AI Sector Study – Methodological Appendix

Approach & Methodology

1.1. Approach

The study uses a mixed methods approach, combining spanning academia, policy and investment spheres.

Stage 1 – Collation of initial data inputs: a long-list of 3,421 AI companies deemed to be potentially within the scope of the study was identified from numerous sources, predominantly via web intelligence generated by Glass.ai's web-reading capabilities. Just under one third of companies were also identified via other sources including but not limited to Bureau van Dijk's FAME, Beauhurst, Crunchbase, Lightcast and FDI Markets.

Stage 2 – Initial classification and filtering: A set of 217 key words and categories were identified through a combination of automated classification using Glass.ai language models and workshop sessions with representatives from academia, industry, government and the core study team. The long-list of potentially in-scope firms was refined and filtered to provide a shortlist of 3,170 in-scope AI companies.

Stage 3 – Survey design and administration: a detailed business survey was designed with input from the study steering group, including representatives from DSIT and academic and commercial research expertise. The survey was administered via multiple channels, including via telephone, e-mail and web-hosting. A total of 250 responses were received.

Stage 4 – Data augmentation: a series of manual data quality checks were conducted across key metrics (revenue, employment, location, classification) by both the core study team and DSIT analysts. Company data was then augmented using multiple data sources, providing a consistent set of key metrics for each UK AI business.

Figure 1.1 – Shortlisting & Augmentation Overview

Stage 5 – Regional & sub-sectoral analysis: more granular data on the trading locations of in-scope AI companies was gathered through web-intelligence and proprietary data sources, enabling a more detailed analysis of the trading presence of UK AI companies locally, and internationally. Employment is estimated through the use of company accounts where known or estimated. AI related employment is estimated at firm level using both survey estimates and web data.

Stage 6 – Sector modelling: The short-listed AI company set was used to produce analyses of, inter alia, the number, scale and location of UK AI companies, incorporations, investment, R&D expenditure and exports. GVA is calculated using GVA per FTE estimates based on a combination of company data, web intelligence and survey data.

Stage 7 – Qualitative interviews & case studies: in-depth follow-up interviews were conducted with 10 AI companies that responded to the survey. Findings were combined with those from 10 in-depth semi-structured strategic stakeholder interviews to address qualitative research questions regarding strengths, weaknesses, opportunities, challenges and risks to the UK AI sector.

Stage 8 – Analysis & reporting: findings from the quantitative and qualitative research were synthesised through steering group discussions and qualitative analysis sessions and triangulated to inform this baseline report.

1.2. Interpretation of Data

Artificial Intelligence activity in the UK is not defined by a formal Standard Industrial Classification (SIC) code[1]. This study therefore uses experimental methods to identify and quantify AI activity across traditional economic sectors. The approach and methodology are consistent with those employed to deliver analyses of the UK cyber security sector annually since 2018[2]. The data used to inform the study includes:
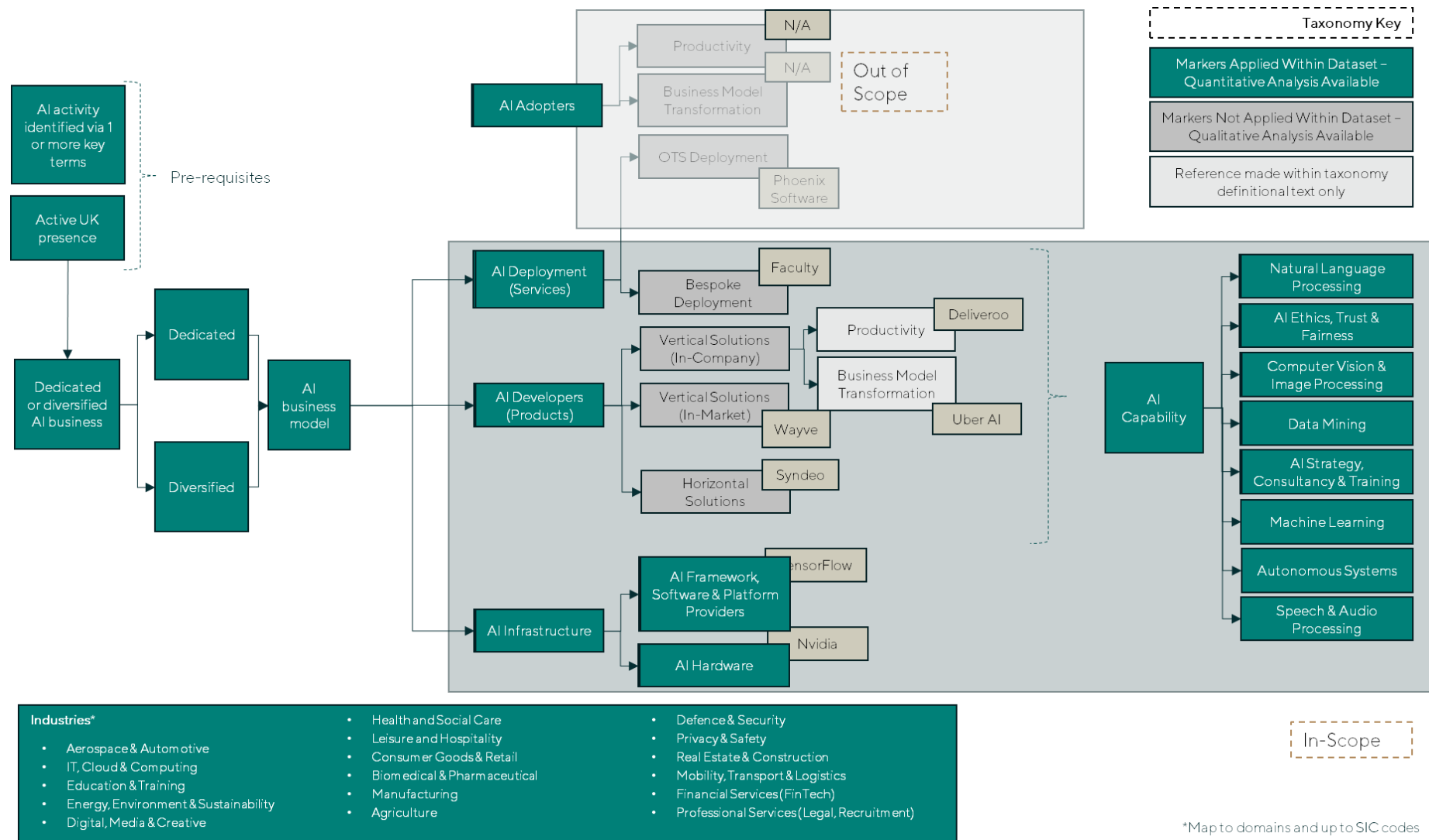
- Identification of AI firms according to an agreed taxonomy using AI driven language models applied across websites, news, social media, academic and official sources.

