

Leeds Institute for Data Analytics

2024 - 29

STRATEGY

For multidisciplinary data science and artificial intelligence at the University of Leeds

LIDA



Table of Contents

1. Foreward	3
2. Executive summary	4
3. Introduction	7
4. Meet the team	9
5. Our values	14
6. Research & Innovation	16
7. Education & Training	22
8. Research Technology	29
9. Operational Enablement	36
10. Equity, Diversity & Inclusion	40
11. Environmental Sustainability	45

Data science and AI is one of the fastest-moving areas of technological and societal advancement, with the rise of generative AI driving a dramatic surge in visibility. Higher education, government and business organisations are recognising the need to accelerate the development of data science and AI platforms and models to be at the forefront of innovations and new discoveries. This is creating vast opportunities for innovative academic data science and AI research, education and training to tackle some of the most pressing global challenges, including across Health, Society and the Environment. It is also leading to a rapid expansion, not just in the need for data scientists, but new roles such as data engineers to process and wrangle data and machine learning engineers to build models.

LIDA was established in 2014 and celebrates its 10-year anniversary in 2024/25. Funded originally by two external grants from the Economic and Social Research Council and Medical Research Council it now enables and contributes directly to a multi-disciplinary data science and AI research portfolio worth over £90m. This 5 year strategy builds on these strong foundations by harnessing the University's collective knowledge and skills in data science and AI to increase research impact and develop the leaders of the future in this field. During this period there has also been a rapid advancement in the demand for, and capabilities of, data security and artificial intelligence technologies.

LIDA's in-house Trusted Research Environment, LASER, will remain at the forefront of this dramatic change, providing cutting edge infrastructure for multidisciplinary academic teams, both internal and external, to continue to undertake research involving both AI models and highly sensitive data, securely and ethically.

As we begin to understand the policy agenda for the new Government, it is clear that AI and data science research and education will remain a critical national priority for achieving positive societal impact and UK commercial success.

This strategy recognises the critical importance of LIDA's collaboration with our partners, N8 and Alan Turing Institute, to this national mission, enabling us to collaborate with fellow academics across disciplines and the UK to generate knowledge and impact at scale.

The University of Leeds produces high-quality, challenge led, interdisciplinary research and education to meet societal challenges, at scale. LIDA is a perfect example of how the University of Leeds aims to address these challenges and the University and I are proud to endorse this 5 year strategy.



Prof Nick Plant
Deputy Vice-Chancellor:
Research and Innovation

2

Executive summary



LIDA's vision is to generate societal impact and improve global wellbeing and prosperity through internationally outstanding research, education and training, in interdisciplinary data science and artificial intelligence.

To achieve this LIDA's overarching strategic challenge and opportunity, is to build upon existing strengths across the University so that:

- UoL's Data Science and AI ecosystem is more than the sum of its parts.
- UoL offers a coherent and distinct proposition in data science and AI to compete nationally and internationally.
- UoL provides a home for capacity building and the development of future leaders in the field.

The strategy is broken down into four portfolios; Research and Innovation, Education and Training, Research Technology and Operational Enablement with the following strategic aims (underpinned by short term (1 year), medium (2-3 year) and long term (3-5 year) objectives);

Research and Innovation

- Produce world-class cross disciplinary research.
- Diversify & increase data science & AI research income.
- Deliver impactful research & innovation.
- Develop, support & retain an inclusive research community.

Education and Training

- Make a positive difference to society.
- Be a trusted global collaborator.
- Develop, support & retain a diverse and inclusive education and training community.
- Increase successful outcomes for data science and AI students, and early career data scientists and learners.
- Provide a healthy, safe & inclusive environment and enriching experience for staff and students.
- Diversify and increase data science and AI education and training income.
- Ensure a sustainable and balanced cohort mix in data science and AI education and training.

Research Technology

- Provide high end, secure and scalable research computing infrastructure.
- Develop processes and tools to improve costing and research recoveries.
- Encourage open research practice.
- Deliver impactful research and innovation.
- Be a trusted collaborator.

Operational Enablement

- Developing, supporting & retaining a diverse and inclusive professional services staff community.
- LIDA's facilities provide a best-in-class, collaborative work environment.
- LIDA is a trusted international collaborator.
- LIDA is adaptive, effective and financially and environmentally sustainable.

Equity, Diversity and Inclusion

Research finds that certain groups (such as women, those from minority ethnic backgrounds and people with disabilities) are underrepresented in the data workforce. A lack of workforce diversity has the potential to amplify existing inequalities and prejudices.

LIDA will contribute to [UoL's EDI Strategy](#) to eliminate gaps between groups with protected characteristics and those without, by focussing on the following strategic aims:

- Improve the development, progression, pay and retention rates of our staff and communities across protected characteristics (e.g. race, gender, disability, sexual orientation, age).
- Increase the engagement of LIDA managerial and academic research, education and training community staff with leadership development, to grow capability and help drive forward LIDA's strategy.
- Improve the socio – economic diversity of our research, education and training communities.

Environment Sustainability

LIDA will also contribute to the University's Net Zero targets by 2030 by:

- Reducing the primary energy demand of our building and estate.
- Reducing emissions from business travel.
- Decarbonise energy usage related to LIDA's use of gas and electricity in our building and use of compute infrastructures.

3

Introduction



Data science is an umbrella term for fields that use large quantities of data to discover actionable insights. Data science typically begins with data collection, processing and storage, followed by analysis and modelling. The final stages may involve AI, a branch of computer science concerned with developing systems that carry out tasks normally requiring human intelligence.

Data Science and AI are increasingly seen as major drivers in the growth of the UK economy over the next 20 years. AI in particular represents a £550 billion opportunity for the UK over the next 10 years[1]. With the most advanced tech sector in Europe, second globally to the US, the UK is well set up for AI leadership.

