INNOVATION & RESEARCH CAUCUS

SCIENCE AS SUBSTRATE

Mapping the maturity of research supporting Innovate UK's innovation portfolio

IRC Report No: 079

REPORT PREPARED BY

George Richardson

Innovation Growth Lab, Nesta

Maria Brackin

Innovation Growth Lab, Nesta

Delivered with ESRC and Innovate UK

Authors

The core members of the research team for this project were as follows:

- George Richardson - Innovation Growth Lab, Nesta
- Maria Brackin- Innovation Growth Lab, Nesta

This document relates to IRC Project FFGtR010: What Work and When: Funding timing and Research Maturity of Innovate UK Investments

Acknowledgements

This work was supported by Economic and Social Research Council (ESRC) grant ES/X010759/1 to the Innovation and Research Caucus (IRC).

We are very grateful to the project sponsors at UK Research & Innovation (UKRI) for their input into this research. The interpretations and opinions within this report are those of the authors and may not reflect the policy positions of UKRI or its constituent councils. We would also like to acknowledge and appreciate the efforts of the IRC Project Administration Team for their support in preparing this report.

We would like to thank Abigail Stocker, Innovate UK, and Jess Ridsdalesmith and Robin Polding, Department of Science Innovation and Technology for their thoughtful discussion and guidance in this project.

About the Innovation and Research Caucus

The Innovation and Research Caucus supports the use of robust evidence and insights in UKRI's strategies and investments, as well as undertaking a co-produced programme of research. Our members are leading academics from across the social sciences, other disciplines and sectors, who are engaged in different aspects of innovation and research systems. We connect academic experts, UKRI, IUK and the (ESRC), by providing research insights to inform policy and practice. Professor Tim Vorley and Professor Stephen Roper are Co-Directors. The IRC is funded by UKRI via the ESRC and IUK, grant number ES/X010759/1. The support of the funders is acknowledged. The views expressed in this piece are those of the authors and do not necessarily represent those of the funders.

Contact

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: info@ircaucus.ac.uk

Cite as: Richardson, G. & Brackin, M. 2026. *Science as Substrate: Mapping the maturity of research supporting Innovate UK's innovation portfolio*. Oxford, UK: Innovation and Research Caucus

CONTENTS

Executive Summary

Background

Innovation agencies face a persistent strategic dilemma: when to support new products and services. Innovations do not emerge in a vacuum, but build in part upon the cumulative body of scientific knowledge, and upon secondary outputs of the research ecosystem such as talent and know-how. Intervening early, while the underlying science is still forming, can secure first-mover advantages and a national stake in emergent domains, but risks supporting innovations the knowledge base cannot fully advance. Supporting innovations built upon more established knowledge lowers technical risk, but may sink resources into companies late to market. Whether intentionally or not, innovation agencies may adopt a deliberate position along this spectrum, or diversify across it through the innovations and areas they choose to support.

Innovate UK is tasked with supporting business-led innovations to start and grow, including those which have emerged from scientific research. Over the course of its existence, the organisation has undergone strategic shifts that have altered the nature and domains of innovation it has prioritised, and the instruments that it has used. Its most recent set of priorities place high importance on deep tech, based on breakthrough discoveries. Such shifts highlight the need for Innovate UK, and other innovation agencies, to understand the academic foundations upon which new innovations are being built.

Approach

This report asks how advanced the underlying knowledge is upon which Innovate UK-funded innovations are built, and how this has changed over two decades. We examine this by analysing whether the funding portfolio of Innovate UK projects has aligned with trends in the maturity of the relevant research areas.

We linked 29,262 Innovate UK-funded projects to the global scientific literature through research topics using GtR+ – an enhanced version of UK Research and Innovation's Gateway to Research funding data. A large language model assigned each project a research topic from its abstract. We then assembled annualised histories for each of the resulting 1,878 topics based on publication volumes and growth rates. Topics are categorised as niche, emerging, active, and mature. Weighting these categories by Innovate UK grant spend, we established where public funding for innovation sits relative to the developmental stage of the underlying science.

Key Findings

Our analysis reveals, first, that over its history, Innovate UK's portfolio spans innovations building on the full range of research maturity rather than concentrating within a single stage. Examining the year in which each topic received its maximum Innovate UK grant funding, peak investment points are distributed across innovations drawing on niche (25.1%), emerging (23.4%), active (31.2%) and mature (20.2%) research.

Second, the balance of the portfolio has shifted over time, revealing two main strategic eras. Between 2006 and 2014, resources were concentrated on innovation projects with foundations in emerging and active research – high-growth areas, consistent with bridging the translation gap for current research trends. From 2015 onwards the portfolio diversified substantially, extending to include innovations built on niche and mature research, with lower growth topics accounting for approximately two thirds of funding by 2022. The most recent years indicate a return towards a balanced portfolio, with support for active and emerging domains.

Third, the six frontier technology areas specified by current UK government industrial policy, and now being prioritised by Innovate UK, present distinct maturity profiles. They rest upon 146 research topics, a substantial proportion of which have no innovation projects associated within Innovate UK's portfolio according to our dataset. Within artificial intelligence, 90% of funding since 2020 has supported projects underpinned by low-growth niche or mature topics. Cybersecurity funding rests almost entirely upon mature research. Semiconductors present the inverse profile, having a portfolio of innovation projects built on a predominantly mature research base that has recently moved towards niche and emerging topics.

Limitations and Future Work

These findings should be read with three significant qualifications. Innovate UK rarely references the maturity of the underlying research base within its strategies or delivery plans; this analysis therefore does not measure a gap between intent and achievement, but offers a lens the agency has not previously applied. Important further research should examine the link between research maturity in Innovate UK's portfolio and company outcomes.

We also measure maturity only relative to the topics associated with Innovate UK projects, not against a global benchmark. A holistic study would expand the baselines selected, identifying where Innovate UK builds on lifecycle phases of science within its own remit, across global research, and among UK research.

Additionally, GtR+ is an experimental dataset and as such, several data limitations posed challenges to this study. Only 35.8% of Innovate UK grant funding could be linked to a research topic, principally because many project records lack valid abstracts, leaving a partial view of the portfolio. Perhaps more consequential, is the absence of post-project outcome metrics for Innovate UK awards within GtR+. Tracking these would move beyond describing where public capital sits within the scientific lifecycle towards testing which timing strategies succeed commercially.

Recommendations

By mapping how Innovate UK's portfolio corresponds to the underlying maturity of academic fields, we can describe the organisation in terms of its role in knowledge translation and implicitly characterise its strategy in terms of the scientific foundations it seeks to exploit. In doing this, we are able to present an analysis of Innovate UK's past activities, and an outlook for its future investments to support a targeted focus on frontier technologies.

Extending this analysis to study the impact that scientific foundations of innovation have had on commercial outcomes would allow Innovate UK to understand the risk and opportunities in its future portfolios. In programme delivery, such the mapping presented here could be used in portfolio analysis to map opportunities and unexploited translation gaps in the research landscape, in conjunction with other decision making criteria.

These possibilities, as well as future policy evaluation, would benefit from an improved data landscape. GtR+ is a promising demonstration of the potential for enriched funding data, but remains experimental. We suggest two main actions that Innovate UK (and UK Research and Innovation) might take to begin bridging the gap between administrative and analysis-ready data. One is that some augmentations should be generated at the point of application, rather than applied retrospectively. For example, topic tagging by applicants or manually or using a human-AI approach would generate a better classified dataset for both funded and unfunded proposals, at very low cost to all parties.

The other is that some level of outcome tracking for Innovate UK projects should be integrated directly within data that can be used by internal analysts or published in Gateway to Research. This would unlock the ability to create significantly richer policy insights that can inform future decisions.

1. Introduction

Innovation agencies in the UK fund and support innovative activities that lead to new solutions which address societal and economic goals. As the UK's national innovation agency, Innovate UK is responsible for supporting innovative firms to grow and scale, and therefore plays a critical role in translating research and development into market solutions. Its core mission is to help UK businesses thrive through the development and commercialisation of new products, processes, and services. New innovations do not happen in a vacuum. They rely, in part, on the cumulative body of scientific literature and knowledge accrued through non-commercial research activities (Narin et al. 1997; Ahmadpoor and Jones 2017). While academic research generates foundational knowledge documented in peer-reviewed publications, Innovate UK targets downstream, business-led applied research and experimental development that builds directly upon this evidence base. Whether directly or indirectly, Innovate UK-funded projects may draw on prior scientific literature, sometimes recombining existing ideas, sometimes relying on entirely new seeds provided by research and development (Nelson 1959; Arrow 1962; Weitzman 1998; Fleming 2001). They can also draw on secondary outputs of scientific research such as talent and know-how, which are a product of the size and strength of academic communities.

In general, innovation agencies face a persistent strategic dilemma: when to support new products and services. Intervening early to support nascent innovations and pioneering firms can yield first-mover advantages and secure a national stake in emergent domains, but risks backing immature ventures that fail commercially (Lieberman and Montgomery 1988). Conversely, supporting innovations built on more established knowledge may be safer from a technology perspective, but risks companies being late to market. Innovation agencies must wrestle with the choice between delivering programmes that support innovations primarily characterised by a particular level of knowledge maturity, or spreading their bets by supporting a diverse portfolio. This dilemma can be resolved onto two distinct dimensions rather than one: how much research exists in an area, and how fast it is accumulating. The two do not always move together, and the areas with the least research are not always the newest.

Addressing how to allocate funding to innovation according to this problem might be an explicit consideration, because an agency is tasked with supporting particular kinds of innovative activity, or an implicit directive, because government priorities push strategic focus towards industrial sectors of national importance and other societal goals.

Since the inception of Innovate UK's predecessor, the Technology Strategy Board (TSB), the strategic focus and narrative of the UK's public innovation agency has undergone several significant revisions. Some of those changes have included shifts to the relationship between the organisation and the research system, such as the point along the research-to-market pipeline at which it would intervene. In 2008, TSB's *Connect and Catalyse* 2008-2011 strategy put scientific readiness of an innovation within its core investment criteria, and specified a Collaborative Research and Development grant mechanism, whose award size would be modulated according to whether a project was market ready or “coming from the knowledge base” (Technology Strategy Board 2008). It also set out plans for Knowledge Transfer Networks that would focus on exchange of knowledge between academia and business, implicitly requiring sufficiently developed academic fields to draw upon.

In 2017, Shaping the Future, an Innovate UK delivery plan, presented a dual focus on both emerging technologies, defined as “newly developed science whose best applications and value have yet to be realised” and enabling technologies “based on more established research” (Innovate UK 2017). This was contrasted by the 2021 Plan for Action, which references the need for a strong innovation ecosystem “from idea through to commercialisation, adoption, and diffusion” (Innovate UK 2021). Alongside this wide lens, Innovate UK targets seven technology areas, noting that they cover a range of technology maturities “from early-stage emerging technologies, such as quantum technologies, to the latest refinements of enabling technologies, such as electronics, photonics, advanced materials, robotics, and biosciences.”

Innovate UK’s current strategy places central importance on supporting “deep tech” innovations - those which rely heavily on recent scientific breakthroughs and substantive advances in engineering (Innovate UK 2026). It also places strong emphasis on six *frontier technologies* prioritised in UK government industrial strategy, and that overlap strongly with contemporary areas of academic research: advanced connectivity technologies, artificial intelligence, cyber security, engineering biology, quantum technologies, and semiconductors (Department for Business and Trade and Department for Science Innovation and Technology 2025). This reinforces the importance of the underlying research base as an input to the activities that Innovate UK aims to nurture.

Together these snapshots of Innovate UK’s past and present offer a picture of an organisation that supports innovations at varying stages of maturity, and which therefore draw on science in different ways. In several cases, specific mechanisms are designed to ensure that the organisation has the flexibility to support projects that are more emergent versus those which are based on more established knowledge. These shifts in high level strategy, and the need for responsiveness at a programme level, create a requirement for the agency to understand the size, pace and nature of developments in fields upon which new deep tech innovations are built.

The purpose of this project is to use the Gateway to Research “plus” (GtR+) dataset - an enhanced version of the UKRI funding data from the original Gateway to Research - to better understand how Innovate UK makes funding decisions, and how it can steer future decisions. We ask a deceptively simple question: *how advanced is the underlying knowledge upon which Innovate UK-funded innovations are built*?

To answer it, we must define and measure *research maturity* - the stage of evolution and readiness of the scientific knowledge base upon which an innovation rests - and apply this definition to understand past and future strategic trends within the Innovate UK portfolio.

We approach this by establishing a link, both theoretically and in concrete data, between Innovate UK-funded innovation projects and the broad scientific base that underpins them. Rather than relying on narrow case studies, we took a comprehensive approach, constructing a dataset of over 40 million academic publications from OpenAlex to map the knowledge supporting Innovate UK projects. By measuring how investments correspond to the underlying maturity of academic fields, we can describe what Innovate UK looks like as a funder and its role in knowledge translation, beginning to build a profile of their investments. In doing this, we are able to present an analysis of Innovate UK's past activities, and an outlook for its future investments and support toward a targeted focus on frontier technologies.

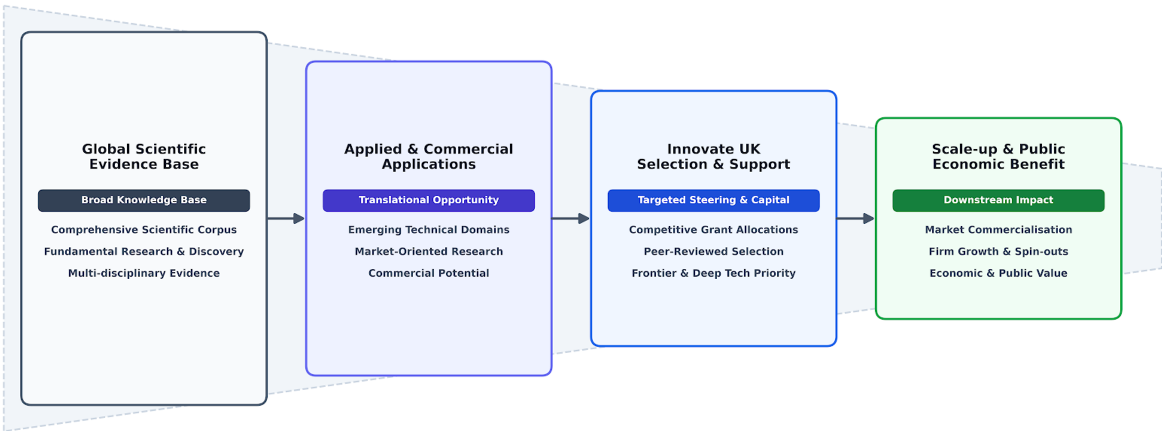

Figure 1.1 - Diagram showing the global scientific knowledge base as an (indirect) input into innovation activity, of the kind funded by Innovate UK.

In Section 2, we present our methodology, describing our approach (2.1), the datasets assembled (2.2) and methods used for topic classification and research maturity (2.3) and reporting of the outputs from our large language model (LLM) classifications (2.4). In Section 3, we describe how fields and topics evolve. We then present a profile of Innovate UK as a funder in Section 4, describing the type of innovations funded (4.1) and the maturity of the underlying research base (4.2), including how Innovate UK strategically invests. In Section 5, we discuss how the six frontier technology areas have previously been supported, looking ahead towards future potential. In Section 6 we discuss our experience using the GtR+ dataset for this research. Finally, in Section 7 we discuss our findings, acknowledge limitations and conclude with suggestions for future research directions.

2. Measuring research maturity

2.1 Design

Our approach rests on the principle that many innovations - regardless of how commercial they become - build, whether directly or indirectly, upon a foundation of academic research. By linking the Gateway to Research "plus" (GtR+) dataset directly to OpenAlex, we can systematically measure the maturity and dynamics of the global scientific base underpinning Innovate UK projects.

To operationalise this, our analytical framework comprises four steps:

1. Topic classification: Assigning a specific OpenAlex *topic* to each Innovate UK project in order to link it directly to its underlying scientific base.
2. Frontier technology classification: Identifying which of the OpenAlex topics are associated with one of the 6 *frontier technologies*.
3. Scientific base extraction: Collecting comprehensive *publication data* from OpenAlex for the identified topics.
4. Volumetric maturity measurement: Calculating the readiness and evolution of that scientific base using a *volumetric framework* for research maturity.
5. Frascati classification: Assigning one or more *Frascati categories* to every journal in which papers in the Innovate UK-linked topics have been published, to further understand research maturity.

2.2 Data

The datasets resulting from the steps above cover 29,262 Innovate UK-funded projects between 2006 and 2025 and an underpinning global scientific corpus of over 40 million OpenAlex articles spanning 1,887 topics.

This study relies on two primary data sources: OpenAlex for global scientific outputs and GtR+ for UK public funding records. Academic publications serve as our primary unit of analysis to track the evolution of research topics, while GtR+ provides an enriched database of projects supported by Innovate UK.

OpenAlex is an open, comprehensive catalogue of global scholarly literature. To categorise research, it uses a four-level hierarchical taxonomy developed in collaboration with the Centre for Science and Technology Studies (CWTS) at Leiden University (Jan van Eck and Waltman 2024):

1. Domains (4): The highest level, dividing science into broad realms (Health Sciences, Life Sciences, Physical Sciences, and Social Sciences).
2. Fields (26): Major disciplinary branches within each domain (e.g. Medicine, Computer Science, Materials Science).

3. Subfields (254): Specialised branches within a field (e.g. Health Informatics, Chemical Engineering).
4. Topics (~4,500): Specific, highly granular research clusters representing active research communities, targeted problems, and specialised technological domains (e.g. Artificial Intelligence in Medicine).

We selected *topics* as our primary unit of scientific resolution. Broader disciplines (fields or subfields) aggregate vast areas of research, masking critical shifts in specific technologies. Topics offer the necessary granular precision to observe emerging scientific breakthroughs and measure maturity in specialised technical domains.

GtR+ provides enriched metadata on UK publicly-funded research and innovation. While the complete dataset spans the broader UK Research and Innovation (UKRI) portfolio - encompassing awards across all constituent Research Councils, Research England, and Innovate UK - our study focuses strictly on Innovate UK-funded projects to isolate business-led R&D and commercialisation funding. A key element of this data is the extract of project subjects, which classifies Innovate UK projects into subfields using an approximation of the OpenAlex classification schema. This was a valuable starting point for our own classification.

2.3 Methods

2.3.1 Topic classification

Our first task was to assign a unique, primary topic to every Innovate UK project, by interpreting the text within the abstract. Projects without valid abstracts were excluded from this classification.

Evaluating 29,262 Innovate UK projects directly against more than 4,500 potential OpenAlex topics presents a significant computational challenge. Running a single-pass classification across the entire topic hierarchy for every project is computationally inefficient and likely to result in degraded classification accuracy. To resolve this, we designed a two-stage LLM classification pipeline using OpenAI's gpt-4o-mini:

- Stage 1 (subfield routing): The model identifies the primary OpenAlex subfield corresponding to the project's methodology. This drastically narrows the search space from thousands of topics to a focused subset.
- Stage 2 (granular topic classification): The model evaluates the project abstract strictly against the candidate topics within that pre-selected subfield to determine the final, granular topic.

