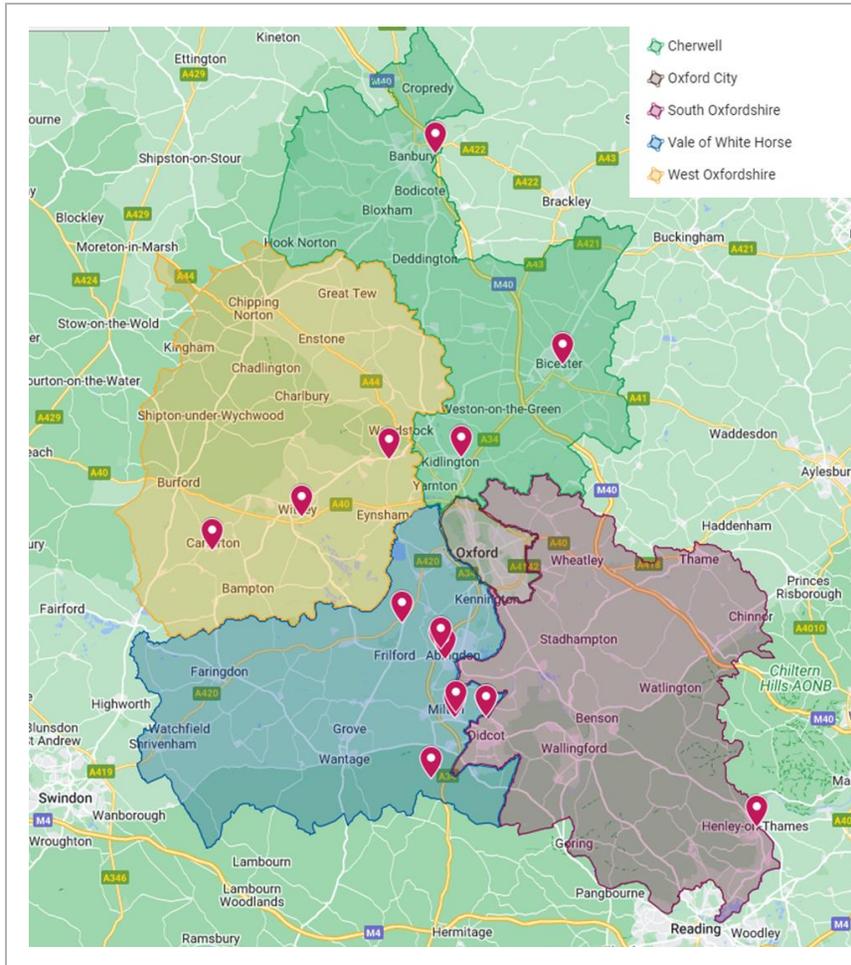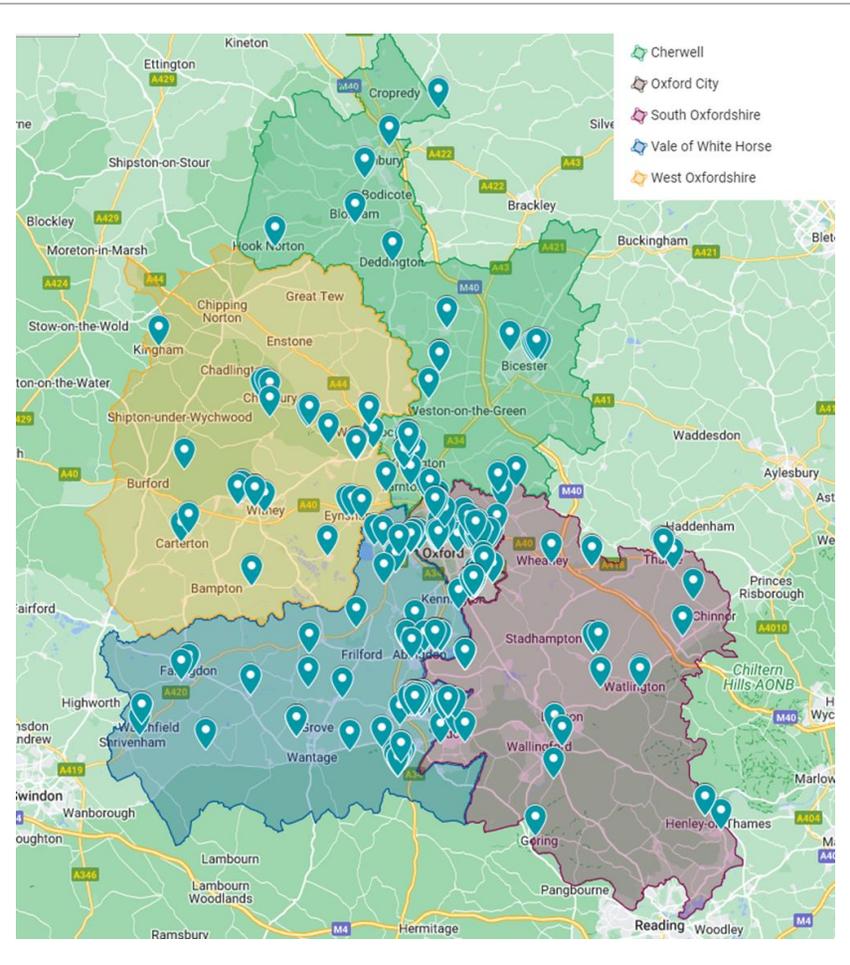