In 2021, the UK Government launched the National AI Strategy, a 10-year action plan to make the UK an 'AI superpower'. Digital technology is already one of the biggest drivers of UK growth and the country has attracted over £20 billion in private investment in AI since 2016. Potential social benefits range from climate change mitigation, to modelling social behaviour in response to the Covid-19 pandemic, to improving early detection and diagnosis of cancers.

Leeds is recognised as an important and growing national centre for 'deep tech' with AI and data science at its heart, with national organisations such as The Bank of England, Channel 4 and ASDA accessing the city's skilled pool of data scientists and software engineers.

At the University of Leeds, data science and AI is used cross-faculty, becoming ever more prominent in all research from Earth & Environment, to Transport Studies, to Health and Medicine.

LIDA is the cross-faculty institute at the university with expertise in multi-disciplinary data science and artificial intelligence (AI). With our support, AI techniques will be disseminated and immersed across the University for the benefit of students, researchers, academics and wider society.

LIDA's vision is to generate societal impact and improve global wellbeing and prosperity through internationally outstanding research, education and training, in interdisciplinary data science and artificial intelligence.

This strategy sets out how LIDA will lead and nurture the University's data science and AI ecosystem over the next 5 years, through the development of research and innovation, knowledge and skills.



Prof Chris Gale,
Co-Director

[1] <https://microsoftuk.publicfirst.co.uk/>

Overview of LIDA

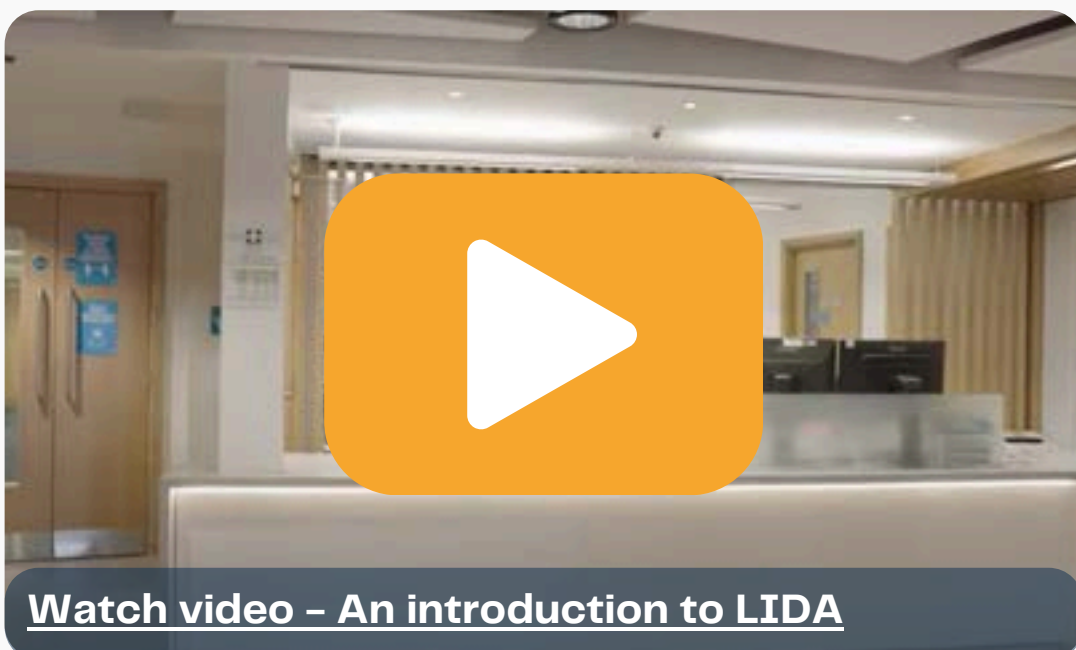
LIDA brings together data science and AI researchers, students and professionals from all disciplines, opening up new opportunities to understand health and human behaviour and casting light on the action required to tackle a wide range of problems.

Connecting academic research, education and training with external partners in business, government and the third sector, LIDA matches the world class data science and AI capabilities of UoL with regional and national priorities.

LIDA's overarching strategic challenge and opportunity, is to build upon existing strengths across the University of Leeds (UoL) so that:

- UoL Data Science and AI ecosystem is more than the sum of its parts
- UoL offers a coherent and distinct proposition in data science and AI to compete nationally and internationally.
- UoL provides a home for the development of future leaders and capacity building capacity building

LIDA is funded by 8 Academic Faculties to support data science and AI research, education and training across the University. We also work closely with the University's Research Innovation Service to promote innovation and build collaborative partnerships and across Leeds, West Yorkshire and beyond.



A summary of our core Data Science and AI portfolio and services

Research & Innovation

Alan Turing Institute partner for the University of Leeds

Programmes

Artificial Intelligence
Immersive Technologies
Data Science Infrastructures
Data Visualization
Science of Data Science

Communities

Food
Health
Societies
Environment

Research portfolio worth >£90million

Research Centres

Consumer Data Research Centre (CDRC) | Centre for Immersive Technologies

Professional Services Staff

Support day to day research collaboration inc. funding proposals, events and office administration

Collaborative physical space

Desks, meeting rooms and offices for the hybrid working of 100+ co located researchers and learners

Education & Training

10x MSc programmes with 685 students

2x PhD Centres for Doctoral Training

CPD short-courses and seminars

Data Scientist Development Programme

A 12-month, paid, on the job training role for aspiring data scientists

Open Data Science for Schools

Outreach initiative to encourage more children to consider data science as a career path

Research Technology

Data Analytics Team (DAT)

Support researchers through the full project lifecycle

Information Governance Team

Ensure sensitive data is protected, working directly with the DAT

LASER

Our Trusted Research Environment for storing and processing sensitive data for research projects

Safe rooms for research

available in house at LIDA and across the UK via the SafePod Network

4

Meet the team



Executive team



Prof Chris Gale, Co-Director

I am jointly responsible for the strategic direction and oversight of LIDA, to grow world class data science and AI at Leeds, in alignment with the Universities 10 year strategy.