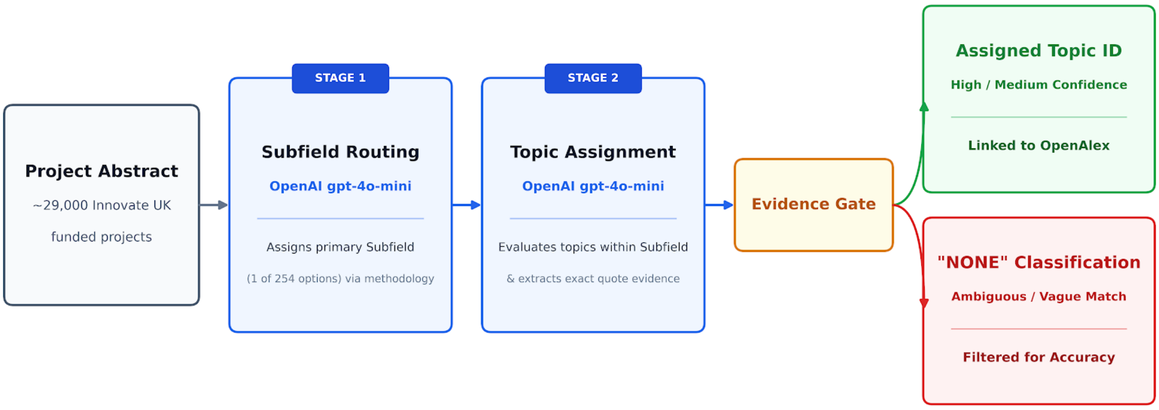

Figure 2.1 - Methodological diagram summarising the multi-step subfield and topic assignment methodology.

A core principle of our classification design was to prioritise *accuracy* above total portfolio *coverage*. Rather than forcing every project into a topic, the classifier was required to output explicit technical reasoning and direct textual evidence for its decisions. If an abstract was too vague or lacked technical alignment with the available topics, the model was instructed to assign a “NONE” classification. This strict filtering mechanism deliberately removed ambiguous matches, ensuring that subsequent maturity calculations relied on high-confidence links (see the Appendix for more details on classifier implementation).

2.3.2 Frontier technology areas

To align our research maturity analysis with current UK policy priorities, we evaluated every topic in the global OpenAlex taxonomy (~4,500 topics) against Innovate UK’s six frontier technology areas using an automated LLM classification workflow. The pipeline evaluated topics within their parent hierarchy (*Domain* > *Field* > *Subfield*) against explicit criteria, applying refined boundary logic to distinguish core technical R&D and domain-specific governance—such as core machine learning algorithms alongside AI ethics, policy, and legal impacts—from broad traditional engineering or generic software tools. This workflow assigned a binary flag (TRUE or FALSE), technology category, and technical justification to every record, identifying topics across the global research base that directly underpin the six frontier technology areas. For full details of the classification procedure, see the Appendix.

2.3.3 Scientific base extraction

Publication records were gathered directly via the OpenAlex API to perform two primary analytical functions:

1. Volumetric time-series extraction: We retrieved annual publication counts for every topic, underlying subfield, and overarching field within the Innovate UK project portfolio. These macro- and micro-level counts provided the data required to compute our volumetric research maturity metrics.

2. Article-level metadata extraction: We extracted granular article records for over 40 million individual publications, capturing every paper published during the 1996–2025 period across all topics linked to Innovate UK projects. These individual article records captured publication dates, abstracts, and journal titles, from which a deduplicated set of unique journal names was compiled to execute our Frascati R&D classification pipeline.

2.3.4 Volumetric maturity

To measure the structural evolution of the research landscape, we applied a volumetric approach grounded in established scientometric principles. The underlying premise, long established in bibliometric literature, such as Derek de Solla Price's foundational work on the exponential growth of science (1963) and more recent analyses by Bornmann and Mutz, is that the maturity and momentum of a scientific discipline can be quantitatively observed through its output and expansion rates over time (Price 1963; Bornmann et al. 2021). By tracking both the absolute volume of activity and its growth, we can systematically classify the developmental stage of any given technological topic. We operationalised this using two distinct lenses: a scientific baseline (Approach A) and a capital commitment lens (Approach B). This approach is derived from previous work used to investigate research and innovation trends (Kanders and Smith 2021).

Rather than looking at single-year spikes - which can be misleading - we measure growth over a rolling 3-year window. Each year, we consider all active research topics and find the exact middle point (median) for both volume (total papers published) and growth (3-year growth rate).

As a result, topics are categorised into four maturity bins:

1. Niche (low volume, low growth)
2. Emerging (low volume, high growth)
3. Active (high volume, high growth)
4. Mature (high volume, low growth)

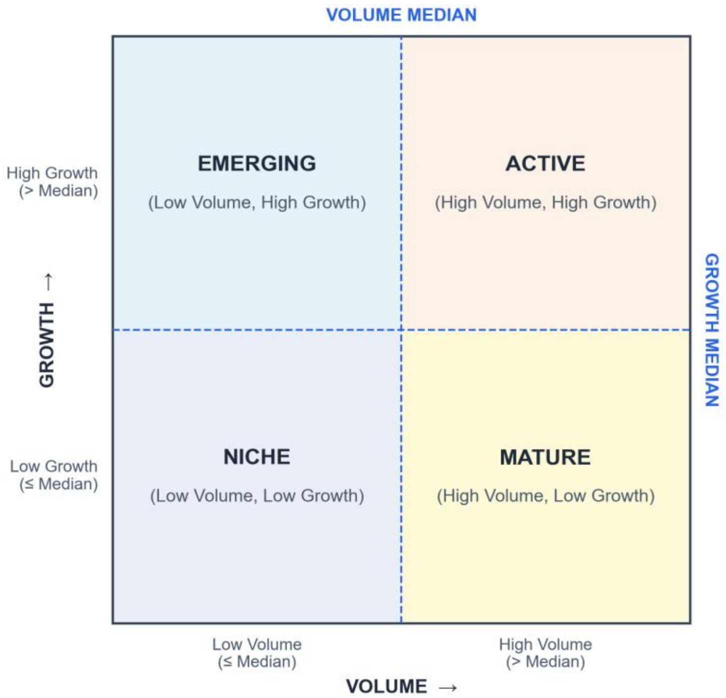

Figure 2.2 - Conceptual diagram demonstrating how research areas are categorised into maturity definitions according to their growth and volume dynamics.

It is important to note that the categories exist along two independent dimensions, not a single ordering. In particular, a topic that is genuinely new will typically show very high growth from a small base and will therefore appear as emerging. Niche denotes areas that are both small and not currently expanding: specialised, plateaued or declining literatures rather than nascent ones.

Scientific maturity

The first method, Approach A, measures publication volumetrics to establish a baseline of scientific maturity, based on activity levels. This approach asks *how mature a research topic is within the Innovate UK project portfolio*. The inputs for this measurement are the total number of academic papers published in a specific topic each year, alongside the 3-year growth rate of those publications.

Capital-weighted maturity

The second method, Approach B, applies a capital commitment lens to measure the funding-weighted maturity of the active portfolio. It shows *what percentage of the annual Innovate UK budget went to topics of different scientific maturity*. Rather than treating every research topic equally as in Approach A, this approach evaluates how Innovate UK allocates its financial capital - measured in total active grant spend (£) - across the scientific maturity stages defined by global publication trends. The inputs and cutoff points are the same as in Approach A.

The approaches are summarised below:

Table 2.1 - Summary of approaches to measuring maturity with a volumetric methodology.

Dimension	Approach A: Scientific maturity (publications)	Approach B: Capital-weighted maturity (investments)
Primary Question	How mature are research topics within the Innovate UK project portfolio?	What percentage of the annual Innovate UK budget went to topics of different scientific maturity?
Portfolio	Unweighted count of unique active topics in each maturity bin	Capital-weighted (£): total active grant funding stock allocated to topics in each bin
Volume	Total global academic papers published per topic per year	
Growth	3-year percentage change in global published papers	
Cutoff Points	Annual median of global paper volume and 3-year publication growth rate	
Strategic Insight	Thematic portfolio breadth and risk exposure across the Innovate UK research ecosystem	Capital concentration, financial scale, and prioritisation of public funding relative to scientific momentum

Strategic archetypes

By comparing Innovate UK's funding trajectory on a specific topic to trends in global scientific publishing, we assess whether funding is growing or shrinking relative to scientific activity. This analysis seeks to answer a central question: *For each individual topic, how is the Innovate UK funding trajectory aligned with scientific momentum within the portfolio?*

To evaluate strategic alignment, our framework retains the same publication-derived maturity categories from Approach A (niche, emerging, active, mature) and cross-references them with Innovate UK's internal 3-year grant funding growth trajectory. To establish a reliable baseline for the funding growth trajectories, the analysis begins in 2009, using the 2006–2008 period as the initial window to calculate baseline investment trends.

By overlaying IUK's spending momentum onto global research baselines, we classify funding behavior into 5 "strategic archetypes":

1. Early bet / Anticipatory: UK public funding is scaling into a topic that remains small and slow-growing globally. This may represent an anticipatory bet on an area before the literature responds, or continued reliance on a small area that is not expanding.

2. Ecosystem co-alignment: Both global research momentum and UK grant funding are high-growth. Global paper growth is at or above the global median (emerging or active), and UK grant spend is expanding (3-year funding growth > 0%).

3. Capital gap / Lagging: Global academic momentum is surging, but UK grant funding is low or stagnant. Global paper growth is at or above the global median (emerging or active), but UK grant spend is flat or declining (3-year funding growth ≤ 0%).

4. Sustained sector support: Global publication growth has leveled off, but UK grant funding remains high. Global paper volume is at/above median but global growth has dipped below median (mature), while UK grant spend continues to expand (3-year funding growth > 0%).

5. Peripheral / Low priority: Low volume and growth across both global publications and UK funding. Global science is either low-activity (niche) or plateaued (mature), and UK grant spend is flat or contracting (3-year funding growth ≤ 0%).

By evaluating these signals together, the framework points towards, though by no means confirms, strategic intent and portfolio gaps - for instance, distinguishing between funding that runs ahead of the science and funding that lags behind it.

Peak investments

In addition to tracking continuous annual funding flows, our framework evaluates the specific year of *peak investment* for each topic across the portfolio. A topic's peak investment year is defined as the single calendar year in which Innovate UK committed its maximum volume of grant funding (£) to projects within that topic. While longitudinal metrics illustrate portfolio trends, isolating the peak funding year pinpoints the moment of maximum capital concentration for a given technology. By cross-referencing a topic's scientific maturity stage at its exact investment peak, we can assess the timing posture of public interventions. This diagnostic tells us where Innovate UK's largest commitments fall against both dimensions: whether they arrive while a literature is expanding rapidly, after it has reached scale, or in areas that are neither large nor growing.

Portfolio maturity balance

To quantify the overall structural posture of Innovate UK's portfolio, we translate complex funding distributions into a single, accessible balance score using a spatial distance calculation (*vector analysis*). Rather than examining hundreds of individual topics independently, this technique summarises the entire portfolio by treating each of the four scientific maturity stages—niche, emerging, active, and mature—as its own dimension. The agency's actual allocation in any given year forms a single coordinate in this four-dimensional landscape, defined by the percentage of total topics or funding assigned to each stage. To evaluate where Innovate UK lands, we compare its position against a neutral benchmark: a theoretical, perfectly balanced portfolio where capital or attention is distributed evenly, taking up exactly twenty-five percent in each of the four quadrants.

This neutral benchmark serves as a statistical baseline, enabling us to observe the extent to which Innovate UK's empirical portfolio distribution deviates from a uniform spread. If the portfolio is balanced, then the distance is zero; if not, the distance increases as the portfolio becomes concentrated in fewer categories. For more details about the vector analysis, see the Appendix.

2.3.5 Frascati maturity

The OECD Frascati Manual (2015) serves as the international standard for defining and classifying Research and Development (R&D). Used widely by governments, research councils, and statistical agencies, the framework provides a standardised vocabulary to categorise scientific activities according to their intent, methodology, and proximity to practical application.

Unlike our volumetric approach, which counts individual published papers, the Frascati classification exploits information about the journal that each paper was published in, assigning it a category from the Frascati Manual. The classification of a journal does not change over time. This venue-level approach addresses our question: *what proportion of research underpinning a given topic within the Innovate UK portfolio is published in basic, applied, and experimental development journals?*

To assign Frascati categories, we first extracted unique journal names from individual publication records across the OpenAlex corpus for each year, deduplicated them across the dataset, and linked each journal to its corresponding journal-level OpenAlex metadata. Next, we used an automated LLM classifier to divide the journals into five distinct categories:

1. Basic research: A journal publishing experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any specific practical application or commercial use in view.

2. Applied research: A journal publishing original research conducted to acquire new knowledge, directed primarily towards a specific, practical aim or commercial objective.

3. Experimental development: A journal publishing systematic work drawing on existing knowledge gained from research and practical experience, directed towards producing new or improved products, processes, or services.

4. General: A journal publishing a mixture of basic, applied, and experimental development without a single dominant focus.

5. NONE: A journal publishing research not clearly related to Research and Development (R&D), for example many topics within the arts and humanities.

All journals were assigned a primary classification, representing over 50% of the journal's mission or target output. If relevant, journals were also assigned a secondary and tertiary category (General and NONE journals received only a primary classification). NONE journals were excluded from this maturity analysis.

Rather than subjecting the classifications to complex quantitative transformations - such as Euclidean vector space modeling - the Frascati framework is intentionally applied as a straightforward, complementary diagnostic tool. We implement two distinct aggregation approaches at the topic level. The first approach evaluates topic maturity using the *primary category only*, attributing a journal's publication volume entirely to its main operational focus. The second approach applies a *weighted measure* that incorporates primary, secondary, and tertiary classifications according to fixed proportional weights (50%, 30%, and 20%, respectively).

Further details about classifier implementation, and the results of our basic analysis, are presented in the Appendix.

2.4 Classification outcomes

Following the method described in Section 2.3, we used a Large Language Model (LLM) to identify the topics associated with the Innovate UK-supported projects, to flag topics associated with the 6 frontier technology areas, and to assign Frascati categories to the journals in which articles were published within these topics.

First, we removed 8,921 project records in GtR+ for which there was no valid abstract. Then, we used the LLM classifier with its two-step procedure to assign the topic. In Step 1, it successfully identified 269 subfields, covering 100% of the 29,262 projects. In Step 2, the LLM identified 1,878 topics, finding a match for 70.0% of the projects. Of these, 69.3% were assigned based on a "strong single match" with a specific topic among the options within the subfield, and only 0.3% based on "primary", "marginal" or "ambiguous" matches. For the 30.0%

of projects unassigned, all were assigned the “NONE” category because there was “no relevant match” between the project description and the topic options.

The LLM next identified 137 OpenAlex topics as related to the 6 frontier technology areas. See Section 5 for summary statistics about the topics identified for the frontier technologies, and the Appendix for a full list of the topics associated with each technology.

Using a list of deduplicated journal names and the journal-level metadata on OpenAlex, the LLM assigned Frascati categories to 68% of the journals covering topics attributed to Innovate UK projects, excluding 32% because they were not clearly linked to R&D. Considering the primary classifications, the LLM identified 3% of journals as basic, 52% as applied, 0% as experimental and 13% as general (32% were given the classification of NONE and excluded). When taking into account the weighting (inclusive of secondary and tertiary classifications), 12% were identified as basic, 39% as applied, 4% as experimental and 13% as general.

For visualisations of the classifications, and the top Subfields by number of publications, see the Appendix.

3. How Research Maturity Evolves

As discussed in Section 2.2, academic research can be described in terms of broad categories (domains, fields), more specific subsets (subfields), and granular areas (topics). We focus on topics because they are most relevant when considering specific technological breakthroughs, targeted commercial applications, and active innovation communities where business-led R&D takes place. However, we sometimes consider field level dynamics, to provide context about the broader disciplinary stability and underlying research capacity.

Every area of research has a different maturity profile, with changes over time shifting research between states. Rather than remain static, a field or topic will undergo gradual structural transitions driven by internal dynamics, such as breakthrough discoveries and technological advances, and external drivers, including institutional investment, societal needs, and economic signals. A field may originate within a niche state characterised by low output and low growth. They may remain niche indefinitely, but can transition to emerging or active phases, for example as a result of technical breakthroughs triggering a rapid surge in scientific interest. Over longer horizons, as fundamental scientific questions are resolved and research methods standardise, publication growth can moderate, shifting the field into a mature state defined by high publication volume alongside steady, incremental progress.

However, these field-level transitions are rarely linear or uniform across disciplines. While certain broad fields exhibit long periods of stability in high-volume states, others demonstrate cyclical movements or structural plateauing as academic focus redistributes into newer sub-disciplines. Establishing this macro-level profile provides context about the structural trajectory of an overarching field, which defines the baseline capacity, institutional infrastructure, and evidence base within which more granular subfields and technological topics evolve.

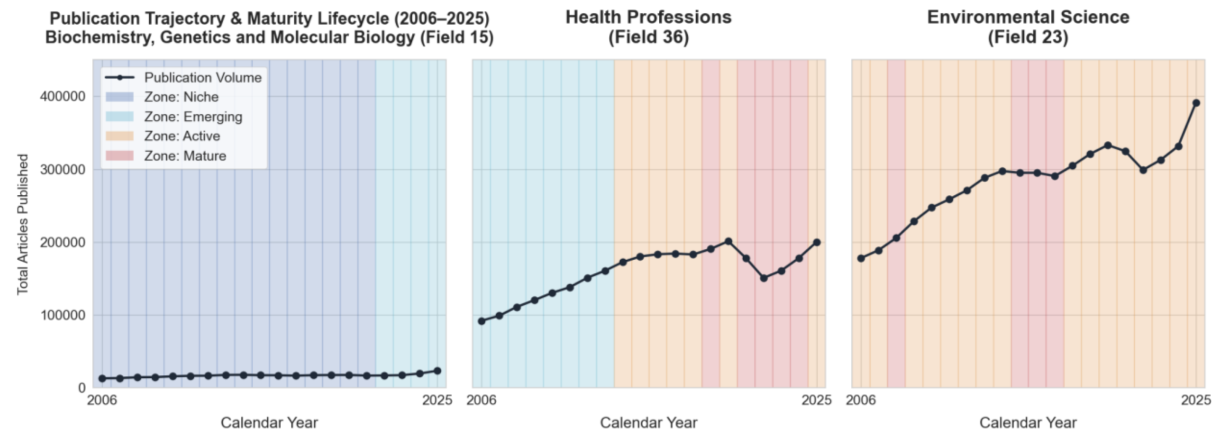

Chart 3.1 - Field evolution examples: Biochemistry, Genetics and Molecular Biology, Health Professions and Environmental Science.

Research topics also undergo transitions, although these are modified by their granularity in comparison to field level dynamics. While fields can only shift from one state of maturity to another if multiple academic communities advance multiple topics, the topics themselves are much narrower and therefore sensitive to changes in research trends. For example, the immediate societal pressures of a particular health crisis might accelerate research in that topic area, because additional funding crowds in new researchers and therefore changes activity, while the overall containing field will only experience a small shift.