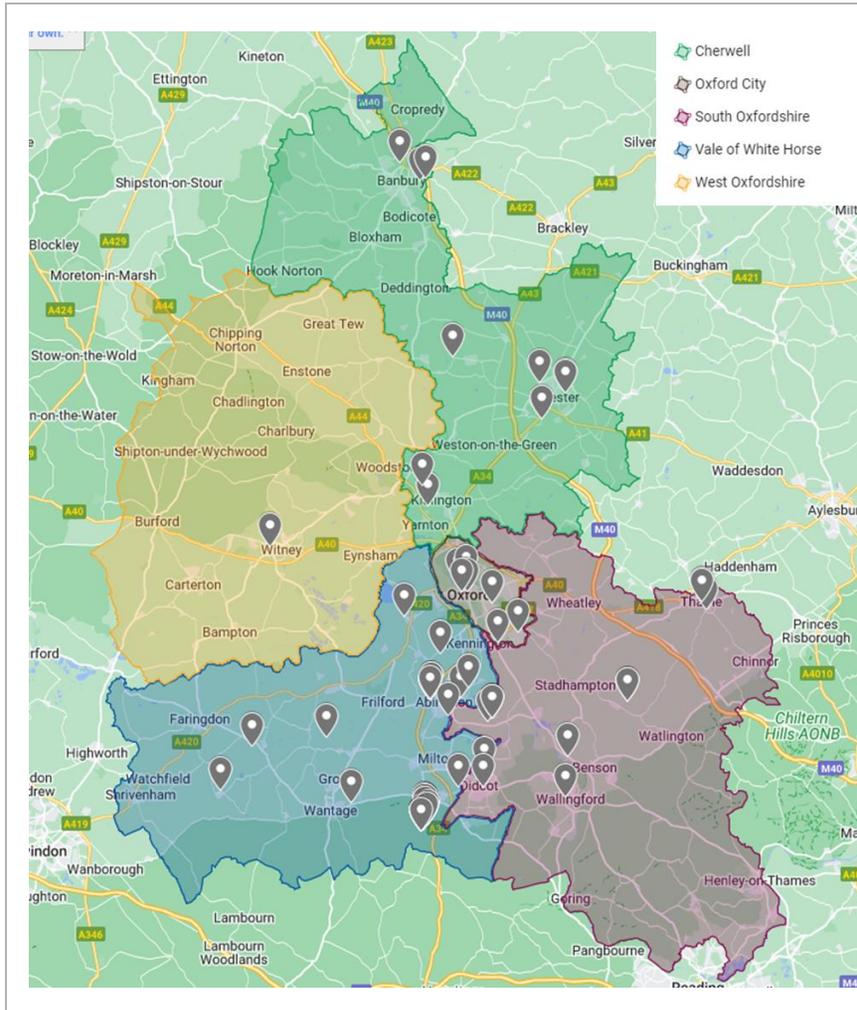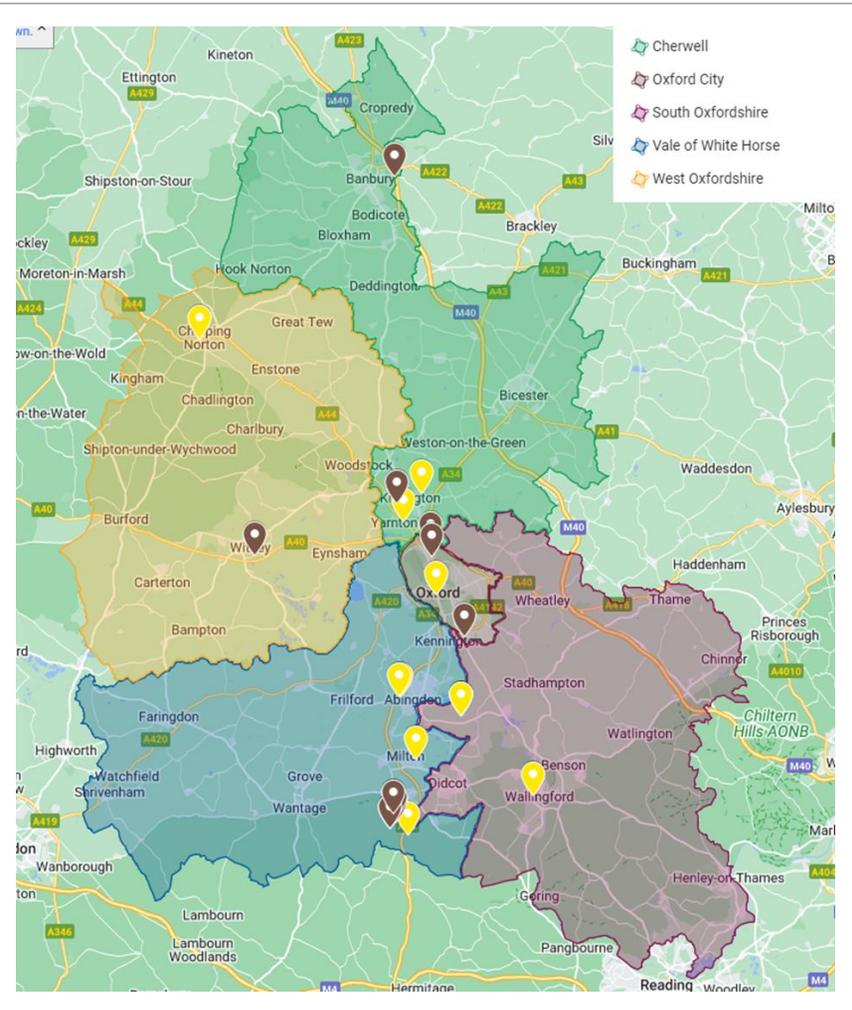
Map 2: AI Businesses in Oxfordshire