Prof Susan Grant-Muller, Co-Director

I am jointly responsible for the strategic vision and leadership of LIDA, ensuring we continue to work in collaboration with all faculties and multi-disciplinary academic communities.



Dr Roger Beecham, Director of Research & Innovation

I lead LIDA's research and innovation strategy, working in partnership with the pan faculty research Communities, Programmes and Special Interest Groups.



Prof Paul Baxter, Director of Education & Training

I lead on the enhancement and growth of Data Science & AI (DS&AI) education across the University in partnership with our Schools and Faculties.



Prof Roy Ruddle, Director of Research Technology

I am responsible for the computer hardware, software and wetware (technical expertise) that are used by researchers in LIDA.



Dominic Frankis, Director of Operations

I lead the professional services team in LIDA, who manage our intellectual property, data, digital and physical environment to support our academic objectives and community wellbeing.

Meet the team



Kim Wright
Business Co-ordinator
/ Turing Liaison



Karen Fletcher
Office
Administrator



Renata Lopes
Executive
PA



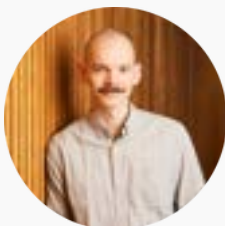
Andrea Rylands
IT Governance
Manager



Hannah Jackson
Communications
Manager



Kylie Norman
Data Scientist Development
Programme Manager



Adam Keeley
Data Analytics
Team Manager



Jenny Sexton
Lecturer in Data
Science



Jodi Gunning
Research & Innovation
Development Officer



Fojan Ilderem
Research Software
Engineer



Ifeanyi Chukwu
Research Software
Engineer



**Obosekoghene
Eselebor**
Research Software
Technician



Lydia Wharton
Research Software
Technician

5

Our values





Collaboration

Our communities work together across academic disciplines, research and education and organisational boundaries with a shared identity and purpose to make a positive difference to the world.



Compassion

Our communities are caring and considerate in our words and actions. We listen, are empathetic and we treat each other with respect which make our communities happier and more successful.



Inclusivity

Our communities are equitable, diverse and inclusive, welcoming underrepresented groups, enhancing both their sense of belonging and their and LIDA's success.



Integrity

Our communities are open and honest to foster innovation and build trust by providing an environment where individuals comfortably recognise our own and each other's areas of strength and need for support.



Contribution

Our communities will focus our efforts on the research and education areas in which we can be a global leader. We will combine our global outlook with a keen awareness of the importance national, regional and local impact and links with business and the civic community through outreach and co-production.

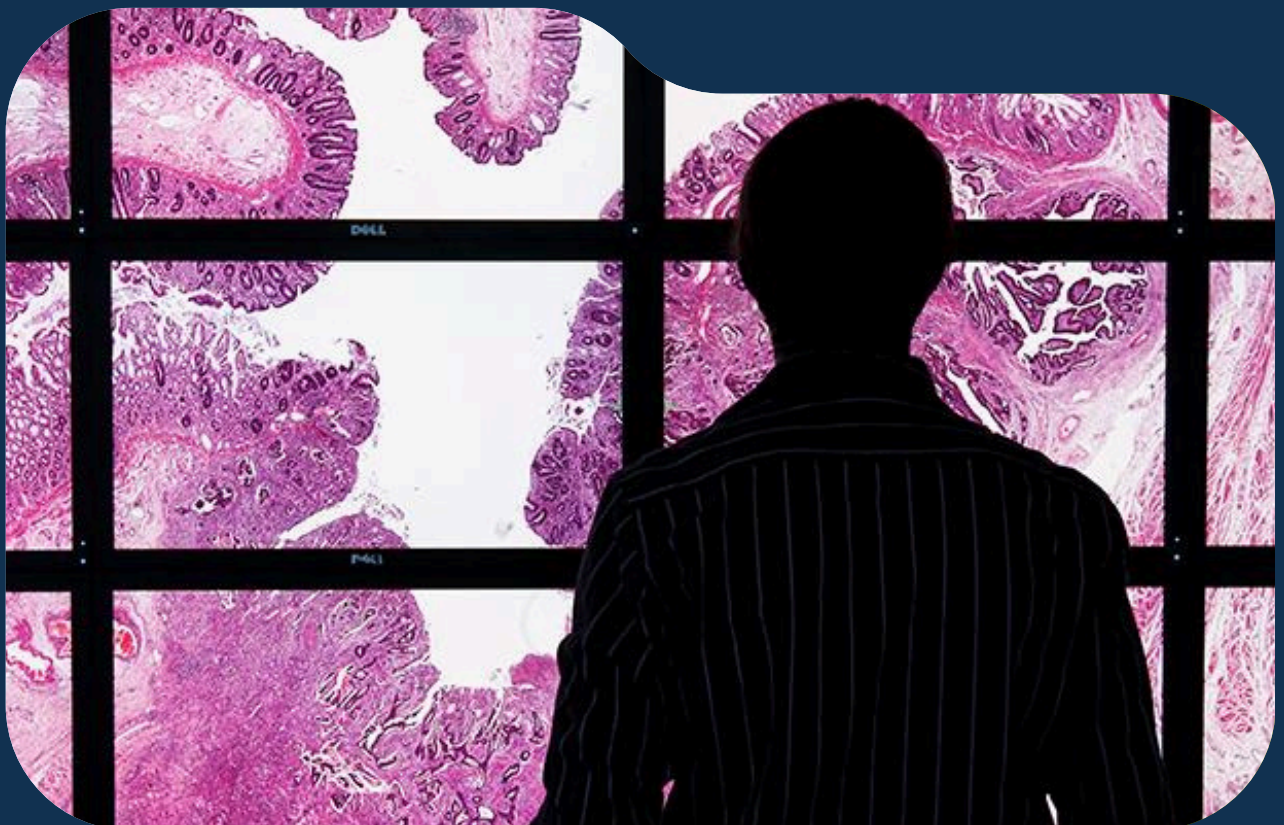


Sustainability

We will make a positive difference to the world in a way that delivers value for money for the University of Leeds and is environmentally sustainable.