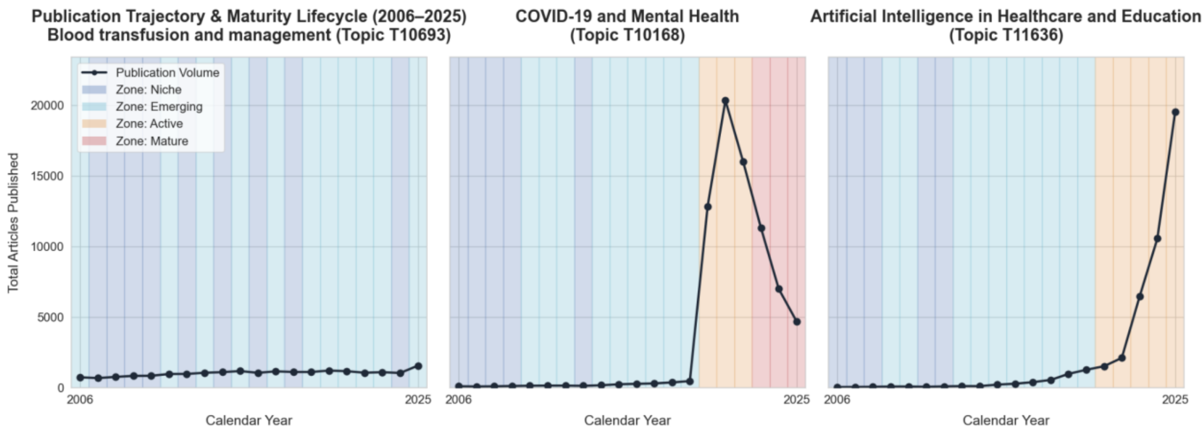

Chart 3.2 - Topic evolution examples: Blood transfusion and management, COVID-19 and Mental Health, Artificial Intelligence in Healthcare and Education.

4. Innovation funding and the research base

In this analysis, we focus exclusively on funding awarded by Innovate UK, which accounts for £17.73 billion (26.0%) of the total financial commitment captured within the GtR+ dataset. Overall funding levels expanded gradually between 2006 and 2021, before reaching a peak in 2023 and subsequently declining.

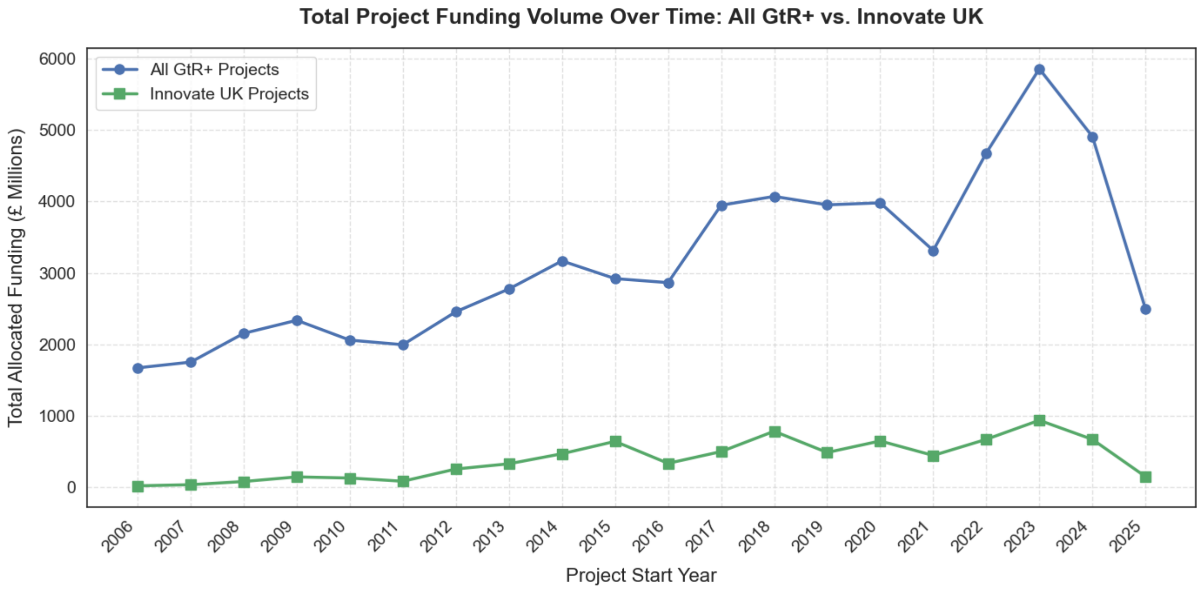

Chart 4.1 - Innovate UK funding versus all source funding within GtR+ dataset, 2006-2025.

Across the evaluated timeline, Innovate UK has awarded £17.73 billion in overall funding. The distribution of this capital across projects associated with individual topics varies considerably over time, alternating between periods of broad thematic dispersion and concentrated capital commitment. Topic level funding intensity reached its lowest point in 2011, averaging £326,717 per topic, while capital concentration peaked in 2018 at £1.53 million per topic. These investment figures reflect only the portfolio mapped to successfully classified topics, accounting for 35.8% of overall Innovate UK grant funding.

The downward trend in recent years reflects a 78% decrease in the overall project funding allocated by Innovate UK between 2023 and 2025, according to official UKRI figures (UKRI, 2026a; UKRI 2026b). However, our sample - and the results presented below - cover only part of this total funding. The GtR+ dataset includes 43.4% of the total Innovate UK funding spend, and our topic-linked analysis dataset includes only 27.0% of funding spend. For further details of Innovate UK funding offered and allocated, see the Appendix.

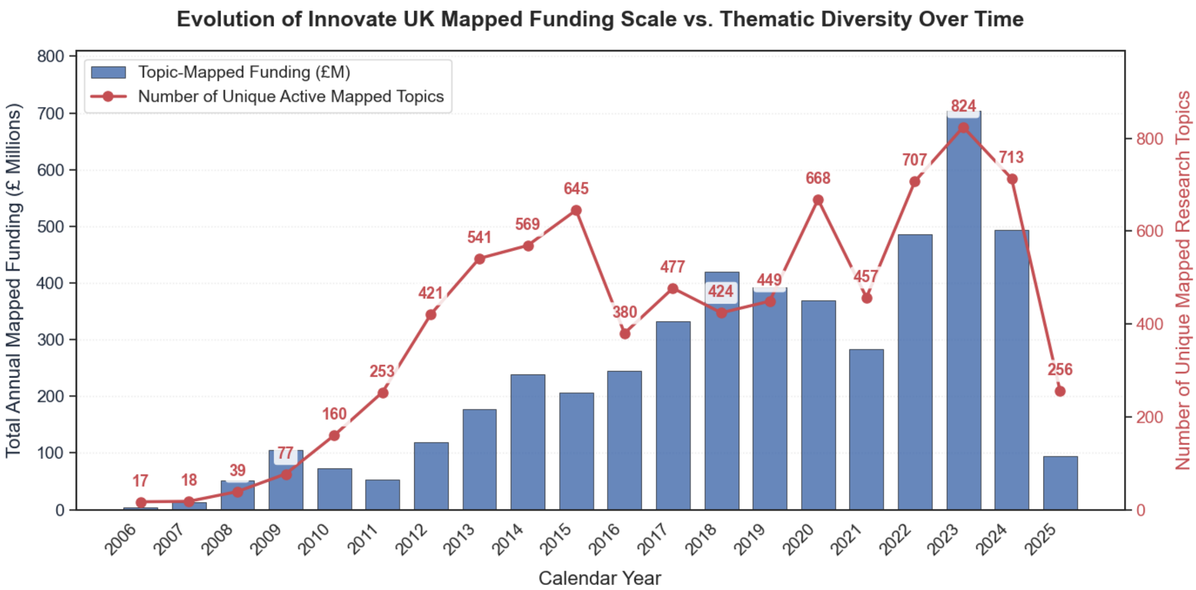

Chart 4.2 - Innovate UK funding amount and number of topics associated with supported projects, 2006-2025.

4.1 Profiling Innovate UK - what research supports Innovate UK fund?

During funding rounds, Innovate UK selects individual projects for support, implicitly choosing innovations that build on specific areas of R&D alongside broader strategic considerations. A central objective of our analysis is to identify which topics are relevant to the funded innovation projects and to evaluate the maturity of that underlying research base. To contextualise these topic selections within the wider research ecosystem, we examine granular topics and trends at the subfield level.

When examining the composition of Innovate UK's investments over time, we see high portfolio fluidity at the granular topic level, and relative stability at the subfield level. Specific topics funded vary significantly from year to year, with only 3 topics remaining in the top 20 over a 10 year period. However, funding remains fairly consistent within a core group of underlying subfields, with 13 staying in the top 20 over the same period. This is consistent with Innovate UK having long term strategic priorities that may be aligned with policy priorities and broad, national industrial capabilities, while making tactical programme level decisions based on the shifting state of innovation within fields.

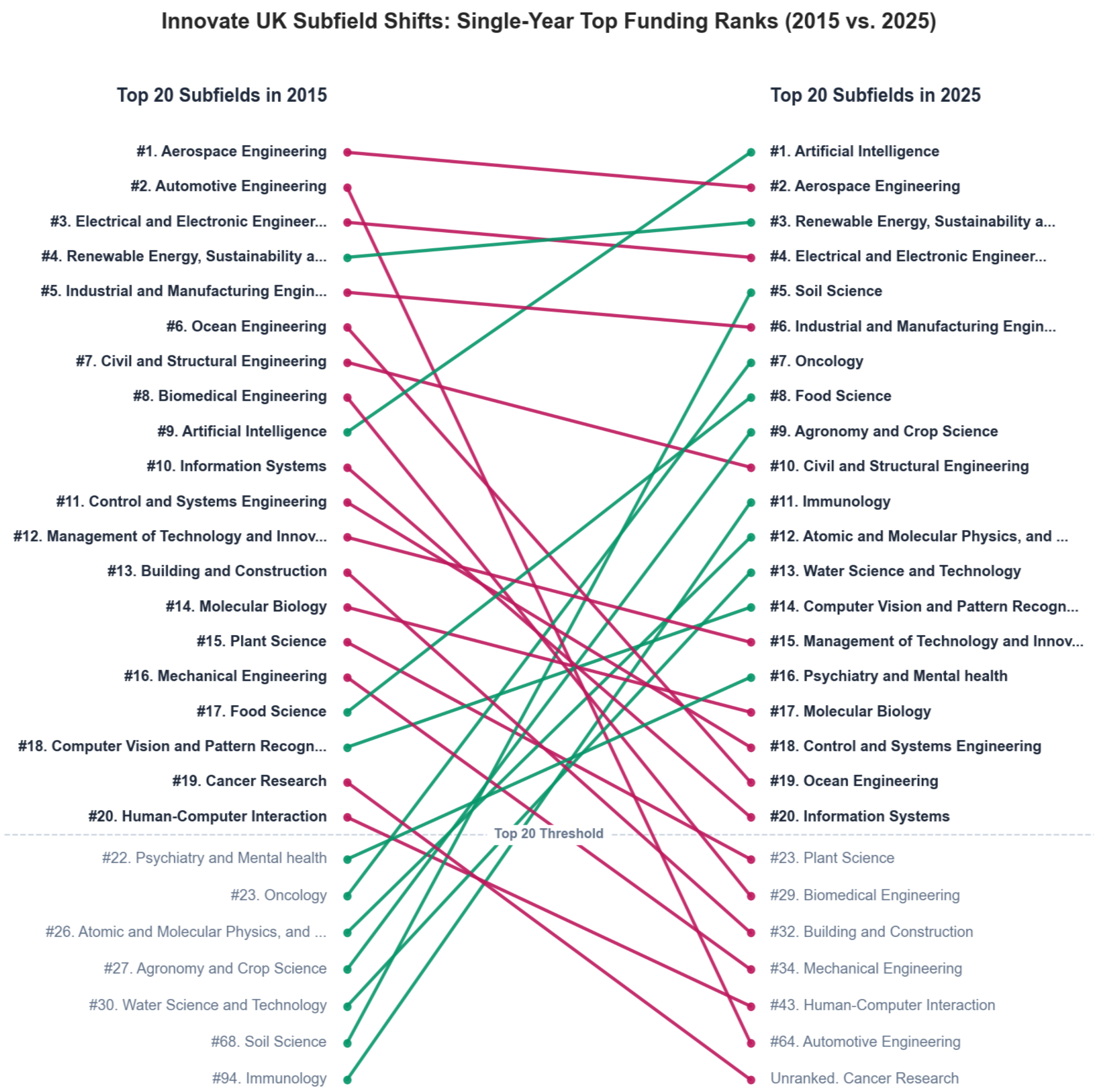

Chart 4.3 - Subfields associated with projects Innovate UK has funded - top 20 today and 10 years ago.

Chart 4.4 - Topics associated with projects Innovate UK has funded - top 20 today and 10 years ago.

By categorising research topics linked to Innovate UK projects into niche, emerging, active, and mature categories (Approach A), we obtain a view of the research base upon which publicly funded UK innovations are built. This analysis shows that the underlying topics associated with Innovate UK projects - the global scientific knowledge base supporting these innovations - were initially concentrated in active and emerging research. Post-2015, however, this pattern shifted markedly: the proportion of emerging topics contracted as scientific output consolidated, with the portfolio increasingly leaning toward mature, high-volume domains alongside an expansion in niche areas.

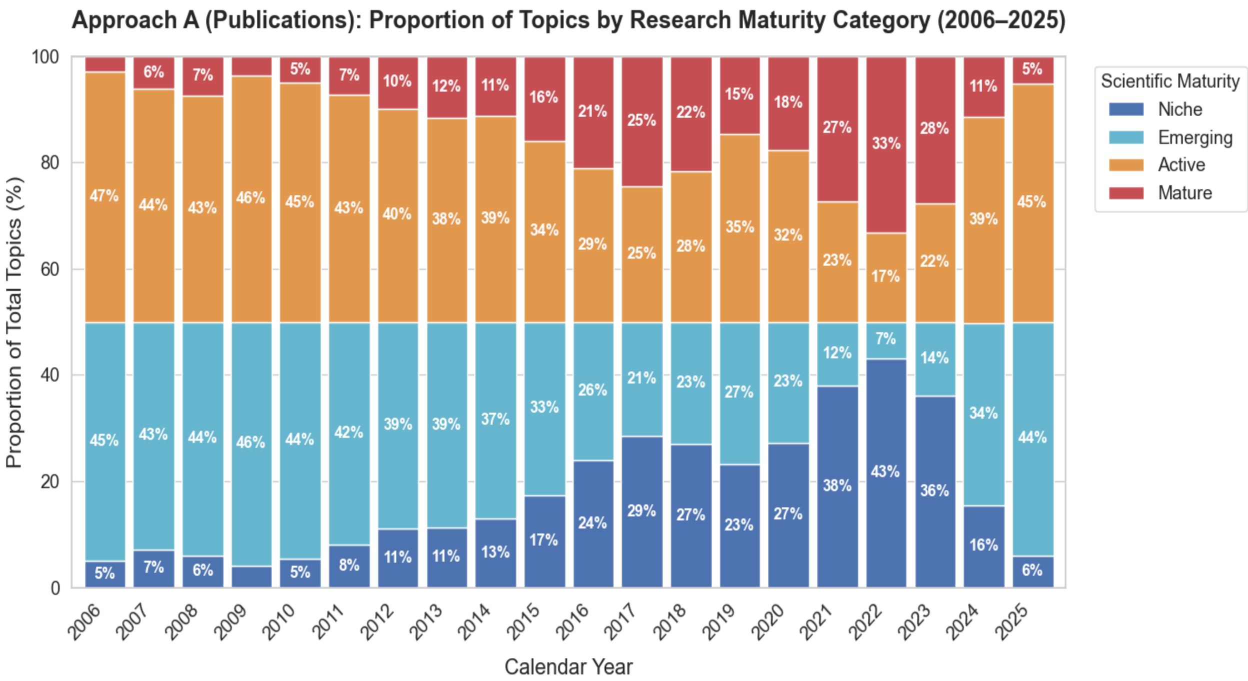

Chart 4.5 - Maturity of overall Innovate UK funding portfolio over time (number of publications).

The static 50% boundary separating the low-volume categories (niche and emerging) from the high-volume categories (active and mature) is an intentional property of the annual volume median, which strictly partitions the topic population into two equal halves in every calendar year. Prior to 2015, the structural symmetry observed between emerging and active reflects scale-independent growth dynamics (Gibrat's Law), wherein small and large research topics possessed *comparable probabilities of attaining above-median growth rates*. This symmetry breaks post-2015: emerging topics contract significantly to a low of 6.8% (with niche topics expanding to 43.2%), while active and mature topics maintain larger overall shares (16.8% and 33.2%, respectively). This divergence indicates a transition from scale-neutral research expansion to *scale-dependent growth momentum*, where publication activity is increasingly consolidated within larger, well-resourced topics, while smaller, low-volume areas struggle to sustain high growth rates.

Linking real project support to underlying research activity (Approach B) shows that investment support varies substantially over time across this research base. Weighting the exposure of

Innovate UK projects to research base maturity by financial expenditure (£) reveals how capital has been historically concentrated. Annual funding allocations exhibit sharper fluctuations between maturity stages than paper counts alone, driven by significant funding into targeted areas. As a result, even in years where the count of funded topics appears evenly spread, financial capital often leans heavily towards high-volume active domains or specific emerging breakthroughs.

At least two distinct strategic eras for Innovate UK are present in this analysis. In the first decade, the agency concentrated capital on innovation projects that were supported by an underlying research base consisting of emerging and active topics. This is consistent with Innovate UK having a primary interest in bridging the translation gap between research and innovation. From 2015/2016 onwards, Innovate UK's portfolio diversified substantially, with innovation projects built on both niche and mature research having a much stronger presence. This produces a more even spread across the portfolio. Both additions are supported by areas of low publication growth, at opposite ends of the volume scale: well-established literatures where the science is largely settled, and small literatures that are not currently expanding. Significantly, neither corresponds to the early-mover position described in Section 1, which in this framework is the emerging category.

Our spatial vector analysis shows that the structural posture of Innovate UK's portfolio varies considerably over time relative to a uniform distribution across research maturity stages. Tracking the Euclidean distance between actual annual allocations and the theoretical baseline—where capital is divided evenly at 25% across niche, emerging, active, and mature domains—highlights key shifts in portfolio balance. In 2017, Innovate UK's investment behaviour reached its closest alignment with this neutral benchmark, reflecting a broad dispersion of capital across all four developmental stages. Since 2017, however, the portfolio has deviated substantially from an even split. This widening distance metric indicates an increasing degree of capital concentration toward specific maturity stages in recent years, moving away from a uniform spread across the scientific lifecycle.

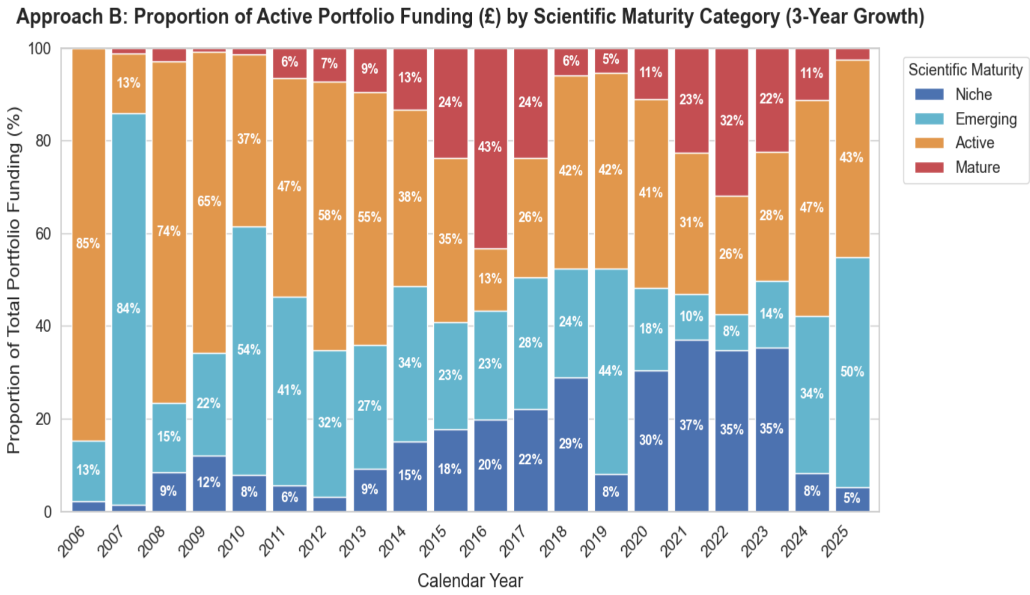

Chart 4.6 - Maturity of overall Innovate UK funding portfolio over time (weighted by distribution of funding).