**Map 3: Connected Autonomous Vehicles Businesses in Oxfordshire**

Note: More detailed information on these maps and the specific businesses within each cluster can be found here

Map 4: CleanTech Businesses in Oxfordshire　　　　**Map 5: Creative & Digital Businesses in Oxfordshire**

Map 6: Cryogenics Businesses in Oxfordshire

Map 7: Life Sciences Businesses in Oxfordshire

Note: More detailed information on these maps and the specific businesses within each cluster can be found here

Map 8: Space Businesses in Oxfordshire

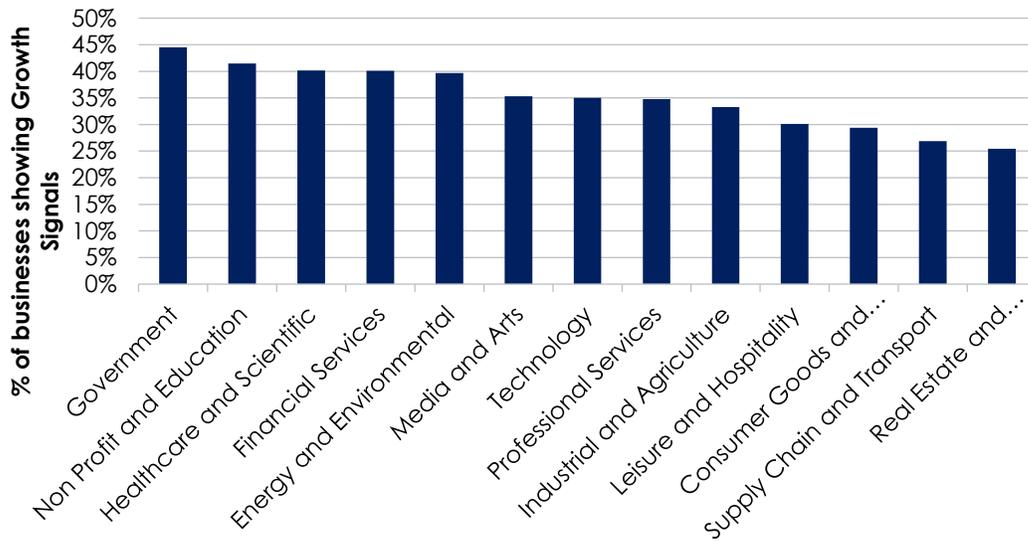**Map 9: CleanTech (Yellow) And Quantum (Grey) Businesses in Oxfordshire**

Note: More detailed information on these maps and the specific businesses within each cluster can be found here