6

Research & Innovation



Challenges & Opportunities

LIDA's vision is to deliver world class and impactful research and innovation.

To date in the current Research Excellence Framework (REF2029) period, LIDA's grant portfolio comprises 155 projects, supported by £120 million of external awards, from 43 funders. LIDA's research projects span more than 58 schools and institutes. Two thirds of those projects are interdisciplinary, bringing together the necessary expertise from a network of 400+ academics and researchers.

LIDA's research communities are focused on four main challenge areas that require interdisciplinary working: Food, Societies, Health and Environment. Supporting these are four cross cutting programmes of work that focus on developing our capabilities and capacity; Statistical Data Science, Artificial Intelligence, Immersive Technologies, Data Visualisation and Data Science Infrastructures (see Fig 1.)

In their REF2021 environment statements, 40% (11 out of 28) of Leeds units of assessment (UOAs) flagged their LIDA research as strategically important. To strengthen that wide-reach, we need to understand the manifold ways in which research becomes world class across different disciplines, including the approaches to and expectations of interdisciplinarity, and achieve critical mass in certain key areas.

Here the opportunity is to use our vibrant research communities and methodological programmes to create a further step-change in interdisciplinary collaboration. Complementing these research strengths there are opportunities to increase LIDA's research capability in new areas such as engineering, maths and computing, physical, social and biological sciences and arts, humanities and cultures.



Strategic Aims & Objectives

Strategic Aim	Short term objectives (1 year)	Medium term objectives (2-3 years)	Long term objectives (4-5 years)
1. Produce world-class cross-disciplinary research	<p>A) Scoping to characterise what makes LIDA research world-leading and specify a blueprint for successful data science and AI research. In-focus will be composition of academic teams, expert partners and stakeholder organisations, and output types produced - fundamental papers, software tools, datasets</p>	<p>B) Develop a pipeline of new research activity, via e.g. DSDP, that recognises the blueprint for success in 1A).</p> <p>C) Establish critical mass for LIDA in all Schools and Faculties engaged in core and applied data analysis.</p>	<p>D) Make strong LIDA research output contributions that deliver a positive impact to society and to REF2029, across many units of assessment.</p>
2. Diversify & increase data science & AI research income	<p>A) Develop new models for generating and increasing data science and AI research income, working with RIS.</p>	<p>B) Strengthen LIDA’s research portfolio, in underrepresented faculties/departments, and include full range of research councils and other funders.</p>	<p>C) Develop new research programmes.</p> <p>D) Secure long-term high status research funding streams.</p>
3. Deliver impactful research & innovation	<p>A) Identify exemplars of impact-in-the-small and impact-in-the-large, from applied and underpinning data science disciplines;</p> <p>B) Identify key and emerging research agendas, technologies and methods in modern data science, specialised to research discipline.</p>	<p>C) Establish a pipeline to impact for data science research.</p> <p>D) Engage with internal and external stakeholders, through LIDA’s data analytics communities, programmes, CDTs and DSDP.</p>	<p>E) Make tangible contributions to REF2029 impact case studies, across several units of assessment. Address global challenges and those with national and regional socio-economic impact.</p> <p>F) Build and sustain strategic impactful global partnerships, and a strong academic reputation, in specific areas e.g. food science.</p>
4. Develop, support & retain an inclusive research community	<p>A) Attract and nurture LIDA researchers. Proactively encourage new and established staff to partake in LIDA research. Provide guidance and career support.</p> <p>B) LIDA Academic Fellows in Data Science and Artificial Intelligence - support the development of early career researchers in data science, to become Associate Professors and future leaders.</p>	<p>C) Bring together businesses, public sector organisation, researchers, technical and professional services staff.</p> <p>D) Increase opportunities for researchers and professional services staff (especially protected characteristics) for development, progression and leadership</p>	<p>E) Develop career progression for LIDA researchers.</p>

Strategic Partnerships



The Alan Turing Institute

The University became a partner of the Alan Turing Institute in 2018, with LIDA leading the relationship.

We will play a full role in delivering the new [Turing strategy](#) on behalf of the nation and research communities worldwide.

LIDA will contribute to the Turing's four principal strategic actions by:

- **Participating in Turing grand challenges in 'Health' and 'Environment and Sustainability'**, utilising the university's world-leading interdisciplinary research across 'health and the environment', 'socioeconomic status and multimorbidity', and 'health and nutrition'.
- **Advancing foundational research**, particularly in the areas of machine learning and optimisation, computer vision (image segmentation, video analysis), knowledge representation and reasoning, natural language processing (large language models and reasoning), augmenting human intelligence, and robotics.
- **Targeting the national skills gap** through our Centres for Doctoral Training, and building capacity through the DSDP, LIDA Research Fellowships and CPD courses.
- **The practical translation and application** of AI research, education and training to Industry and Government, by linking with Nexus and Policy Leeds.
- **Actively contributing to the future development of the Turing**, such as by promoting the importance of cognitive science and robotics.
- **Working with the wider Turing community** to build capacity for interdisciplinary research with collaboration across institutions, including seeking opportunities to link our other research partnerships such as WUN, N8, White Rose, Met Office Academic Partnerships (MOAP).
- **Contributing to the societal understanding of AI** through the University's research expertise in applied ethics, law, media and communications.



Strategic Partnerships



LIDA and the N8 research partnership

N8 have four data science intensive areas of research - Child of the North, Computationally intensive research, Net Zero North and Policing Research Partnership.

These are aligned with LIDA's research communities and programmes and we will actively support the Universities' data scientists to collaborate with other universities, industry and society on these areas for societal benefit.