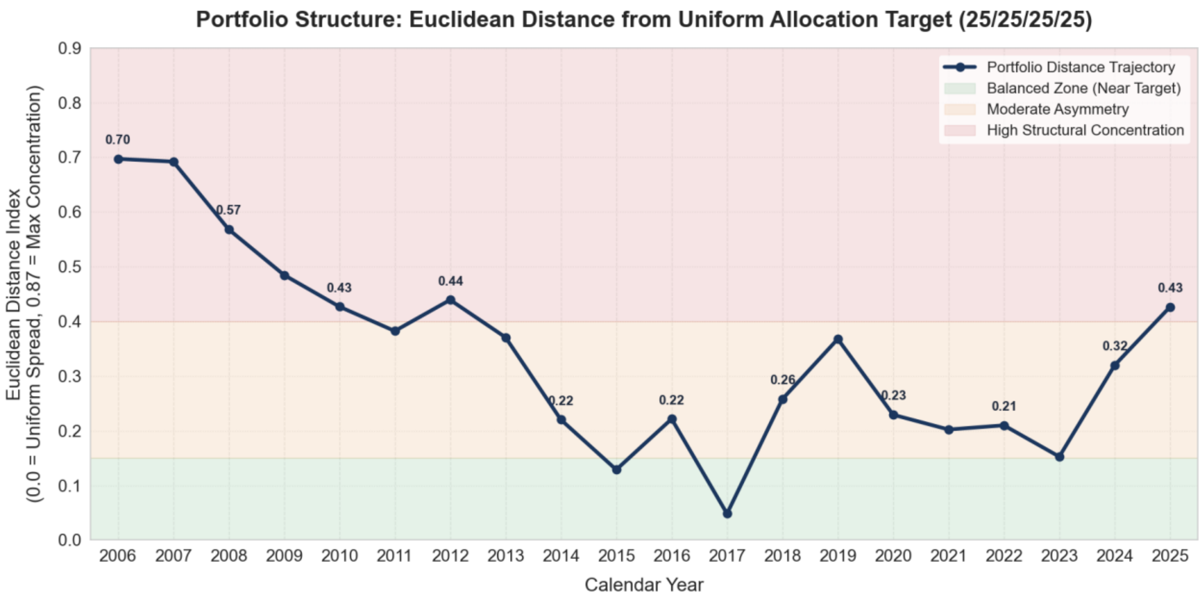

Chart 4.7 - Vector distance from uniform 25% distribution of Innovate UK portfolio, 2006-2025.

The Frascati analysis shows that the scientific evidence base underpinning Innovate UK-supported projects is overwhelmingly anchored in applied research, which consistently accounts for over half of publication volume across all volumetric maturity stages when using the weighted Frascati maturity categories. Applied research expands over time from 55% in 2006 to 59% by 2025 and serves as the single dominant focus for 95.8% of topics (1,799 topics) at the time of peak investment. Basic research provides a persistent foundational baseline—comprising approximately 19–25% of publication share across maturity bins—while general research contracts as topics scale and specialise. Meanwhile, experimental development venues maintain a remarkably stable 12–15% share of total portfolio publications over time, though they never serve as the primary orientation for individual topics. Together, these findings confirm that Innovate UK's capital allocations overwhelmingly concentrate at the exact juncture where theoretical scientific discovery has successfully matured into actionable, applied R&D. For further details of the Frascati analysis, see the Appendix.

4.2 Timing - when does Innovate UK fund research?

When examining individual topic selections, the highest-funded topics receive support across every phase of scientific maturity. Rather than concentrating on a single part of the research landscape, whether fast-growing areas or large established ones, grant allocations in top-funded domains span all four combinations of volume and growth. Major financial commitments are maintained across the full scientific lifecycle, rather than being restricted to a particular phase of underlying research readiness.

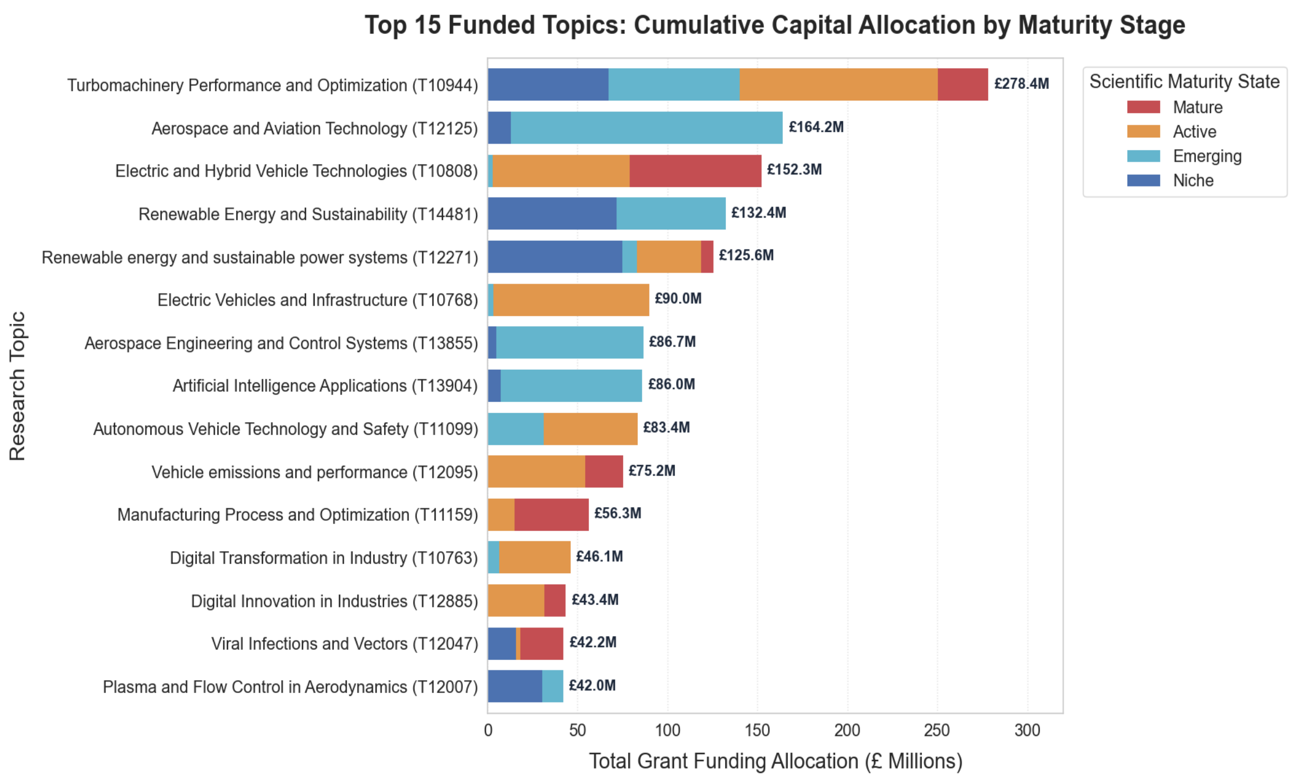

Chart 4.8 - Maturity categories of topics associated with top funded projects.

When categorised into strategic archetypes, Innovate UK's funding distribution primarily follows an *ecosystem co-alignment* pattern, where domestic grant funding expands in parallel with growing global scientific momentum. Across the broader time horizon, the majority of active topics sit within this co-aligned baseline, receiving sustained support as global publication activity scales.

However, between 2019 and 2023, the data reveals a shift in portfolio dynamics. During this five-year period, there is a notable increase in funding directed towards topics that had a small and slow-growing in global research output. Over these five years Innovate UK committed a growing share of capital to areas where the underlying literature was not expanding, alongside its core pattern of funding in step with scientific momentum. Whether this represents anticipation of future growth or exposure to areas that have stalled cannot be determined from the funding data alone. This structural movement highlights a period of temporary portfolio rebalancing, with increased exposure to nascent, early-stage research domains alongside the core co-alignment trajectory.

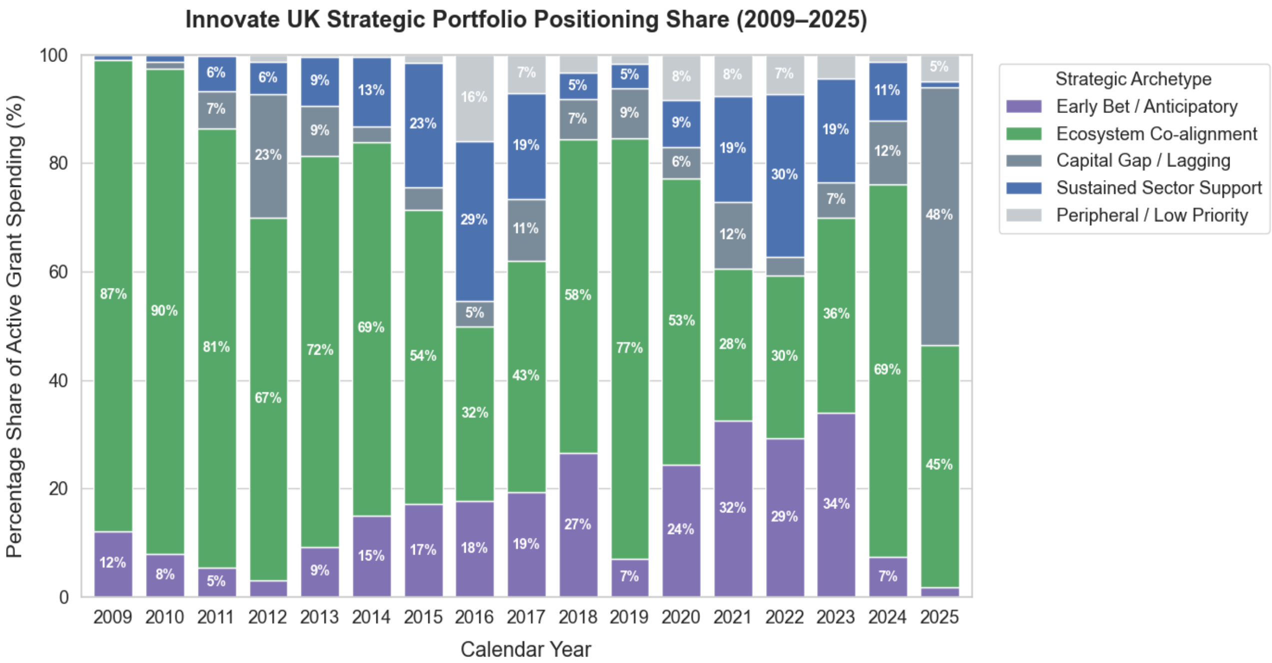

Chart 4.9 - Innovate UK strategic profiles, 2009-2025.

Innovate UK supports topics over multi-year horizons, but the timing of its maximum capital commitment varies relative to the underlying scientific lifecycle. When evaluating the research maturity of individual topics at their exact year of peak investment—the single calendar year receiving the highest volume of grant spend—capital concentration is spread across diverse developmental stages rather than clustering around a single phase.

While a substantial proportion of topics reach peak public funding while operating in mature (20.2%) or active (31.2%) high-volume stages, significant funding peaks also occur when topics are emerging (23.4%) or niche (25.1%). The first shows that innovation projects are not restricted to building on established, high-output domains: almost a quarter coincide with literatures growing rapidly from a small base. The second is a different observation. A comparable share of peak commitments falls on topics that are small and not growing, where the supporting evidence base was neither substantial nor expanding at the point of maximum investment. This demonstrates that funding within the Innovate UK portfolio is not restricted to established domains, but also coincide with earlier stages of potential scientific evolution.

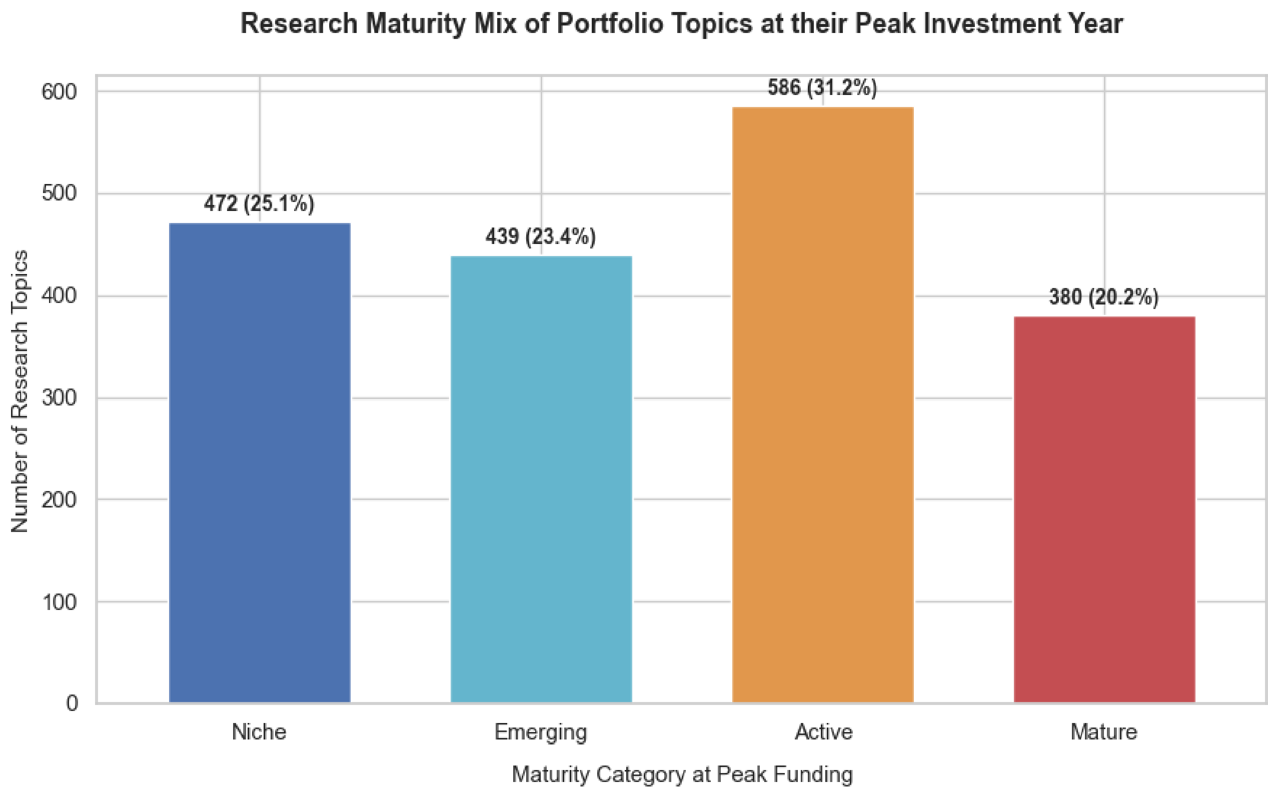

Chart 4.10 - Maturity category at point of peak Innovate UK investment in projects related to topic.

5. Research, Innovation and Frontier Technologies

Innovate UK announced a new strategy to support "deep tech" innovations across six frontier technology areas, in line with the UK government's Digital and Technologies Sector Plan. 137 topics (out of 4,516) were linked to frontier technology areas (for the full list of associated topics, see the Appendix). Many of these topics have already been supported with Innovate UK funding. However, a high percentage remain unsupported, pointing to the fact that Innovate UK does not currently support innovations that exploit the entire research base in these areas.

Table 5.1 - Innovate UK coverage across frontier technology areas

Frontier technology area	Total topics	Topics in Innovate UK projects
Artificial Intelligence	78	56 (71.8%)
Semiconductors	20	16 (80.0%)
Quantum	14	6 (42.9%)
Advanced Connectivity	13	11 (84.6%)
Cybersecurity	12	6 (50.0%)
Engineering Biology	9	7 (77.8%)

Innovate UK has already distributed substantial amounts of funding to innovations within the frontier technology areas, although this is not evenly distributed. Quantum technologies have seen the smallest injection of funding, with £2 million spent since 2006. This contrasts sharply with artificial intelligence projects, which have received over £1.5 billion, more than all other frontier technology areas combined.

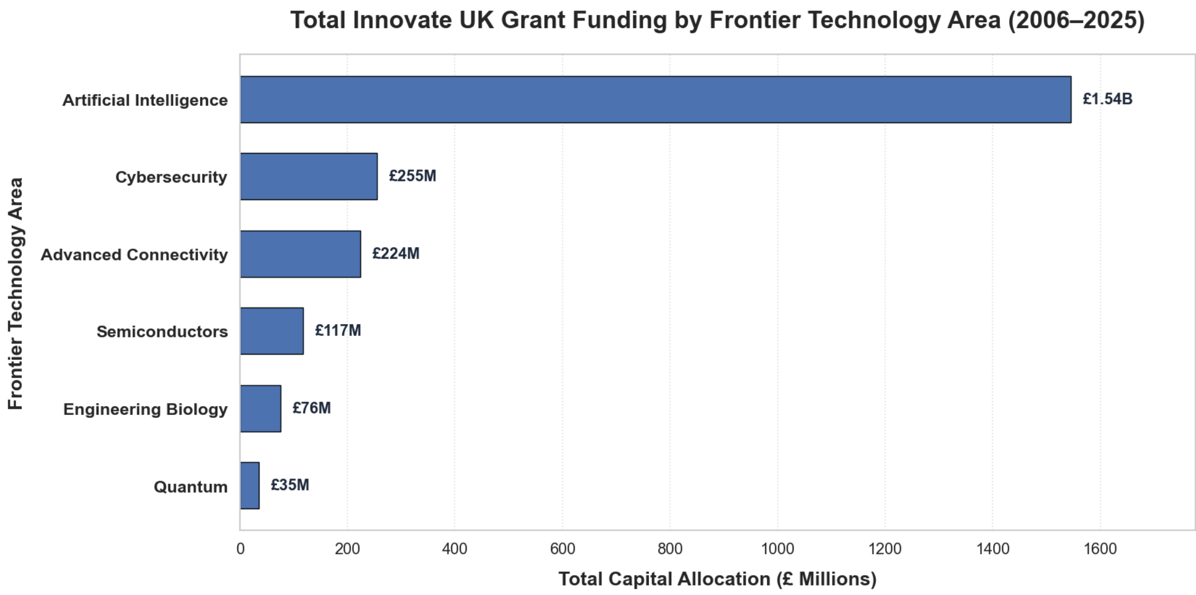

Chart 5.1 - Innovate UK total investment (2006-2025) in projects associated with frontier technology areas.

The topics within the Innovate UK portfolio of frontier technologies present a unique and shifting maturity pattern. Advanced connectivity and artificial intelligence shift over time from being composed of niche and emerging topics to more active and mature topics. This is a product of high historic and ongoing growth in these fields. Cybersecurity has a similar but exaggerated profile, with mature research topics being dominant for the last decade. Semiconductor research provides a contrasting profile, having consisted mostly of mature and active topics in the past but now being dominated by niche and emerging topics. The emerging share points to a new wave of next-generation research relevant to frontier innovation; the niche share may reflect the same shift at an earlier point, or fragmentation of the field into smaller specialisms that are not individually expanding.