- Enrichment of web data using open and proprietary data sources including Companies House (company name, registration number, locations, incorporation date), Bureau van Dijk FAME (revenue, employment, profitability, remuneration, R&D spend) and Beauhurst (external grants, fundraisings, accelerator attendance, M&A activity).

Across this report, percentages from the quantitative data may not add to 100% due to rounding and / or the option to select multiple responses to certain survey questions. It is also important to note that the survey data is based on a sample of AI companies and are therefore subject to sampling tolerances. The overall margin of error for the sample of 250 AI companies (within a population of 3,170 companies) is between c.3 and c.6 percentage points at a 95% confidence level. The lower end of this range (3 percentage points) is used for survey estimates closer to 10% or 90%. The higher end (6 percentage points) is used for survey estimates around 50%. Data from the 22 qualitative consultations is intended to be illustrative of the key themes affecting AI activity in the UK generally, rather than a statistically representative view of AI sector businesses or investors.

---

[1] SIC codes are the current system of classifying business establishments and other statistical units by type of economic activity in which they are engaged.
[2] DSIT (2022) Cyber Security Sectoral Analysis 2022, accessible at [https://www.gov.uk/government/publications/cyber-security-sectoral-analysis-2022]

Figure – Full Taxonomy – Version 1

Taxonomy Key

- Markers Applied Within Dataset – Quantitative Analysis Available
- Markers Not Applied Within Dataset – Qualitative Analysis Available
- Reference made within taxonomy definitional text only

Pre-requisites
- AI activity identified via 1 or more key terms
- Active UK presence

Dedicated or diversified AI business
- Dedicated
- Diversified

AI business model

Out of Scope

AI Adopters
- Productivity — N/A
- Business Model Transformation — N/A
- OTS Deployment — Phoenix Software

In-Scope

AI Deployment (Services)
- Bespoke Deployment — Faculty

AI Developers (Products)
- Vertical Solutions (In-Company)
- Vertical Solutions (In-Market) — Wayve
- Horizontal Solutions — Syndeo
- Productivity — Deliveroo
- Business Model Transformation — Uber AI

AI Infrastructure
- AI Framework, Software & Platform Providers — TensorFlow
- AI Hardware — Nvidia

AI Capability
- Natural Language Processing
- AI Ethics, Trust & Fairness
- Computer Vision & Image Processing
- Data Mining
- AI Strategy, Consultancy & Training
- Machine Learning
- Autonomous Systems
- Speech & Audio Processing

Industries*

- Aerospace & Automotive
- IT, Cloud & Computing
- Education & Training
- Energy, Environment & Sustainability
- Digital, Media & Creative
- Health and Social Care
- Leisure and Hospitality
- Consumer Goods & Retail
- Biomedical & Pharmaceutical
- Manufacturing
- Agriculture
- Defence & Security
- Privacy & Safety
- Real Estate & Construction
- Mobility, Transport & Logistics
- Financial Services (FinTech)
- Professional Services (Legal, Recruitment)

*Map to domains and up to SIC codes