The Digital Health theme shares expertise and builds community across the N8 within health data analytics, vital sign monitoring, sensors, drug discovery and multi-omics. In recent years the focus has been primarily on community building and training, but has recently expanded to include running grant-development workshops – looking to bring together diverse teams to apply for EPSRC / NIHR / MRC grant funding. LIDA will support by hosting and contributing its data science expertise and experience using state of the art secure infrastructure capabilities.

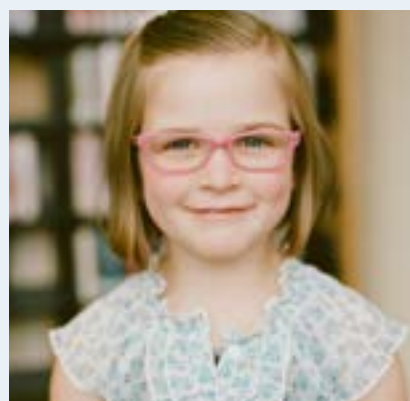
Case study: Child of the North

LIDA will be supporting the action of three key recommendations made in 'Addressing Education and Health Inequity: Perspectives from the North of England'.

1) The methodology used by the Centre for Applied Education Research (CAER) and Child Health Outcomes Research At Leeds (CHORAL) should be rolled out across the North of England.

2) Regional universities and public service organisations should work together to create a positive and inclusive network of R&D departments across the North of England.

3) Universities should be hardwired into programmes from inception so that investments can be properly evaluated and learning what does and doesn't work can be disseminated more effectively at a national level.



LIDA Case Studies

LIDA Academic Fellows in Data Science and Artificial Intelligence

To develop academic leaders of the future and enable the further growth and scale up of our successful communities and programmes, we will invest in LIDA academic fellowships – dedicated time for individuals to be fully immersed in LIDA’s interdisciplinary research environment.

Awarded through open competition, these prestigious fellowships are aimed at mid-career academics from any School or Faculty for a specific period of time to accelerate their academic excellence in Data Science / Artificial Intelligence and a leadership programme aligned to each of LIDA’s strategic priorities.

These fellowships will thus help the rapid development of early career academics’ leadership skills and interdisciplinary research outputs with global impact. They will further act as pump-priming and valuable experience for future large external grants and fellowships in Data Science and AI.

Scientific Machine Learning Group

Born from the LIDA: Environment community, the thriving Scientific Machine Learning (SciML) group continues to lead the way with regular group meetings, events with specialist external speakers (such as the British Antarctic Survey) and hands on training opportunities.

The group also provides learning materials for new ML researchers, including GPU access, coding tutorials, and best practices for running models.

The group has their own playlist within the LIDA YouTube channel, showcasing a wealth of talks, with the most popular on Physics-informed neural networks (PINNs) receiving over 13k views.



7

Education & Training



Challenges & Opportunities

Two-thirds of UK employers currently report difficulty recruiting workers with digital skills. The UK should offer and accredit accessible and valuable lifelong learning to help workers develop their careers and businesses to grow and increase their productivity.

Across the UK workforce, there is increasing demand for specialist data skills (including skills for artificial intelligence), which have the potential to bring economic and social benefits. According to Lightcast, the demand for AI skills has increased since 2014 in six G7 countries, as well as in Spain, Australia, Sweden, Switzerland, The Netherlands, Austria, Belgium and New Zealand.

Compared to other European countries, the UK has one of the highest demands for AI professionals. Internationally, demand is predicted to continue increasing. In 2023, the World Economic Forum surveyed 803 global companies and found that 'AI and Machine Learning Specialists' and 'Data Analysts and Scientists' roles were in the top 10 jobs expected to grow fastest between 2023 and 2027.

Evidence suggests that availability of people with specialist data skills in the UK is insufficient to meet demand. A 2021 study found that in the UK, the supply of data scientists from universities was unlikely to exceed 10,000 per year, yet there were potentially at least 178,000 unfilled data specialist roles.

In 2020, KRC Research consultancy surveyed 12,000 people working within enterprise companies (with more than 250 employees), including 600 in the UK. It found that 17% of UK employees were part of an upskilling process – compared to the 38% average reported globally. Respondents to the World Economic Forum's 2023 survey of global companies highlighted 'AI and big data' as one of the top three upskilling priorities.

To help address this gap we will increase the number of people with data skills by:

- **Academic pathway:** those with STEM degrees could complete a Masters or Doctorate (e.g. a PhD via one of LIDA's affiliated Centres for Doctoral Training). Non-stem students could enrol in a conversion course to learn a new discipline.
- **Upskilling pathway:** upskilling existing employees, from roles outside of data science, via the [Consumer Data Research Centre Annual Training Programme](#)
- **Partnership pathway:** where skilled individuals including international workers work in collaboration with senior researchers and / or external partners on real world research challenges via our Data Scientist Development Programme.

Strategic Aims & Objectives

LIDA's vision is to develop research-based knowledge and skills in data science and AI that secure successful outcomes for students, on-the-job learners and industry, making a positive impact in the world.

Recognizing the critical need for skilled professionals in data science and AI, LIDA is committed to provision and support for top-quality, fully inclusive, education and training programmes. Building upon its award-winning Data Scientist Development Programme (DSDP), and its newly approved fully online MSc in Data Science (Statistics), LIDA will continue to offer comprehensive training to students, early career researchers and data scientists, and industry professionals.

The fully online MSc in Data Science (Statistics) is a specialist master's, combining core data science disciplines with statistics specialisms created in collaboration with the School of Mathematics at the University of Leeds.

The DSDP is an opportunity for graduates and career changers looking to pursue a career in data science. It provides on-the-job training, mentoring and experience in the form of a fixed 12-month, paid role working with industry and third sector partners on social, health and environmental data challenges.

These programmes equip individuals with the necessary skills and knowledge to thrive in a data-driven world, fostering a new generation of data scientists and AI experts. The educators developing these programmes will be supported to innovate their pedagogy and practice as world leaders in their fields.