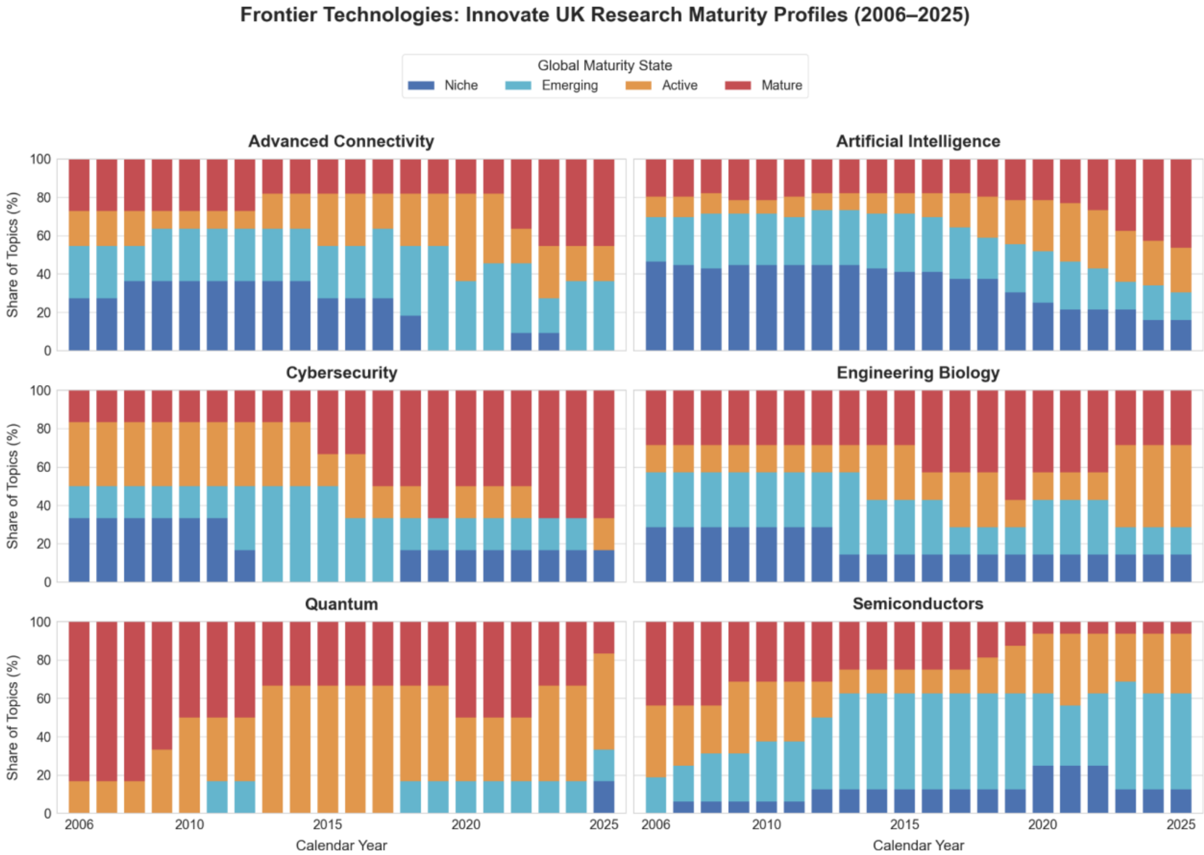

Chart 5.2 - Maturity of Innovate UK-supported topics underlying the frontier technology areas, 2006-2025.

The maturity distribution of research in recent years provides a lens through which Innovate UK is able to understand the nature of the areas it is now tasked with building upon. However, it is also important to understand how their support for innovation in these areas has aligned with funding distribution. Alignment would suggest that the organisation is already supporting projects and using mechanisms that are suitable for supporting innovations that draw on the range of knowledge academic research has to offer. Misalignment could be the result of explicit targeting of areas that suit particular kinds of innovation or an inadvertent “tunnel vision”. In either case, this view offers Innovate UK an opportunity to adjust its strategy to suit the role that it is attempting to play in supporting deep tech to grow and scale in the UK.

Artificial intelligence and cybersecurity are the two technology areas characterised by the most notable misalignments. In the case of artificial intelligence, recent shifts in the research base have created larger mature and niche areas segments, but still with strong emerging and active components. Since 2020, 90% of Innovate UK’s project funding in the field has been for projects underpinned by low growth niche and mature topics. The result is a portfolio concentrated almost entirely on research that is not expanding, at both ends of the volume scale, while the fastest-growing areas of the AI literature attract almost no funding. In the case of cybersecurity, Innovate UK has had an almost singular focus on innovations supported by mature research topics, eschewing projects that are premised on the niche and emerging knowledge that also exists.

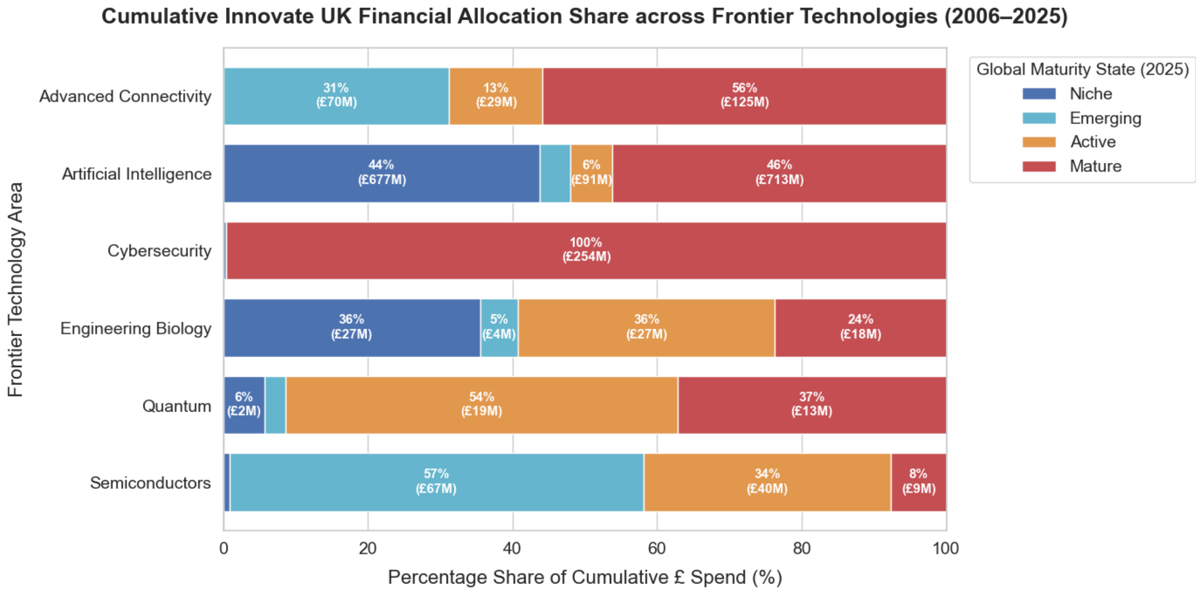

Chart 5.3 - Cumulative Innovate UK investment into projects associated with frontier technology areas, 2020-2025.

Across the frontier technology funding portfolio, we can examine Innovate UK's tendency towards the strategy archetypes. We find that in recent years, (excluding 2025), the organisation's funding has been largely dominated by the Ecosystem co-alignment category. This contrasts with the more blended strategy that we observe across Innovate UK's entire portfolio in the same period. We also see that prior to 2023 this was complemented by a significant minority of funding to topics without corresponding scientific momentum, which has since diminished. If Innovate UK's objective is to support breakthrough discoveries, the more direct question is its exposure to the emerging category, where literatures are expanding rapidly from a small base.

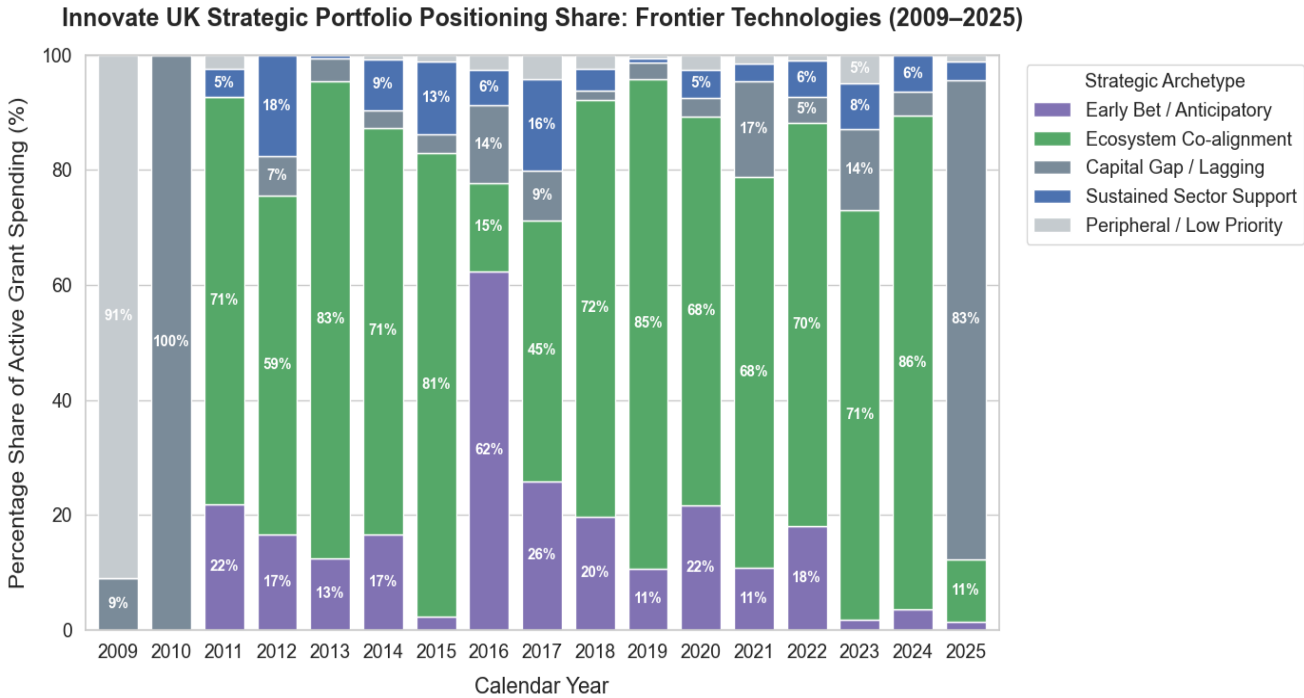

Chart 5.4 - Innovate UK strategic profiles for frontier technology areas, 2009-2025.

6. GtR+: An Experiment

The original Gateway to Research (GtR) dataset represents a valuable public resource, offering clear visibility into project details, partner organisations, and financial award allocations. The primary value added by the enhanced GtR+ dataset lies in its supplementary metadata subject extract, which connects award records with taxonomy-based subject classifications. In principle, this added classification layer enables researchers to bridge public innovation grants directly with global scientific outputs using standardised bibliometric taxonomies.

In practice, however, using the GtR+ topic extract revealed structural limitations. The dataset provides OpenAlex subfields rather than granular research topics, and the classification is applied inconsistently, occasionally defaulting to broad disciplinary fields or drawing upon non-standard schemes outside the OpenAlex framework. Relying on raw GtR+ labels without modification therefore risks distorting portfolio analysis. To maintain analytical rigour, research relying on this extract requires consistent project re-classification using a unified taxonomy—a requirement that led us to construct our two-stage LLM routing pipeline.

Beyond taxonomy inconsistencies, the re-classification process was further constrained by data quality issues within the award descriptions themselves. A significant proportion of GtR+ records lacked valid project abstracts entirely or contained truncated, non-technical summaries, forcing their upfront exclusion prior to the LLM classification pipeline. This highlights a fundamental dependency of post-hoc bibliometric enrichment: the reliability of automated taxonomy matching remains strictly bounded by the completeness and descriptive fidelity of primary grant records.

To improve the utility of GtR for future policy evaluation, metadata accuracy and granularity should be strengthened at the point of application. Rather than relying on post-hoc re-classification, funding workflows could integrate standardised taxonomy choices or semi-structured topic tagging directly into grant submissions. Applicants and peer reviewers could, for instance, select a primary subfield and subsequently choose from the approximately twenty specific topics nested within that subfield. This could be done manually or facilitated with AI assistance. This might also be an opportunity for Innovate UK to link projects with strategic outcomes from the beginning of the support process. Capturing precise metadata and strategic alignment at submission would eliminate taxonomy ambiguity and allow policymakers to track public funding flows across frontier technologies in real time.

A more fundamental constraint of GtR+ in its present form is the systemic absence of post-project outcome metrics for Innovate UK awards. While GtR+ tracks longitudinal outputs for Research Council grants, equivalent tracking is not currently published or linked for business-led innovation funding. Consequently, a wide spectrum of downstream technical and commercial indicators—including intellectual property, spinouts, software and technical products, research tools, clinical trials, databases, key findings, publications, and creative outputs—remains unrecorded for Innovate UK projects. This reporting gap also extends to broader impact indicators such as follow-on funding, policy influence, engagement activities, collaborative networks, and narrative impact summaries.

While the absence of these outcome metrics in GtR+ did not preclude evaluating commercial outcomes altogether, the utility of GtR+ within our study was constrained to evaluating funding inputs and baseline scientific maturity, rather than downstream performance. Integrating native outcome tracking for Innovate UK projects directly within GtR would make measuring commercial success substantially easier and more streamlined. Should UKRI expand GtR to incorporate comprehensive outcome tracking for business-led grants, this would unlock significantly richer policy insights. GtR+ attempts to remedy this by providing information from Companies House, which supports downstream data linking, but remains limited for signals of commercial success. Decision-makers would then be able to evaluate not only when and where public capital is deployed across the research lifecycle, but how those investment timing strategies directly influence market commercialisation, firm growth, and economic return. This is critical for a science and innovation strategy that is more tightly linked to wider industrial policy.

7. Conclusion

7.1 Discussion

We began by considering a strategic dilemma faced by innovation agencies: either to support innovations built on knowledge that is still accumulating, or to wait until the research around an area of innovation is more established. In practice the measure separates two questions that this dichotomy combines: how much research exists, and whether it is still growing. When viewed through this simplified framework, Innovate UK has evolved beyond picking a single strategy. Our analysis reveals that the portfolio of projects funded by Innovate UK over the last decade maps on to research topics that are diversified in terms of their maturity, breaking the binary choice.

Indeed, reality is more complex than this single dichotomy and the picture that emerges from our analysis is nuanced to match. One key observation is that while Innovate UK does not have an obvious tendency towards areas of innovation built on research at a specific level of maturity, the distribution of underlying research maturity has varied over time, defining at least two eras. The first, lasting from 2006 to 2014, is a period in which high growth emerging and active areas of research provided the underlying knowledge base surrounding Innovate UK funding. The second, lasting from 2015 to the present day, still draws strongly on high growth knowledge areas, but also extends towards innovation activities that have a basis in niche and mature research. This indicates that the portfolio has come to span the full range of research volume and growth, rather than concentrating in one part of it. Whether this reflects a deliberate widening of delivery mechanisms cannot be established from the funding data alone.

At points in this era, niche and mature topics, the two low-growth categories, have accounted for a substantially larger share of Innovate UK funding, peaking at around two thirds in 2022. The most recent years however have seen something of a return to reliance on active and mature areas. This is worth noting, because Innovate UK is now tasked with advancing innovations across six frontier technology areas, each of which has a distinct maturity distribution profile in the research base.

The method we have presented offers a new tool that Innovate UK, and other innovation agencies, can use to measure alignment between their stated ambitions, their investments, and the changing nature of scientific inputs to innovation. Innovate UK's recent commitment to support deep tech built on scientific breakthroughs in the six frontier technology areas is a fruitful context for this analysis. Artificial intelligence is one of the areas of focus, and has benefitted from large volumes of Innovate UK funding. However, innovations underpinned by emerging research, defined by rapid growth from a small base, and the closest available proxy for a knowledge base in active formation, are almost entirely absent from the recent AI portfolio, while those resting on small, slow-growing niche topics have received substantial funding. If the objective is to support innovations built on scientific breakthroughs, this is the segment of the knowledge base that appears least exploited.

We consider research as an important input to invention, but only one of several factors that Innovate UK will consider when forming its strategies and delivery plans, and making programme level decisions. Across strategy documents and delivery plans from the last two decades, Innovate UK itself rarely references the relationship between the innovations it seeks to support and the nature of the research base that might underpin them. Our study therefore does not necessarily measure a gap between intent and achievement, but suggests that as a measurable concept, research maturity can inform future decision making.

7.2 Limitations

Our study had several limitations, both in its theoretical framing and practical implementation. We linked the projects supported by Innovate UK investment to the underlying research base, mapping a theoretical pathway from the global scientific evidence base, to applied and commercial applications, Innovate UK selection and support, scale-up and ultimately public economic benefit. To empirically validate this translation pathway and measure downstream commercial impact, further data would be required to establish links between these stages.

We do not claim that Innovate UK explicitly refers to the scientific evidence base in their funding decisions. However, our theoretical framework does imply that project application reviewers and other funding stakeholders implicitly consider the potential of individual projects in relation to this evidence base. Consequently, the observed portfolio patterns reflect implicit alignment with scientific momentum rather than a deliberate, pre-planned target allocation based on bibliometric indicators. Further work would be required to measure the true attitudes and expectations of those reviewing applications and funding innovation projects.

Innovate UK offers different funding instruments (e.g. grants, loans) for innovation projects associated with teams or companies across different stages of commercialisation potential and business readiness, from start-ups to scaling of established businesses. Due to data limitations, the analysis presented above does not distinguish between these types of funding, treating all support equally. This limits our ability to observe whether specific funding mechanisms (such as early-stage feasibility grants versus scale-up loans) systematically target different phases of underlying research readiness.

When implementing the topic classification pipeline, we considered only a subset of the Innovate UK-supported projects due to data quality issues (e.g. missing or substantially truncated abstracts). Because we prioritised accuracy, the classifier only assigned topics to 70% of these projects. The consequence for our analysis is that 17,659 projects, representing 64.2% of all Innovate UK funding, could not be classified and are not included in the analysis, limiting the validity of our findings.

More generally, our use of automated tools raises concerns about the extent to which LLMs are able to approximate experts. In classification tasks, multiple attempts were often necessary to achieve plausible results, and these tools lack deep subject knowledge and the strategic insights of a human evaluator.

We focused on OpenAlex topics, in order to map innovation projects to the research base. While much more granular than the higher taxonomy levels, topics are still imprecise. For instance, “Artificial Intelligence in Healthcare and Education” includes research relevant for both medical and teaching professionals who will likely focus on different areas of inquiry, research goals and benchmarks, and could plausibly have divergent maturity trajectories.

Finally, the maturity measures we used are based only on the topics related to Innovate UK projects, not on all topics on OpenAlex. This means that when we categorise a topic as “mature” this means it is high volume, low growth *relative* to the other topics associated with Innovate UK projects. This approach is internally consistent and valuable for understanding Innovate UK investment choices, but it does not identify potential new investment areas or relate public investment to a global maturity benchmark.

7.3 Future directions

This work has applied methodologies for measuring the maturity of research areas and provided a descriptive analysis of the relationship between those areas and innovation funding. The work has been motivated by the premise of research as an important input to invention. While limited and exploratory, this project opens up new policy-relevant research opportunities.

An original aspiration for this project was to analyse the relationship between funding timing, research maturity, and commercial success. This could be done by linking companies funded by Innovate UK with their entries in comprehensive business datasets (ORBIS, Crunchbase) and measuring outcomes such as company survival rates and follow-on investment or funding. A carefully designed regression analysis, which also accounted for company characteristics, would reveal whether the underlying research maturity for a particular innovation is predictive of company success. This would validate the utility of the methodology presented here as a useful tool for strategic decision making within innovation agencies.