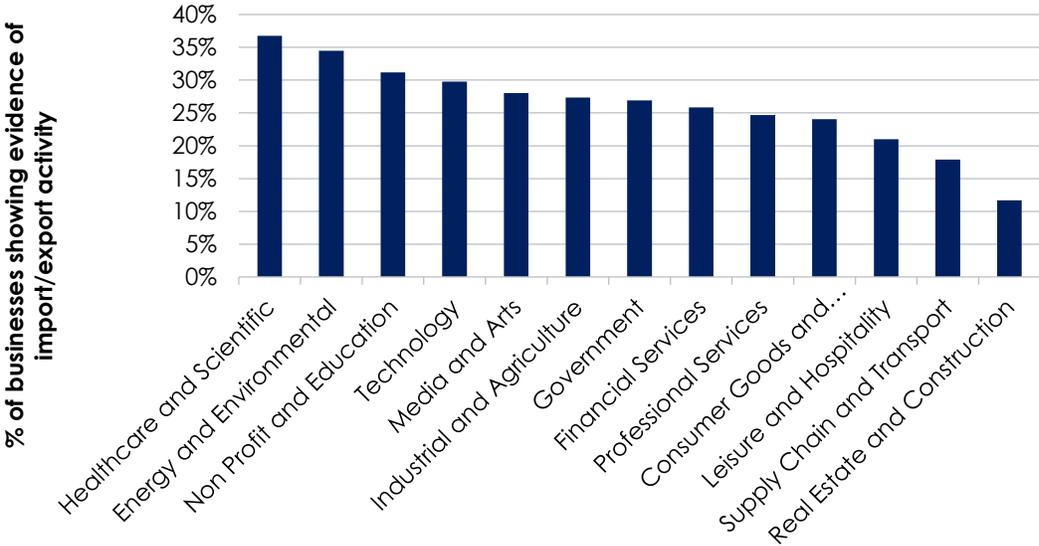**Chart 5: % of Businesses (by Sector) showing Adoption of Net-Zero Practices**
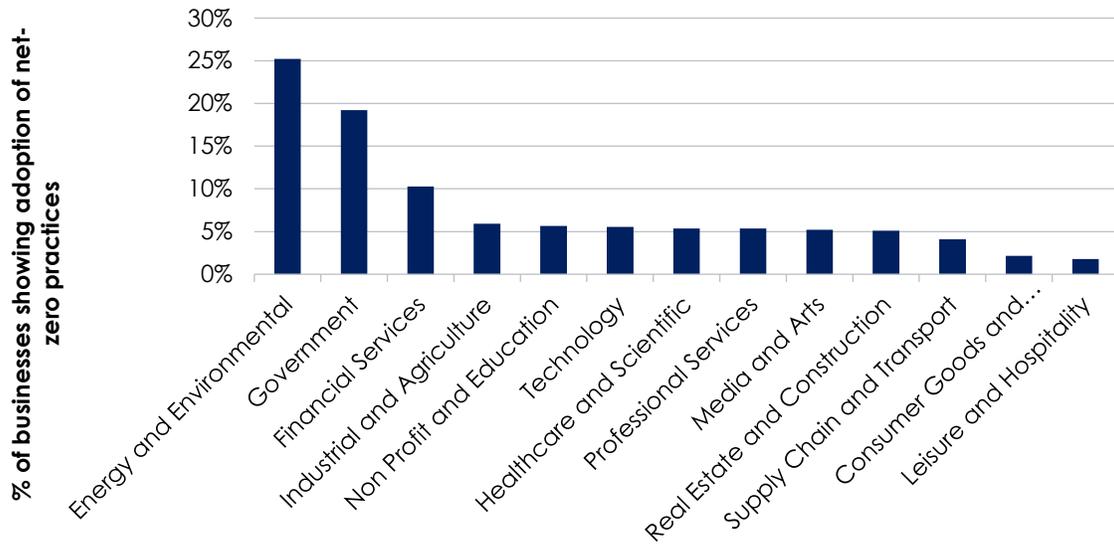
Source: glass.ai data
Number of records: 18,240

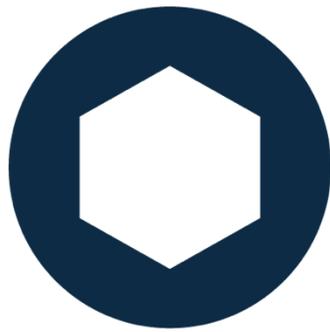**Chart 6: % of Businesses (by Sector) showing Growth Signals**

Source: glass.ai data
Number of records: 18,240

Chart 7: % of Businesses (by Sector) showing Import / Export Activity

Source: glass.ai data
Number of records: 18,240

glass.ai