Strategic Aims & Objectives

Strategic Aims	Short term objectives (1 year)	Medium term objectives (2-3 years)	Long term objectives (4-5 years)
1. Make a positive difference to society	<p>A) Build on success of LIDA careers event for students and make annual (or twice yearly) and on a larger scale, e.g. full day or half day aligned to national initiatives such as Turing Skills.</p> <p>B) Support DSDP academics and partners to take forward data science outcomes from projects into next-stage through CDTs to embed partners across LIDA's E&T portfolio.</p>		<p>C) Undertake research and teaching relevant to current global challenges. Enable our scientists and students to make an important and measurable difference to society.</p>
2. Be a trusted global collaborator:	<p>A) Use our E&T structures to build on our success as a global collaborator. Continue with coordination of LIDA E&T Committee and Board, and support for DS&AI Education Local Interest Group and DSDP interdisciplinary Advisory Group (LIPAG).</p>	<p>B) Continue to foster industry-academic collaboration to develop multi-stakeholder research through the DSDP and taught programmes. Explore the potential to foster industry-academic PhD studentships and postdoctoral fellowships.</p>	<p>C) Work with RIS and Digital Education Service (DES) to enable key international collaborations.</p>
3. Develop, support & retain a diverse and inclusive education and training community	<p>A) Build on the success of the commissioned LITE fellowship "DS&AI Data Skills Literacy for Educators". LIDA to commission further fellowships through the Leeds Institute for Teaching Excellence (LITE).</p>		<p>B) Be recognised as a global innovator in data science education. Develop the next generation of educators and leaders in data science education.</p>
4. Increase successful outcomes for data science and AI students, and early career data scientists and learners	<p>A) Provide stewardship to AI and data science students and mentorship to data scientists on the DSDP so that they are equipped to thrive in the workforce. Coordinate and evaluate data science and AI educational activities so that students and learners understand and are prepared to undertake meaningful research that has societal impact.</p>	<p>B) Provide training and resourcing for Data Science Skillset development across disciplines (inc. non-STEM subjects). Develop materials that could be imported into many UG or TPG programmes (e.g. following the funding model employed by IDEA). Material to be offered, e.g. as a module for the Leeds International Summer School."</p> <p>C) Develop a fully-resourced and best practice mentoring model on the DSDP, resulting in increased data scientist employability and retention in the data workforce.</p>	

Strategic Aims & Objectives

Strategic Aims	Short term objectives (1 year)	Medium term objectives (2-3 years)	Long term objectives (4-5 years)
<p>5. Provide a healthy, safe & inclusive environment and enriching experience for staff and students</p>	<p>A) Use our E&T structures to build on our success as a global collaborator. Continue with coordination of LIDA E&T Committee and Board, and support for DS&AI Education Local Interest Group and LIPAG.</p>	<p>B) Be a University and national exemplar of inclusive practice in data science education. Grow our support for those at masters, doctoral and postdoctoral levels and on the DSDP.</p>	
<p>6. Diversify and increase data science and AI education and training income</p>	<p>A) Lead on the developments of fully online education following the successful launch of the fully online MSc Data Science (Statistics). LIDA to initiate and support further variants e.g. MSc Data Science (Health Informatics) or MSc Data Science (Environmental Analytics), sharing material that is already due for development.</p>	<p>B) Micro-credentials for CPD to be developed from the innovative material on our fully online programmes.</p>	
<p>7. Ensure a sustainable and balanced cohort mix in data science and AI education and training</p>	<p>A) Develop and expand OfS MSc studentships for underrepresented groups in DS&AI. Expand the number of MSc programmes and students included. Continue successful positive action recruitment on the DSDP.</p>	<p>B) Ensure support for underrepresented groups is self-sustaining in the long term. Expand student numbers on the pre-sessional access course developed as part of the OfS MSc scheme.</p> <p>C) Work with LIDA's educational partners to improve access to its DSDP for those coming from non-traditional educational pathways.</p>	

LIDA Case Studies

Data Scientist Development Programme

Our award-winning data science capacity-building programme meets data workforce needs by training graduates and career changers on-the-job via 12-month paid roles. Data scientists answer challenges that serve the public good, in collaboration with interdisciplinary academic and external partner teams.



Since it started in 2016 the programme has trained 73 data scientists through 126 projects in collaboration with 37 industry and third sector partners. 9 of these partner organisations and academic teams make up 34% of the total data scientists' next employers.

Data Science MSc programme

Our online Data Science (Statistics) masters degree offers students the opportunity to learn in-demand data skills such as data acquisition, data preparation, data wrangling, modelling and analysis, and how to deal with missing data.

The online programme welcomed its first cohort of 29 students in January 2024.



LIDA Case Studies

NHS Graduate Health Analysis Scheme Partnership

The NHS Graduate Health Analysis Scheme trains individuals to identify opportunities using data to improve patient experience, develop tools and assist in service provision planning.

From 2024 LIDA is offering Flexi-Placements of 8 weeks to those on the 2 year scheme to work with our researchers on real-world health data projects.



Led by the LIDA Health Community, we will contribute to the learning journey of these individuals who will go on to make a positive impact in society. This new collaboration incorporates LIDA's core values of collaboration, integrity and impact and will contribute to the delivery of our Education and Training strategic objective of being a trusted collaborator.

8

Research Technology



Challenges & Opportunities

Approximately half of LIDA's research is dependent on our state-of-the-art and highly secure Trusted Research Environment (TRE), known as LASER. It is the University's only ISO27001/NHS DSPT accredited computing platform and provides a cost-effective and reliable service across all faculties. However, LASER is currently limited to Windows virtual machines.

Ever-increasing requirements and expectations surrounding data governance mean in the foreseeable future, researchers will no longer be permitted to conduct some types of research with anonymised data on less secure platforms they use today. This is why LASER needs greater capabilities in Linux and machine learning, which are significantly different from the University's current Research IT investments.