We might also consider a stock perspective in addition to a flow perspective. The research considers the maturity of only new annual investments, but the overall portfolio of active projects is also strategically important. Because projects run over several years, this view would give visibility into the mix of all investments at various points in time (new projects and ongoing projects). Future work could also consider an expanded range of maturity categories, making finer-grained distinctions within the niche, emerging, active and mature profiles.

As described above, our analytical approach was internally consistent, but lacks a specific global baseline against which to compare the maturity profiles and dynamics. Future research could identify which comparator(s) are appropriate for determining research maturity, either by taking a global stance or sourcing input from Innovate UK about their most relevant competitors.

A closely-related extension would consider the domestic and global research base distinctly. The UK remains a global leader in the production of scientific research, delivering outputs that far exceed its demographic size. Despite representing approximately 0.9% of the global population and 4.1% of the world's active researchers (DSIT, 2025), the UK led on 4-6% of all global publications between 2006 and 2025. In this work, we have considered the maturity of knowledge arising from the global research system. This knowledge is formed from literature which might diffuse through various mechanisms to innovators in the UK and presents the frontier of theoretical research and innovation possibilities. However, the domestic research base is potentially more important to local innovators as it is also a signal for talent, communities know-how, facilities and infrastructure that could more directly contribute to new innovations that build on research frontiers.

Future research could also include qualitative work to understand how strategic shifts were implemented in practice, and to determine whether and to what extent Innovate UK reviewers are aware of and motivated by publication activity in the underlying academic literature (scientific base). Human experts could also be enlisted to validate automated classifications, such as which topics are associated with the 6 frontier technologies.

A potential outcome from future research might be practical tool(s) to identify promising gaps in public funding support (e.g. a topic within a frontier technology area in which most papers are published in applied/experimental journals). This would incorporate other relevant indicators for strategic and programmatic decision making. This, alongside extensions in the analysis, could also be extended to the activities supported by research funders (such as those within UKRI), who may also have an interest in the distribution of the maturity of the fields they support.

REFERENCES

Ahmadpoor, Mohammad, and Benjamin F. Jones. 2017. 'The Dual Frontier: Patented Inventions and Prior Scientific Advance'. *Science* 357 (6351): 583–87. https://doi.org/10.1126/science.aam9527.

Arrow, Kenneth. 1962. 'Economic Welfare and the Allocation of Resources for Invention'. *NBER Chapters*, 609–26.

Bornmann, Lutz, Robin Haunschild, and Rüdiger Mutz. 2021. 'Growth Rates of Modern Science: A Latent Piecewise Growth Curve Approach to Model Publication Numbers from Established and New Literature Databases'. *Humanities and Social Sciences Communications* 8 (1): 224. https://doi.org/10.1057/s41599-021-00903-w.

Department for Business and Trade and Department for Science Innovation and Technology. 2025. *Digital and Technologies Sector Plan*. Policy paper. UK Government. https://www.gov.uk/government/publications/digital-and-technologies-sector-plan.

Fleming, Lee. 2001. 'Recombinant Uncertainty in Technological Search'. *Management Science* 47 (1): 117–32. https://doi.org/10.1287/mnsc.47.1.117.10671.

Innovate UK. 2017. *Shaping the Future: Innovate UK Delivery Plan 2017–2018*. Delivery plan. Innovate UK. https://www.ukri.org/wp-content/uploads/2017/12/IUK-101221-IUKDeliveryPlan2017To2018.pdf.

Innovate UK. 2021. *Building the Future Economy: Plan for Action for UK Business Innovation 2021–2025*. Strategy. UK Research and Innovation. https://www.ukri.org/wp-content/uploads/2021/11/IUK-18112021-Plan-For-Action-for-UK-Business-Innovation_FULL_WEB-FINAL-26.10.21-1.pdf.

Innovate UK. 2026. *Innovate UK: Turning Breakthrough Ideas into Industry Giants*. Prospectus. UK Research and Innovation. https://www.ukri.org/wp-content/uploads/2026/03/Innovate-UK_Prospectus_v14_DIGITAL_Accessible_Spreads.pdf.

Jan van Eck, Nees, and Ludo Waltman (Leiden Madtrics). 2024. 'An Open Approach for Classifying Research Publications'. January 24. https://doi.org/10.59350/qc0px-76778.

Kanders, Karlis, and Laurie Smith. 2021. 'Innovation Sweet Spots: Which Green Technologies Are at a Tipping Point?' Nesta, Nesta. https://www.nesta.org.uk/feature/innovation-sweet-spots-green-tech/.

Lieberman, Marvin B., and David B. Montgomery. 1988. 'First-Mover Advantages'. *Strategic Management Journal* 9: 41–58.

Narin, Francis, Kimberly S. Hamilton, and Dominic Olivastro. 1997. 'The Increasing Linkage between U.S. Technology and Public Science'. *Research Policy* 26 (3): 317–30. https://doi.org/10.1016/S0048-7333(97)00013-9.

Nelson, Richard R. 1959. 'The Simple Economics of Basic Scientific Research'. *Journal of Political Economy* 67 (3): 297–306.

Price, Derek John de Solla. 1963. *Little Science, Big Science*.

Technology Strategy Board. 2008. *Connect and Catalyse: A Strategy for Business Innovation 2008–2011*. Strategy. Technology Strategy Board. https://webarchive.nationalarchives.gov.uk/ukgwa/20080528124633id_/http://www.innovateuk.org/_assets/pdf/corporate-publications/technology%20strategy%20board%20-%20connect%20and%20catalyse.pdf.

UK Research and Innovation*. 2026. 'Innovate UK funded projects between 2004 and financial year 2015 to 2016'. https://www.ukri.org/publications/innovate-uk-funded-projects-since-2004/.*

UK Research and Innovation*. 2026. 'Innovate UK funded projects from financial year 2016 to 2017 to present'. https://www.ukri.org/publications/innovate-uk-funded-projects-since-2004/.*

Weitzman, Martin L. 1998. 'Recombinant Growth'. *The Quarterly Journal of Economics* 113 (2): 331–60. https://doi.org/10.1162/003355398555595.

APPENDIX

A1. LLM implementation

A1.1 Topic classifier implementation

Overview

The topic classifier follows a two-stage procedure:

- Stage 1: subfield identification
- Stage 2: topic assignment

Both stages of the classifier operated under strict prompt engineering rules and enforced structured JSON formatting to maintain consistency:

Stage 1: Subfield routing

- Role: Technical research router.
- Objective: Map the project abstract to a single OpenAlex subfield based on primary methodology rather than broad thematic keywords.
- Execution Rules: The model evaluated specific technical terms in the abstract, returning the exact subfield string along with a confidence rating (*High*, *Medium*, *Low*), routing evidence, and a single-sentence justification.

Stage 2: Granular topic classification

- Role: Expert research taxonomist.
- Objective: Match the project abstract to the most granular technical topic within the Stage 1 subfield.
- Execution rules:
    - Technical alignment: Evaluated candidate topic summaries and keywords derived from CWTS/Leiden data.
    - Direct evidence: Mandated an exact verbatim quote (5–15 words) extracted directly from the project abstract to justify the classification.
    - The “NONE” fallback: Explicitly assigned selected_topic_id: "NONE" if the abstract lacked sufficient detail or if its core technology was absent from the topic list.
    - Standardised Justification: Categorised confidence reasons into one of five technical definitions (*Strong Single Match*, *Primary Match*, *Ambiguous Match*, *Marginal Match*, or *No Relevant Match*) to facilitate downstream quality filtering.

Prompts

Prompt 1:
You are a technical research router. Your task is to assign a research abstract to the most appropriate OpenAlex Subfield.

DECISION RULES:
- TECHNICAL HIERARCHY: Prioritize the Subfield that would likely contain the most granular technical topics related to the project's primary methodology.
- EVIDENCE: Identify a specific technical term in the abstract that justifies this subfield choice.
- SPECIFICITY: Select the subfield whose research methods or technologies most closely match the abstract. Do not rely on broad thematic similarity.
- SUBFIELD: Return the subfield exactly as written in the list.

RESPONSE FORMAT:
Respond ONLY with a JSON object.

```
{
  "matched_subfield": "Exact string from list",
  "routing_evidence": "phrase from abstract",
  "routing_confidence": "High | Medium | Low",
  "routing_reasoning": "1-sentence justification."
}
```

Prompt 2:

You are an expert research taxonomist. Your task is to classify a UK research project abstract into a granular technical Topic based on the provided taxonomy.

INPUT DATA:
1. TARGET ABSTRACT: {abstract_text}
2. SUBFIELD CONTEXT: {subfield_name}
3. POTENTIAL TOPICS (within {subfield_name}):
{topic_list_with_summaries}

TASK:
Analyze the Abstract and select the most appropriate Topic ID from the provided list.

DECISION RULES:
- TECHNICAL ALIGNMENT: Prioritize specific methods, materials, equipment, or technologies mentioned in the CWTS Summaries and Keywords.
- EVIDENCE: You must provide a 'match_evidence' field containing an EXACT phrase (5-15 words) copied directly from the Abstract that justifies this classification.
- THE "NONE" OPTION: If the abstract is too vague, or if its primary technical focus is not represented in the potential topics list, you MUST set selected_topic_id to "NONE".
- GRANULARITY: Choose the most specific topic possible. Do not choose a topic based on broad sector keywords (e.g., "Sustainability") if a more specific technical match exists.

CONFIDENCE CATEGORIZATION:

For the confidence_reason field, you must select exactly ONE of the following technical justifications:
- Strong Single Match: Direct and dominant overlap with one topic's core methodology.
- Primary Match: Overlaps with several topics, but one is the clear primary focus.
- Ambiguous Match: Overlaps with multiple topics with no clear technical hierarchy.
- Marginal Match: Only weak or tangential technical overlap with one topic.
- No Relevant Match: No technical relationship found between abstract and topics.

RESPONSE FORMAT:
Respond ONLY with a JSON object. No prose or explanations outside the JSON.
{
"selected_topic_id": "string or NONE",
"selected_topic_name": "string",
"match_evidence": "exact phrase from abstract",
"confidence_score": "High | Medium | Low",
"confidence_reason": "select from predefined list above",
"reasoning": "A 1-sentence technical justification."
}

A1.2 Frascati classifier implementation

Overview

We deployed OpenAI's gpt-4o-mini model as an expert research taxonomist trained on Frascati standards. For each journal, the model evaluated three metadata inputs: the journal name, linked scientific concepts, and the official journal description. The classifier's decision logic evaluated linked scientific concepts and editorial descriptions to establish R&D intent. To maintain high auditability, the model was required to extract a 5–15 word verbatim quote from the journal description as direct textual evidence for its primary assignment, while simultaneously outputting a confidence score paired with a standardised rationale reflecting the clarity and specificity of the journal's underlying metadata.

The decision logic for the Frascati classifier followed strict evaluation rules:

- Weighting signals: Scientific concepts provided domain signals regarding typical maturity (for example, "Quantum Mechanics" signaling basic science versus "Civil Engineering" signaling applied research), while descriptions established editorial intent.
- Intent alignment: The model prioritised stated mission goals, distinguishing between a focus on principles discovery (Basic), practical problem-solving (Applied), or implementation and prototyping (Development).
- Textual evidence: The model was required to extract an exact 5–15 word phrase verbatim from the journal description to justify its primary classification.

- Confidence rating: Each classification was assigned a confidence score (High, Medium, Low) paired with a standardised rationale (Pure Focus, Clear Primary, Hybrid/Balanced, or Vague/General).

All outputs were validated and returned in structured JSON format, enabling seamless aggregation across topics to compute the overall Frascati maturity profile of Innovate UK's funding portfolio.

Prompt

You are an expert research taxonomist specialized in the Frascati Manual 2015 standards for R&D classification. Your task is to classify a scientific journal based on its metadata into the appropriate categories of research and development.

FRASCATI DEFINITIONS:
The term R&D covers three types of activity:
1. BASIC RESEARCH: Experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.
2. APPLIED RESEARCH: Original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific, practical aim or objective.
3. EXPERIMENTAL DEVELOPMENT: Systematic work, drawing on knowledge gained from research and practical experience and producing additional knowledge, which is directed to producing new products or processes or to improving existing products or processes.

INPUT DATA:
1. TARGET JOURNAL: {journal_name}
2. SCIENTIFIC CONCEPTS: {concepts}
3. JOURNAL DESCRIPTION: {description}

TASK:
Analyze the provided metadata and determine the weight of the journal's research focus. You must identify a Primary classification and, if the journal is interdisciplinary or "translational," identify Secondary and Tertiary classifications.

DECISION RULES:
- WEIGHTING SIGNAL: Use the **SCIENTIFIC CONCEPTS** to identify the disciplinary domain and its typical R&D maturity (e.g., "Quantum Mechanics" signals Basic; "Civil Engineering" signals Applied). Use the **JOURNAL DESCRIPTION** to identify the specific *intent* of the work (e.g., "fundamental understanding" vs "clinical application").
- INTENT ALIGNMENT: Prioritize the *stated goal* found in the description. Does it emphasize "discovery of principles" (Basic), "solving specific problems" (Applied), or "prototyping/implementation" (Development)?
- MULTI-LEVEL DETERMINATION:

  * PRIMARY: The core operational space representing >50% of the journal's target output or its explicitly stated main mission.
  * SECONDARY: The primary adjacent step in the research translation pipeline or a clear, secondary editorial focus area.
  * TERTIARY: The minor or long-tail footprint of the journal (e.g., a basic science journal that occasionally accepts industrial prototypes or field case studies).
- THE GENERALIST JOURNAL OPTION: If a journal is broad and explicitly balances all three dimensions—covering basic, applied, and experimental development with no single dominant focus—set the `primary_class` to "General".
- MULTI-LEVEL FALLBACKS: If a journal spans multiple types, list them in order of dominance. If only one or two types are present, set the remaining class levels to "NONE".
- EVIDENCE: You must provide a 'match_evidence' field containing an EXACT phrase (5-15 words) copied directly from the **Journal Description** that justifies the Primary classification. If primary_class is "General" or "NONE" and no exact phrase fits, extract the closest structural phrase or mission summary.
- THE "NON-R&D" OPTION: If the metadata refers to professional practice, education, or industry news without an active research or development component, set the primary_class to "NONE".

CONFIDENCE CATEGORIZATION:
For the confidence_reason field, select exactly ONE of the following justifications:
- Pure Focus: The metadata exclusively mentions one type of R&D (e.g., purely theoretical).
- Clear Primary: Multiple types are mentioned, but one is the undeniable core mission.
- Hybrid/Balanced: The journal explicitly positions itself as "Basic to Applied" with equal weight.
- Vague/General: The metadata uses broad language that makes differentiation difficult.

RESPONSE FORMAT:
Respond ONLY with a JSON object. Ensure all classification fields conform strictly to the allowed enum values.

```
{{
  "primary_class": "Basic Research | Applied Research | Experimental Development | General | NONE",
  "secondary_class": "Basic Research | Applied Research | Experimental Development | General | NONE",
  "tertiary_class": "Basic Research | Applied Research | Experimental Development | General | NONE",
  "match_evidence": "exact phrase from description",
  "confidence_score": "High | Medium | Low",
  "confidence_reason": "select from predefined list above",
  "reasoning": "A 1-sentence technical justification mapping the scientific concepts and description to the Frascati hierarchy."
}}
```

A1.3 Frontier technologies implementation

Overview

To determine the alignment of Innovate UK's project portfolio with strategic national priorities, we systematically evaluated every topic in the OpenAlex/CWTS taxonomy (~4,500 topics) against Innovate UK's six defined frontier technology areas:

1. Artificial Intelligence
2. Advanced Connectivity
3. Cybersecurity
4. Engineering Biology
5. Semiconductors
6. Quantum Technologies

To ensure classification rigor, eliminate misclassifications, and accommodate domain-specific nuances, the classification was executed through a two-pass automated pipeline using Large Language Models (LLMs) and programmatic Python auditing:

- Stage 1: Taxonomy hierarchy evaluation: The classifier evaluated every topic within the context of its parent hierarchy (Domain > Field > Subfield > Topic Name). The model applied explicit boundary definitions and inclusion/exclusion criteria for each of the six frontier technologies. For each record, the model generated a binary flag (is_frontier_tech), assigned the corresponding category (frontier_category), and produced a concise 1–2 sentence technical justification (justification).
- Stage 2: Targeted AI audit and domain refinement: Due to the pervasive nature of digital technology across engineering, an initial single-pass run can risk over-generalizing broad traditional disciplines (e.g., aerospace engineering, civil engineering, building and construction) as "Artificial Intelligence" simply due to keywords like "automation" or "control systems." To resolve this, a second audit pass was executed over topics flagged as AI using a Python execution script. This audit applied strict boundary distinctions:
    - In-Scope (True): Core machine learning/deep learning methodologies, computer vision, natural language processing, robotics learning, AI safety, and AI-specific social science/humanities topics (e.g., *AI Ethics*, *AI Safety & Governance*, *AI in Law*, and *AI Policy*).
    - Out-of-Scope (False): Traditional engineering, physical structural modeling, standard CAD/BIM, legacy control systems, or general IT support that merely utilize computing tools without advancing core AI/ML methods or governance.

Outputs across both stages were validated and structured using schema enforcement.

Prompts

Prompt 1: Initial classification

Please read the attached file containing the CWTS / OpenAlex taxonomy dataset.

Using your Python execution environment, evaluate every topic in the dataset against the 6 defined Frontier Technologies below. Add three new columns to the output file:

1. `is_frontier_tech`: `True` if the topic is core R&D or a direct enabler of one of the 6 technologies, otherwise `False`.
2. `frontier_category`: The name of the matched technology (or `"None"` if False).
3. `justification`: A 1–2 sentence factual explanation for the decision.

Frontier Technology Scope & Criteria:

1. Artificial Intelligence: Machine learning, deep learning, computer vision, NLP, neural networks, autonomous AI systems. *(Exclude: standard statistical analysis or basic IT software)*.
2. Advanced Connectivity: 5G-Advanced/6G, optical communication networks, software-defined networking (SDN), LEO satellite communications, edge networking. *(Exclude: legacy Wi-Fi or basic LAN)*.
3. Cybersecurity: Cryptography, network security, threat detection, post-quantum cryptography, secure systems architecture. *(Exclude: general IT support or database admin)*.
4. Engineering Biology: Synthetic biology, CRISPR/gene editing, metabolic engineering, engineered cell therapies, biomanufacturing. *(Exclude: traditional clinical medicine or basic botany)*.
5. Semiconductors: Microchip design, compound semiconductors, photolithography, novel transistor architectures, wafer manufacturing. *(Exclude: generic electrical engineering)*.
6. Quantum: Quantum computing, quantum sensing, quantum key distribution (QKD), quantum algorithms, topological quantum materials. *(Exclude: general physics without explicit quantum tech application)*.

Please output the complete updated dataset as a downloadable CSV file when complete.

Prompt 2: AI Re-classification

Please load the attached dataset in your Python execution environment and perform a classification audit for the "Artificial Intelligence" frontier technology category.
Write and execute a Python script to update the file directly, and return a downloadable CSV when complete.
Classification Rules for Artificial Intelligence
Add/update three columns in the dataset: `is_frontier_tech`, `frontier_category`, and `justification`.

1. **SET TO TRUE (`is_frontier_tech = True`, `frontier_category = "Artificial Intelligence"`) FOR:**
   - **Core AI/ML R&D:** Machine learning algorithms, deep learning, computer vision, natural language processing, neural networks, reinforcement learning, LLMs, generative AI, pattern recognition, and autonomous AI systems.
   - **AI Safety & Responsible AI:** Technical AI alignment, AI ethics, safety frameworks, mechanistic interpretability, AI governance, AI policy, and legal/societal impacts explicitly centered on AI systems.

2. **SET TO FALSE (`is_frontier_tech = False`, `frontier_category = "None"`) FOR:**
   - **Traditional Engineering:** Aerospace, civil/structural, mechanical engineering, building/construction, HVAC, CAD/BIM, ocean engineering, and general signal/power processing—even if "automation" or "control systems" are mentioned—unless the topic explicitly focuses on core AI/ML algorithms.
   - **General Science & Humanities:** General law, standard economics, broad digital transformation, traditional psychology, basic IT management, or standard statistics without a primary focus on AI.
Technical Requirements for Execution:
- Process the full dataset by evaluating `topic_name`, `subfield_name`, `field_name`, `domain_name`, and `keywords`/`summary` where available.
- For reclassified rows, provide a concise 1-sentence justification explaining the decision.
- Save the processed results as a new CSV file and provide a direct download link.

A2. Technical implementation of maturity measures

A2.1 Volumetric analysis

Introduction

The volumetric maturity framework operationalises bibliometric dynamics to evaluate the evolution of scientific research underpinning Innovate UK projects. Grounded in established scientometric literature on publication growth dynamics, the method evaluates research topics along two orthogonal dimensions: total publication scale and short-term growth momentum. To smooth annual volatility and eliminate single-year reporting artifacts, publication growth is calculated across a rolling three-year window. Rather than imposing arbitrary static thresholds, the framework computes annual median baselines across the entire active topic corpus, dynamically partitioning topics into four developmental quadrants: niche, emerging, active, and mature.

This classification architecture is evaluated through two complementary analytical lenses: Approach A (scientific maturity) measures the unweighted distribution of active topics across maturity stages to establish the baseline structural profile of the research ecosystem, and Approach B (capital-weighted maturity) overlays financial expenditure (£) onto these scientific baselines, measuring how Innovate UK commits its capital stock across developmental stages over time.

Equations

For topic i in calendar year t, let $V_{i,t}$ denote the topic's annual publication volume.

Given a lookback window of $k = 3$ years, the 3-Year Relative Growth Rate is defined as:

$$G_{i,t} = \frac{V_{i,t} - V_{i,t-k}}{V_{i,t-k}}$$

Let $\mathcal{I}_t$represent the set of all active thematic topics evaluated in year t. The dynamic median benchmarks for volume ($\tilde{V}_t$) and growth ($\tilde{G}_t$) in year t are calculated non-parametrically across the portfolio:

$$\tilde{V}_t = \text{median}_{i \in \mathcal{I}_t}(V_{i,t}), \quad \tilde{G}_t = \text{median}_{i \in \mathcal{I}_t}(G_{i,t})$$

The maturity classification function $M(i,t)$partitions topic i in year t into one of four discrete categories based on its position relative to the joint volume-growth median thresholds $(\tilde{V}_t, \tilde{G}_t)$:

$$M(i,t) = \begin{cases} \text{Active,} & \text{if } V_{i,t} > \tilde{V}_t \text{ and } G_{i,t} > \tilde{G}_t \\ \text{Emerging,} & \text{if } V_{i,t} \leq \tilde{V}_t \text{ and } G_{i,t} > \tilde{G}_t \\ \text{Mature,} & \text{if } V_{i,t} > \tilde{V}_t \text{ and } G_{i,t} \leq \tilde{G}_t \\ \text{Niche,} & \text{if } V_{i,t} \leq \tilde{V}_t \text{ and } G_{i,t} \leq \tilde{G}_t \end{cases}$$

To analyze financial allocation across research stages, total grant funding $F_{m,t}$ assigned to maturity category $m \in \{\text{niche, emerging, active, mature}\}$ in year t is aggregated across all constituent topics:

$$F_{m,t} = \sum_{i \in \mathcal{I}_t : M(i,t) = m} F_{i,t}$$

where $F_{i,t}$ is the total grant spend (£) allocated to topic i in year t.

A2.2 Vector analysis

Introduction

To quantify structural shifts in Innovate UK’s overall portfolio posture, we model annual funding distributions as coordinates within a four-dimensional vector space, where each axis represents one of the four scientific maturity stages. Evaluating raw percentage distributions independently across four categories makes it difficult to detect broader portfolio-level rebalancing over time. By calculating the Euclidean distance between Innovate UK's empirical allocation vector and a theoretical benchmark of equal dispersion (where spending is divided evenly at 25% across all four quadrants), we condense complex multi-category movements into a single, intuitive structural metric.

A distance score near zero indicates a highly balanced portfolio posture dispersed uniformly across scientific maturity stages. Conversely, an increasing distance score signals capital concentration toward specific maturity quadrants, reflecting periods of targeted strategic steering or thematic prioritization. Tracking this metric longitudinally illustrates whether the agency's overarching investment posture is expanding toward broad lifecycle coverage or contracting into concentrated developmental phases.

Equations

We define a constant reference vector $\mathbf{u}$ representing a uniform, perfectly balanced portfolio allocation of 25% across each maturity quadrant:

$$\mathbf{u} = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

The observed Innovate UK portfolio distribution in calendar year t is represented as a corresponding vector of empirical proportions $\mathbf{p}_t$:

$$\mathbf{p}_t = \begin{bmatrix} p_{\text{niche},t} \\ p_{\text{emerging},t} \\ p_{\text{active},t} \\ p_{\text{mature},t} \end{bmatrix}, \quad \text{where} \sum_{m \in \{\text{niche, emerging, active, mature}\}} p_{m,t} = 1.0$$

The Euclidean distance $d(\mathbf{p}_t, \mathbf{u})$ between the empirical portfolio allocation and the uniform baseline is calculated as:

$$d(\mathbf{p}_t, \mathbf{u}) = \sqrt{\sum_m (p_{m,t} - 0.25)^2}$$

This spatial calculation yields a continuous scalar metric bounded on the interval $\left[0, \frac{\sqrt{3}}{2}\right]$, where 0 represents exact parity with the uniform benchmark and the theoretical maximum of $\frac{\sqrt{3}}{2} \approx 0.8660$indicates total portfolio concentration within a single maturity quadrant.

A2.3 Frascati weighting

Introduction

The Frascati weighting framework quantifies the translational posture of a research topic by converting venue-level R&D classifications into continuous topic-level distribution vectors. To capture the multi-stage nature of interdisciplinary and translational publications, each journal is evaluated across three hierarchical classification ranks—primary, secondary, and tertiary—mapped to the set of valid Frascati categories. Topic-level maturity profiles are computed under two linear aggregation schemes: an unweighted baseline that attributes publication volume exclusively to a journal's primary classification, and a weighted multi-level scheme that distributes volume across hierarchical ranks using fixed proportional weights.

Equations

Let $k \in \{1,2,3\}$denote the classification rank assigned to a project, corresponding to the Primary, Secondary, and Tertiary Frascati fields respectively. The vector of canonical weights $\mathbf{w} = (w_1, w_2, w_3)$is defined as:

$$w_k = \begin{cases} 0.50, & \text{for } k = 1 \text{ (Primary)} \\ 0.30, & \text{for } k = 2 \text{ (Secondary)} \\ 0.20, & \text{for } k = 3 \text{ (Tertiary)} \end{cases}$$

where $\sum_{k=1}^{3} w_k = 1.00$.

If project j is assigned fewer than three Frascati codes, let $\mathcal{K}_j \subseteq \{1,2,3\}$represent the set of active classification tiers present for project j. The normalized tier weight $\tilde{w}_{j,k}$ assigned to tier $k \in \mathcal{K}_j$is rescaled to preserve full financial allocation ($\sum_{k \in \mathcal{K}_j} \tilde{w}_{j,k} = 1.00$):

$$\tilde{w}_{j,k} = \frac{w_k}{\sum_{m \in \mathcal{K}_j} w_m}$$

A3. Innovate UK funding and GtR+

Table A3.1 - Total funding by data source

Source	Total (£M)	% of funding offered	% of funding spent
Official UKRI (award offered)	£21,383	100.0%	118.8%
UKRI Innovate UK (actual spend)	£17,996	84.2%	100.0%
GtR+ portfolio (total)	£7,815	36.5%	43.4%
Analysis dataset (topic-mapped GtR+ portfolio)	£4,862	22.7%	27.0%

Table A3.2 - Annual funding by data source

Year	UKRI offered (£M)	UKRI spend (£M)	GtR+ total (£M)	Analysis (topic-mapped) dataset (£M)	GtR+ % of UKRI offered	GtR+ % of UKRI spend	Analysis dataset % of GtR+
2006	£188.70	£175.43	£20.26	£4.14	10.70%	11.60%	20.40%
2007	£165.88	£152.96	£35.91	£13.58	21.70%	23.50%	37.80%
2008	£216.87	£194.42	£80.57	£51.25	37.10%	41.40%	63.60%
2009	£279.42	£248.35	£145.09	£106.04	51.90%	58.40%	73.10%
2010	£271.59	£256.13	£129.31	£72.78	47.60%	50.50%	56.30%
2011	£339.79	£328.06	£84.12	£53.33	24.80%	25.60%	63.40%
2012	£512.55	£482.31	£255.46	£119.40	49.80%	53.00%	46.70%
2013	£574.63	£540.42	£328.63	£177.87	57.20%	60.80%	54.10%
2014	£781.17	£704.88	£468.95	£238.11	60.00%	66.50%	50.80%
2015	£946.74	£798.06	£641.91	£206.10	67.80%	80.40%	32.10%

2016	£1,005.91	£959.11	£335.04	£244.87	33.30%	34.90%	73.10%
2017	£1,062.38	£1,014.57	£499.04	£331.89	47.00%	49.20%	66.50%
2018	£2,554.51	£2,391.32	£783.57	£419.22	30.70%	32.80%	53.50%
2019	£1,158.24	£1,064.15	£486.52	£391.87	42.00%	45.70%	80.50%
2020	£1,560.20	£1,402.48	£647.82	£369.32	41.50%	46.20%	57.00%
2021	£1,360.12	£1,233.52	£445.10	£282.89	32.70%	36.10%	63.60%
2022	£1,630.91	£1,403.67	£670.60	£486.13	41.10%	47.80%	72.50%
2023	£3,642.32	£2,770.53	£937.25	£704.53	25.70%	33.80%	75.20%
2024	£1,665.00	£1,140.33	£669.71	£493.80	40.20%	58.70%	73.70%
2025	£800.50	£419.48	£149.68	£94.97	18.70%	35.70%	63.40%

A4. Frontier technologies

Table A4.1 - Frontier technologies and associated topics

Frontier technology	# Topics	Topics
Artificial Intelligence	78	Topic Modeling Advanced Neural Network Applications Medical Image Segmentation Techniques Face and Expression Recognition Natural Language Processing Techniques Speech Recognition and Synthesis Recommender Systems and Techniques Neural Networks and Applications Video Surveillance and Tracking Methods Advanced Steganography and Watermarking Techniques Context-Aware Activity Recognition Systems Multi-Agent Systems and Negotiation Reinforcement Learning in Robotics Advanced Vision and Imaging Data Mining Algorithms and Applications Handwritten Text Recognition Techniques Advanced Image and Video Retrieval Techniques Robot Manipulation and Learning Sentiment Analysis and Opinion Mining Emotion and Mood Recognition Image and Signal Denoising Methods Remote-Sensing Image Classification Privacy-Preserving Technologies in Data Generative Adversarial Networks and Image Synthesis Human Pose and Action Recognition Fuzzy Logic and Control Systems Image Retrieval and Classification Techniques Speech and Audio Processing Ethics and Social Impacts of AI Advanced Data Compression Techniques AI-based Problem Solving and Planning

Frontier technology	# Topics	Topics
		Image Enhancement Techniques Advanced Image Processing Techniques Image and Video Quality Assessment Model Reduction and Neural Networks Advanced Graph Neural Networks Music and Audio Processing Hand Gesture Recognition Systems Face recognition and analysis Anomaly Detection Techniques and Applications Text and Document Classification Technologies Artificial Intelligence in Games Visual Attention and Saliency Detection Stochastic Gradient Optimization Techniques Artificial Intelligence in Healthcare and Education Adversarial Robustness in Machine Learning Multimodal Machine Learning Applications Evolutionary Algorithms and Applications Explainable Artificial Intelligence (XAI) Machine Learning and Algorithms Industrial Vision Systems and Defect Detection AI in Service Interactions Time Series Analysis and Forecasting Hate Speech and Cyberbullying Detection Authorship Attribution and Profiling Machine Learning and Data Classification Image and Object Detection Techniques Neural Networks and Reservoir Computing Machine Learning and ELM Vehicle License Plate Recognition Gait Recognition and Analysis Adaptive Dynamic Programming Control Gaussian Processes and Bayesian Inference Advanced Text Analysis Techniques Digitalization, Law, and Regulation Data Mining and Machine Learning Applications Text Readability and Simplification Artificial Intelligence in Law AI and HR Technologies Law, AI, and Intellectual Property Artificial Intelligence Applications Mathematical Control Systems and Analysis Artificial Intelligence and Decision Support Systems Impact of AI and Big Data on Business and Society Currency Recognition and Detection Statistical and Computational Modeling Advanced Technologies in Various Fields Artificial Intelligence in Education
Semiconductors	20	Semiconductor Quantum Structures and Devices

Frontier technology	# Topics	Topics
		Radio Frequency Integrated Circuit Design Photonic and Optical Devices Analog and Mixed-Signal Circuit Design Low-power high-performance VLSI design Advancements in Semiconductor Devices and Circuit Design Radiation Effects in Electronics Particle Detector Development and Performance Advanced Power Amplifier Design Advancements in Photolithography Techniques Advancements in PLL and VCO Technologies VLSI and FPGA Design Techniques 3D IC and TSV technologies Advanced Semiconductor Detectors and Materials CCD and CMOS Imaging Sensors Advanced Optical Sensing Technologies solar cell performance optimization Electrostatic Discharge in Electronics Magneto-Optical Properties and Applications Integrated Circuits and Semiconductor Failure Analysis
Quantum	14	Quantum and electron transport phenomena Diamond and Carbon-based Materials Research Magnetism in coordination complexes Topological Materials and Phenomena Scientific Research and Discoveries Black Holes and Theoretical Physics Quantum Chromodynamics and Particle Interactions Cold Atom Physics and Bose-Einstein Condensates Atomic and molecular physics Quantum Mechanics and Applications Quantum Computing Algorithms and Architecture Quantum chaos and dynamical systems Quantum many-body systems Quantum Electrodynamics and Casimir Effect
Advanced Connectivity	13	Advanced MIMO Systems Optimization Optical Network Technologies Advanced Photonic Communication Systems Advanced Optical Network Technologies Optical Wireless Communication Technologies Millimeter-Wave Propagation and Modeling UAV Applications and Optimization Advanced Wireless Communication Technologies PAPR reduction in OFDM Satellite Communication Systems IoT Networks and Protocols Full-Duplex Wireless Communications Telecommunications and Broadcasting Technologies

Frontier technology	# Topics	Topics
Cybersecurity	12	Quantum Information and Cryptography Cryptography and Data Security Information and Cyber Security Coding theory and cryptography Polynomial and algebraic computation Security in Wireless Sensor Networks Cryptography and Residue Arithmetic Cellular Automata and Applications Cybersecurity and Cyber Warfare Studies Network Packet Processing and Optimization Urban Transport Systems Analysis Data Privacy and Cybersecurity
Engineering Biology	9	RNA and protein synthesis mechanisms Gene Regulatory Network Analysis CRISPR and Genetic Engineering Microbial Metabolic Engineering and Bioproduction CAR-T cell therapy research Plant biochemistry and biosynthesis Porphyrin Metabolism and Disorders Microbial metabolism and enzyme function Biochemical and biochemical processes

A5. Frascati analysis

The Frascati classification provides a direct measure of research orientation and translational proximity, establishing how applied or market-focused the underpinning scientific base is for any given domain. This enables us to evaluate whether Innovate UK's grant allocations align with its core mandate of supporting downstream, business-led R&D and experimental development, rather than upstream fundamental discovery.

Here we present three supplementary analyses:

- First, we descriptively compare the Frascati classification to the volumetric maturity categories, by measuring the proportion of publications in basic, applied, experimental and general research journals within each of the volumetric maturity bins.

- Second, to complement the volumetric maturity model, we map each topic's Frascati research profile across the maturity spectrum. This cross-framework characterisation contextualises the nature of scientific output within each bin - identifying, for example, whether 'Niche' domains represent specialised applied niches or foundational basic research, and how the balance between theoretical inquiry and experimental development shifts as topics scale.

- Third, to support the volumetric timing metrics, we evaluate the Frascati composition of each topic at the exact moment of peak Innovate UK investment to observe the scientific character of the underlying base when public funding is concentrated.

A5.1 Maturity and Frascati categories

Across all four volumetric maturity stages, applied research represents the dominant share of underpinning scientific literature, maintaining a remarkably stable mean within-topic share between 54.8% and 56.8% (with topic-level medians consistently hovering around 55–57%). Basic research displays a modest upward trend across developmental stages, moving from 20.0% in niche and 19.4% in emerging topics up to 21.2% in active and 24.7% in mature topics. Conversely, general research contracts progressively as topics scale, falling from 13.0% in niche domains down to 7.8% in mature domains. Experimental development remains steady across the entire scientific lifecycle, accounting for between 12.2% and 13.6% of publication share across all four maturity bins.