A key limiting factor in some research and impact is the availability of expertise in software engineering. We see that as an opportunity, which we are addressing by continued investment in LIDA's Data Analytics Team, and expanding capacity through new initiatives with staff and PhD students in schools such as Computing and Maths. That will simultaneously open doors for those staff and students to broaden their knowledge and experience.

The move to using cloud computing platforms for research presents an opportunity to increase research recoveries and reduce overheads. LIDA has already captured fine-grained cloud usage data for more than 70 LASER projects. The challenge is to develop and validate a cost prediction model from that data, which will be unique in the university sector.



Strategic Aims & Objectives

LIDA's vision is to provide access to secure, innovative, research-informed data science and AI infrastructures, and academic and research software engineering workforce expertise, that enhance user experience and research and educational outcomes.

LIDA recognises the crucial role of research technology in enabling transformative data-driven and AI research. By supporting researchers to identify and gather user requirements LIDA provides a centre of expertise for the development of advanced data science and AI tools and platforms, accelerating discoveries and enhancing collaborations across disciplines.

To this end, LIDA has a specialist team of Research Software Engineers in the Data Analytics Team (DAT) that can manage research requiring the development of both bespoke technical solutions and/or use of LASER. This cutting-edge, nationally accredited infrastructure empowers researchers with secure access to sensitive and confidential data, facilitating innovative research while safeguarding data privacy and security.

Over the next 5 years LIDA aims to build on this unique capability by:

- Increasing the capacity and capability of academically-led and experienced research software engineers embedded into LIDA, to support its Communities and Programmes deliver and implement data science and AI.
- Increasing the availability of hybrid and scalable AI computing and storage, to enable deep learning (GPU) and simulations (CPU) with large & fast storage.
- Placing user requirements and researcher user experience (UX) at the fore, to drive improvements in infrastructure, software and computational methods for data science, visualization and AI.



Strategic Aims & Objectives

Strategic Aims	Short term objectives (1 year)	Medium term objectives (2-3 years)	Long term objectives (4-5 years)
<p>1. Provide high end, scalable research computing infrastructure</p>	<p>A) Enable researchers to safely and securely self manage and audit the movement and transfer of their data in and out of LASER, to improve research productivity and efficiency</p> <p>B) Collaborate with UoL IT to enhance LASER support and maintenance services and improve bug and issue resolution response times</p>	<p>C) Provide high-end, scalable AI/ML research computing (e.g., based on LEARN).</p> <p>D) Explore access to LASER by masters students for dissertation research.</p>	<p>E) Ensure LIDA continues to keep pace with, and achieves sustainability with all new and relevant innovations in technologies in support of LIDA's stated vision.</p> <p>F) Provide Linux capability with LASER, to allow use of codebases common in maths, computer science and genomics and enable a step-change in researcher engagement from those domains.</p> <p>G) Prioritise user requirements and researcher user experience, to drive improvements in infrastructure, software and computational methods for data science, visualization and AI.</p>
<p>2. Develop processes and tools to improve costing and research recoveries</p>	<p>A) Enhance LASER costing model. Validate the current model, using the results to make it more fine-grained and accurate.</p> <p>B) Implement process for use of LIDA infrastructure by external collaborators, that do not have academic sponsorship within UoL.</p>	<p>C) Develop a data-driven model to optimise costs of HPC and AI/ML research in cloud-based environments.</p> <p>D) Improve transparency and availability of LASER research costs, including guidance on engaging with Schools/Institutes</p>	<p>E) Share or exploit the cost model to increase LIDA's impact.</p>
<p>3. Encourage open research practice</p>	<p>A) Establish process for the open sharing of software, reproducible workflows and analysis methods.</p> <p>B) Provide training to support rigour in open sharing.</p>	<p>C) Establish an approach for Open/FAIR sharing of data from sensitive research.</p>	<p>D) Develop a portfolio of LIDA-branded open software/ workflows/ methods for reproducible and rigorous data science.</p> <p>E) Advance fundamental and applied data science disciplines through promotion of LIDA-branded and associated software/workflows/methods.</p>

Strategic Aims & Objectives

Strategic Aims	Short term objectives (1 year)	Medium term objectives (2-3 years)	Long term objectives (4-5 years)
<p>4. Deliver impactful research and innovation</p>	<p>A) Identify gaps and opportunities in skills/training by mapping usage of data science methods, expertise and requirements.</p> <p>B) Develop knowledge and understanding of existing infrastructure and skills available for research through LIDA, among internal and external stakeholders.</p>	<p>B) Increase access to expertise on data science, visualization and AI methods.</p> <p>C) Develop and grow LIDA's Data Analytics Team capability (people, processes and tools) and capacity to provide best in class RSE expertise to new and diverse areas of income generating data science and AI research activity, both with internal and external partners.</p>	<p>D) Open up leading-edge technologies and methods to researchers for the ever-increasing scale and complexity of their data.</p> <p>E) Researchers routinely adopt methods that are suited to large-scale data.</p> <p>F) Provide guidance about productionising new data science methods, supporting early-adopters and attracting users.</p>
<p>5. Be a trusted collaborator</p>	<p>A) Establish a sector-leading reputation for the operational delivery and security (ISO 27001 and DSPT) of data science and AI IT infrastructure.</p> <p>B) Found an exchange programme for data science professionals in leading public & private sector organisations.</p>	<p>C) Develop relationships with technology partners to provide underpinning tools & technologies.</p> <p>D) Provide multidisciplinary guidance and support about use of AI technology in research</p>	<p>D) Develop a recognised role in leading and shaping the data ecosystem in the city and region.</p> <p>E) Maintain our reputation, standards and accreditation for data security and information governance.</p>

LIDA Case Studies

The Data Analytics Team (DAT) is a group of specialists in data management, data analysis and software engineering, who collaborate with LIDA researchers across all stages of their projects.

The team has a central role in LIDA, managing the secure research platforms, collaborating on research projects, and offering data analytics and, in collaboration with Information Governance, information security expertise to researchers.