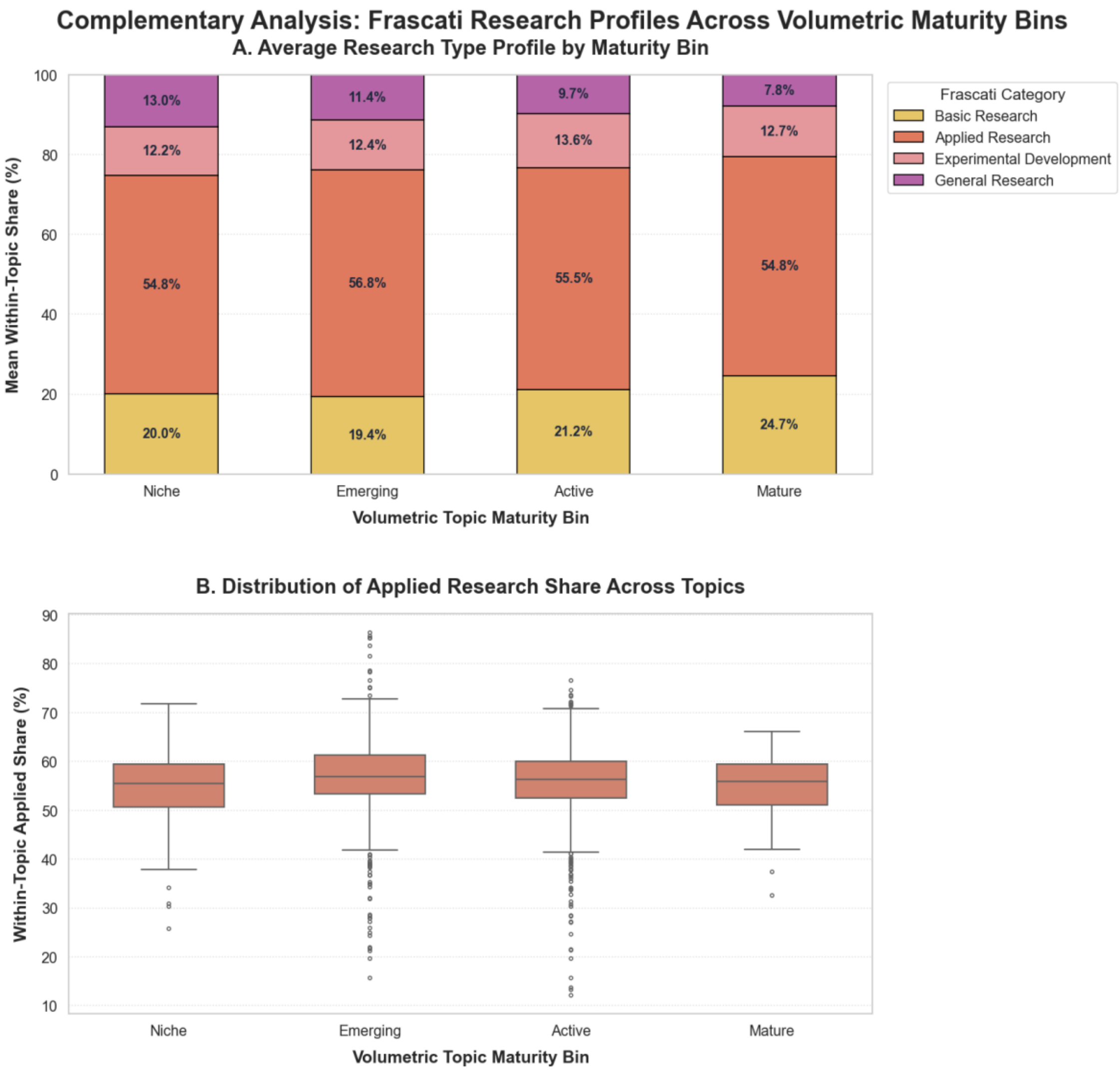

Chart A5.1 - Frascati maturity categories (weighted) in each of the volumetric maturity bins for Innovate UK project topics (R&D only)

A5.2 Frascati profiles over time

Between 2006 and 2025, the longitudinal distribution of weighted Frascati publication shares demonstrates a gradual slight shift toward greater applied concentration within the supporting research base. Applied research expanded its majority share from 55% (maintained between 2006 and 2013) to 59% by 2023–2025. Over the same twenty-year horizon, Basic research experienced a slight contraction from 24% in 2006 down to 21% in 2024 and 2025, while general research decreased from 8–9% down to 6%. Experimental development exhibited notable structural stability across the entire timeline, remaining tightly bounded between 13% and 15% of annual publication output.

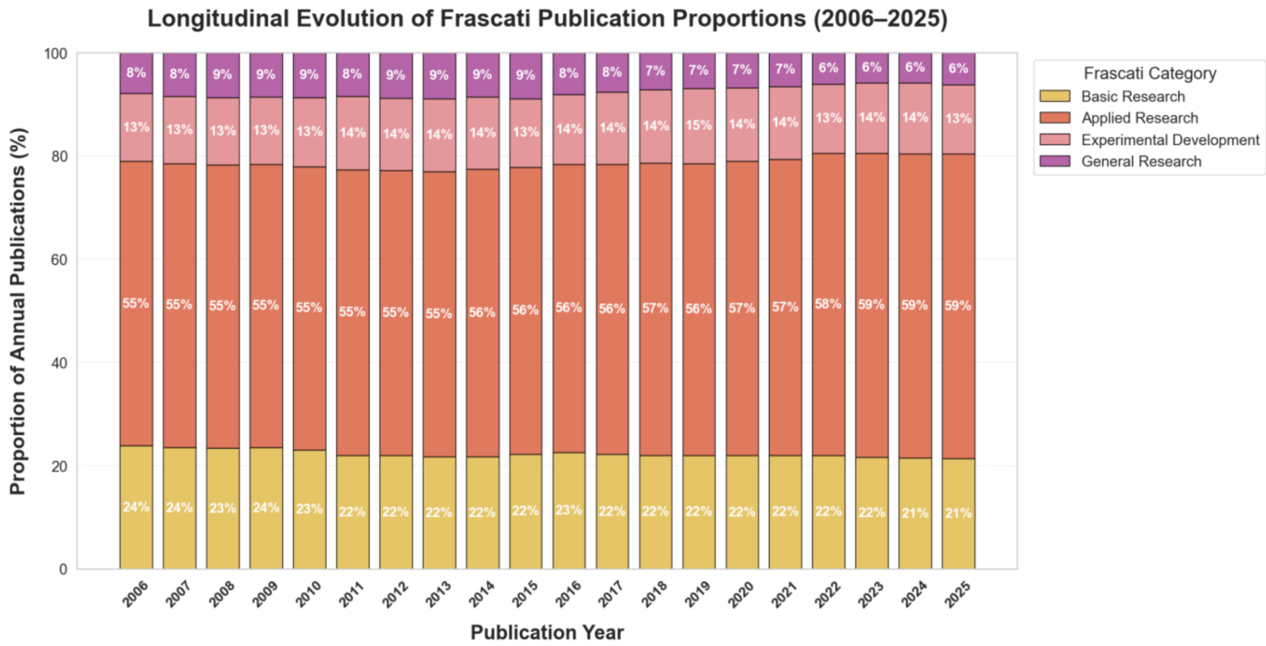

Chart A5.2 - Frascati maturity categories (weighted) of topics associated with Innovate UK innovative projects (R&D only), 2006-2026

A5.3 Peak investment Frascati categories

Evaluating the dominant Frascati research orientation of individual topics at their exact year of peak Innovate UK grant investment reveals an overwhelming concentration in applied scientific domains. Out of all evaluated topics, 1,799 topics (95.8%) are dominated by applied research at their point of maximum capital allocation. In contrast, general research represents the primary category for 53 topics (2.8%), while basic research forms the dominant base for just 26 topics (1.4%). Notably, experimental development serves as the primary orientation for no topics (0.0%). This near-total dominance underscores that public funding scale-ups systematically coincide with research domains whose supporting evidence base has already matured into predominantly applied, practical solutions rather than early-stage fundamental discovery or standalone development venues.

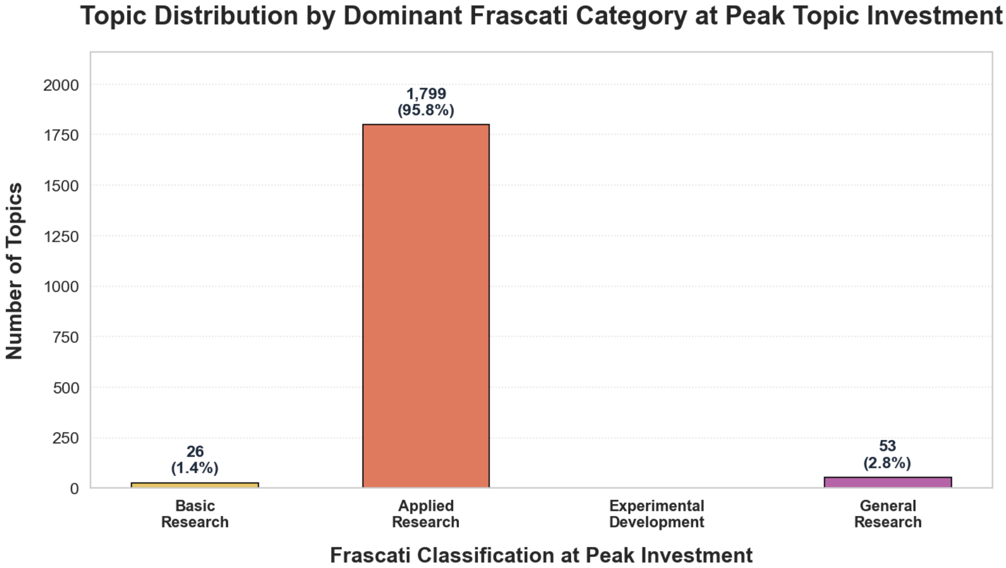

Chart A5.3 - Frascati maturity classification (weighted) in year of peak investment for Innovate UK project topics (R&D only)

A6. Supplementary visualisations

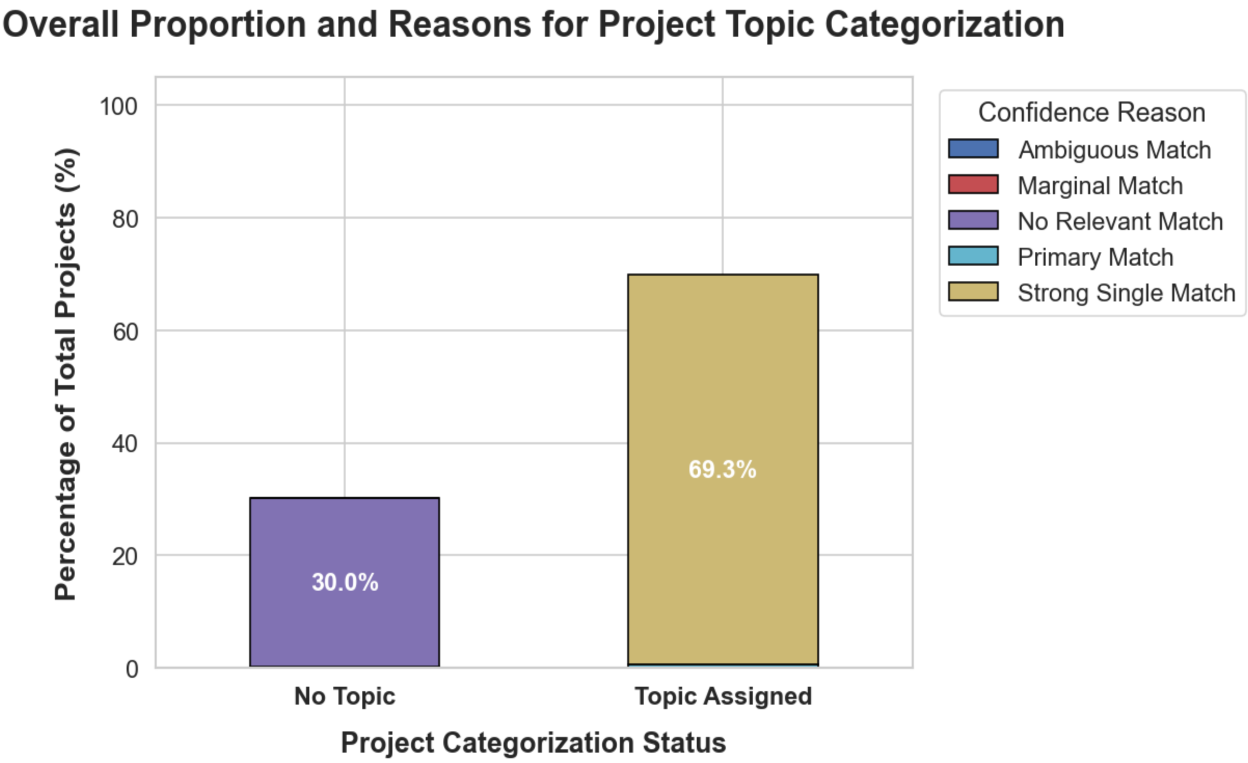

Chart A6.1 - LLM topic classification results and rationale provided

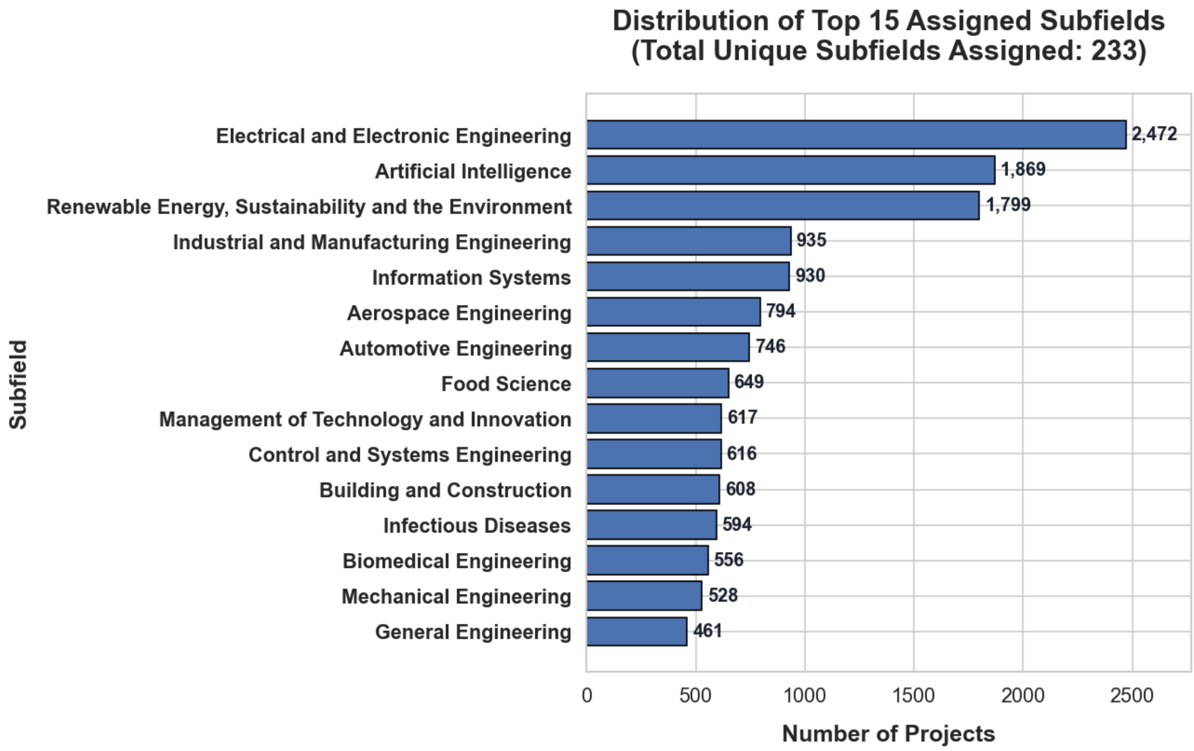

Chart A6.2 - Top 15 assigned subfields (2006-2025)

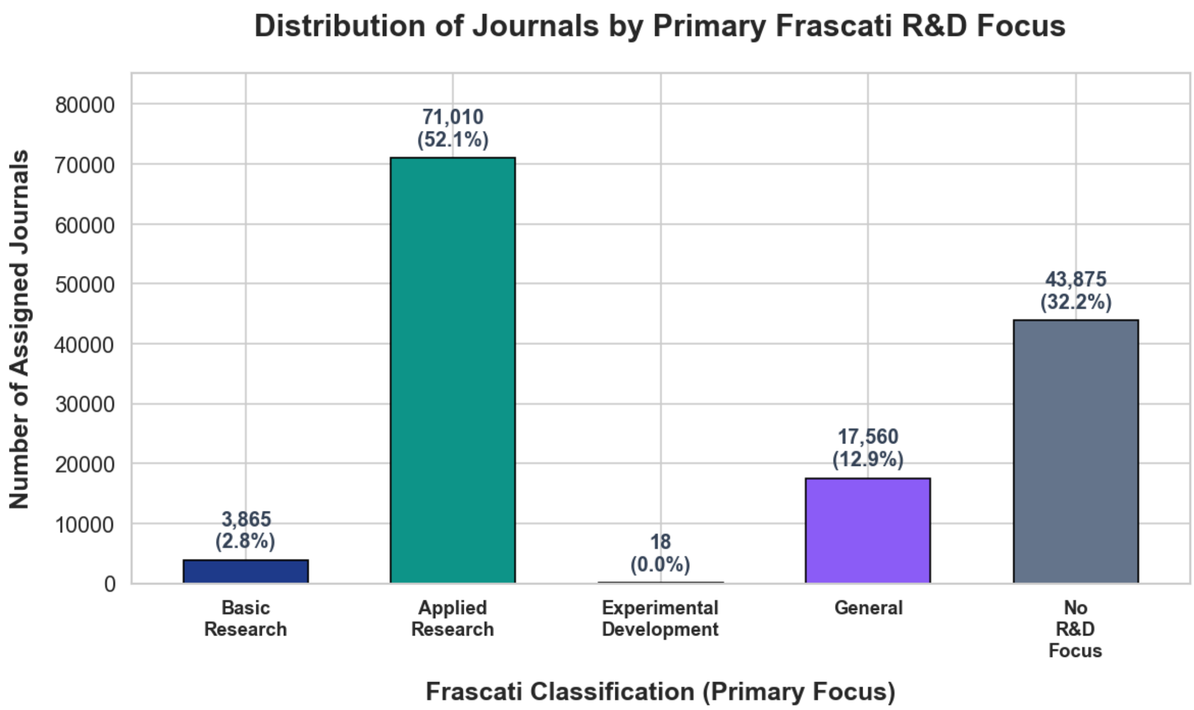

Chart A6.3 - Frascati category assignment, primary only (unweighted)

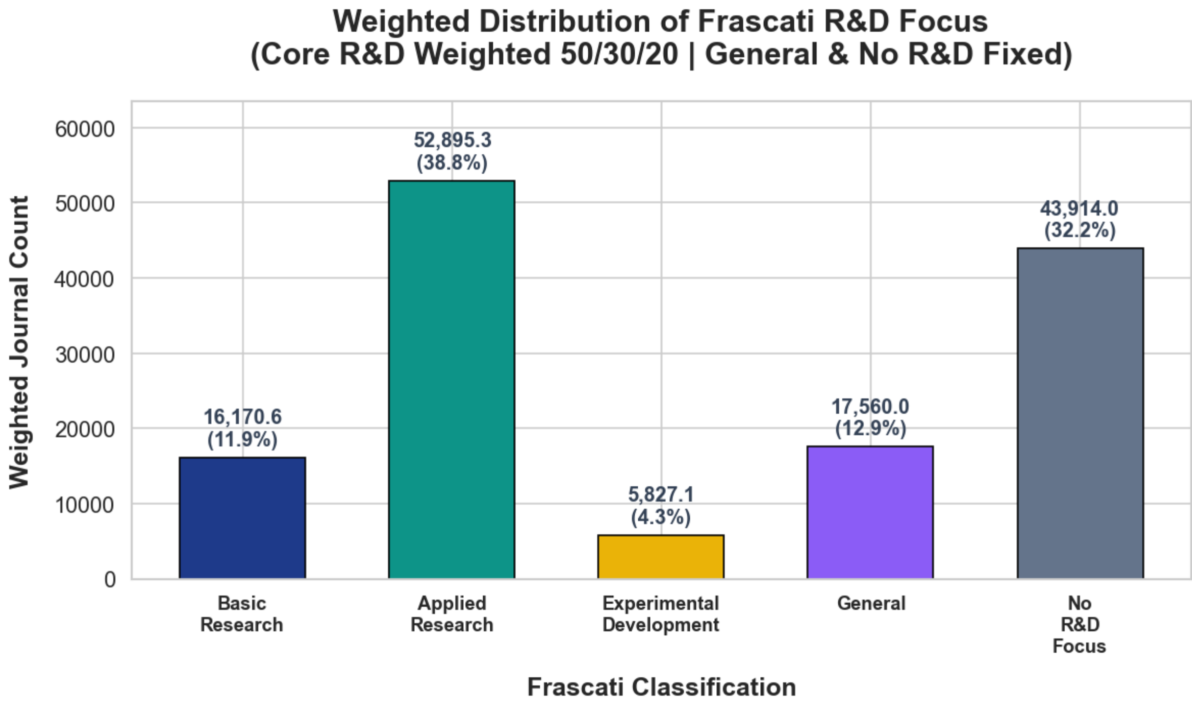

Chart A6.4 - Frascati category assignment, primary, secondary and tertiary - weighted

www.ircaucus.ac.uk

Email info@ircaucus.ac.uk Twitter @IRCaucus

INNOVATION & RESEARCH CAUCUS

Delivered with ESRC and Innovate UK