You can find out more about the work of the DAT in the case studies below and their full skill set in the annexe.

Safe Pod Network

Broadening accessibility of collaborative research

Research at the University, that uses the most sensitive data, is normally restricted to Safe Rooms within LIDA offices.

We have worked with Safe Pod Network to broaden access to this data in LASER, our trusted research environment.

Using purpose-built 'Safe Pods', such sensitive research is now accessible from 23+ other locations across the UK, whilst maintaining the level of security required. We aim to build on this for researchers to collaborate more between the University and organisations across the UK; which is otherwise restricted by data security requirements.



LIDA Case Studies

FIND-AF API & Web App

Putting academic results to public use for societal benefit

We worked with health researchers to ensure their predictive model can be used by doctors and nurses, to benefit public health.

The DAT are developing an API with a web app interface so that doctors and nurses can log on, enter patient details, and receive a risk prediction of irregularities of the heart of their patients to prevent stroke.

We aim to expand on this type of work with other research groups in LIDA to continue making research outputs publically accessible.



Natural Language Processing

Expanding capabilities of LASER

We worked with IT Services to enable native Azure resources to be deployed to research projects in LASER. Initially this focussed on Azure Cognitive Services, an AI for data scientists.

Applying Cognitive Services to research in LASER, the DAT also provide a service for transcribing and de-identifying audio and video files at scale, without transferring to a third party organisation. This reduces risk footprint and costs and we aim to expand on Cognitive Services to explore its potential and other resources available.



9

Operational Enablement



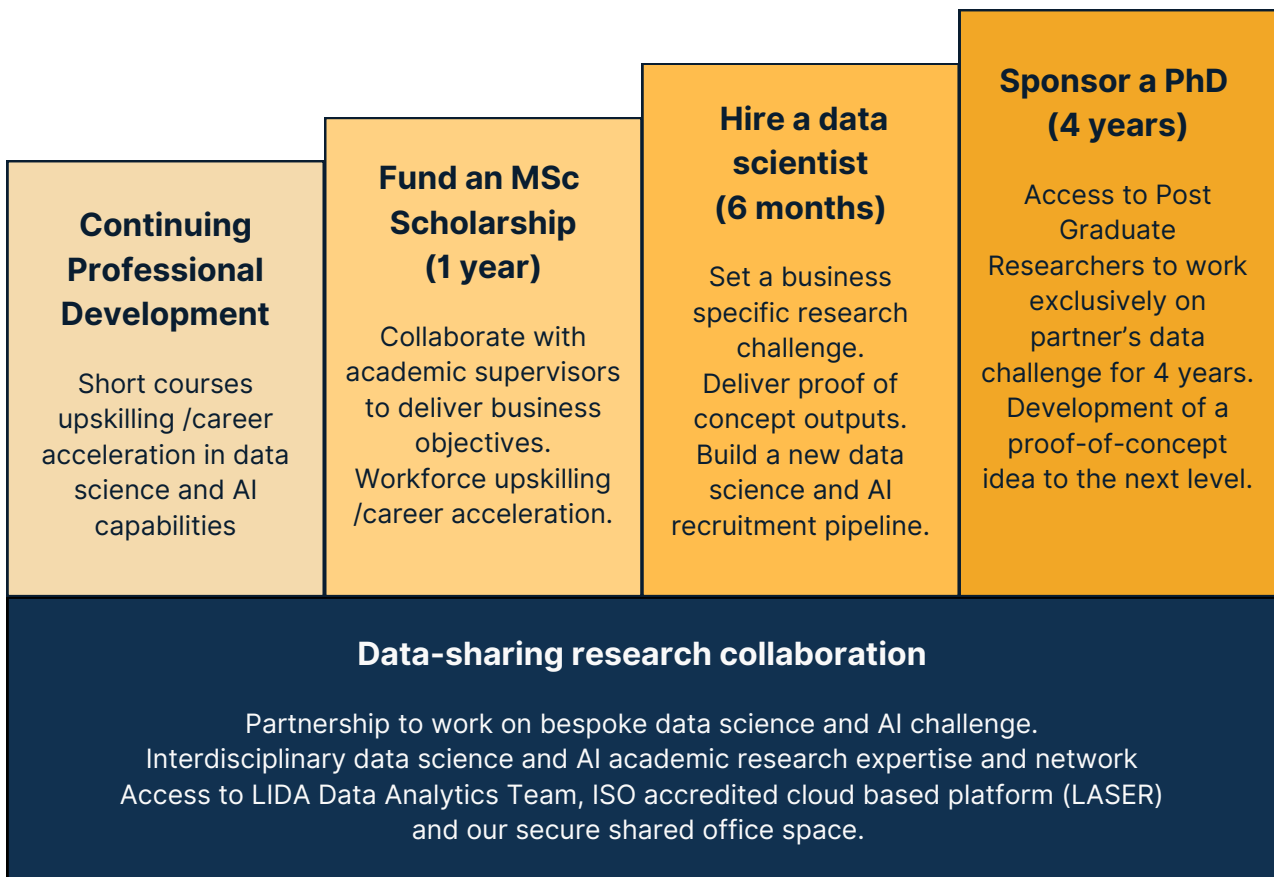
Strategic Aims & Objectives

LIDA's vision is to be academically led, but operationally enabled to be interdisciplinary, diverse and inclusive, with a positive working culture that is financially and environmentally sustainable.

LIDA understands the importance of operational enablement in effectively executing its academic data science and AI strategy. By fostering a collaborative and digitally supported on-campus and hybrid environment, nurturing partnerships with academic, industry and government entities and streamlining administrative processes, LIDA aims to create a supportive data science and AI environment that encourages innovation and knowledge exchange.

LIDA's 5 year strategy supports additional investment in our people, academic communities, platforms and processes whilst also achieving a balanced budget in 2029.

LIDA service benefits for partners



Strategic Aims & Objectives

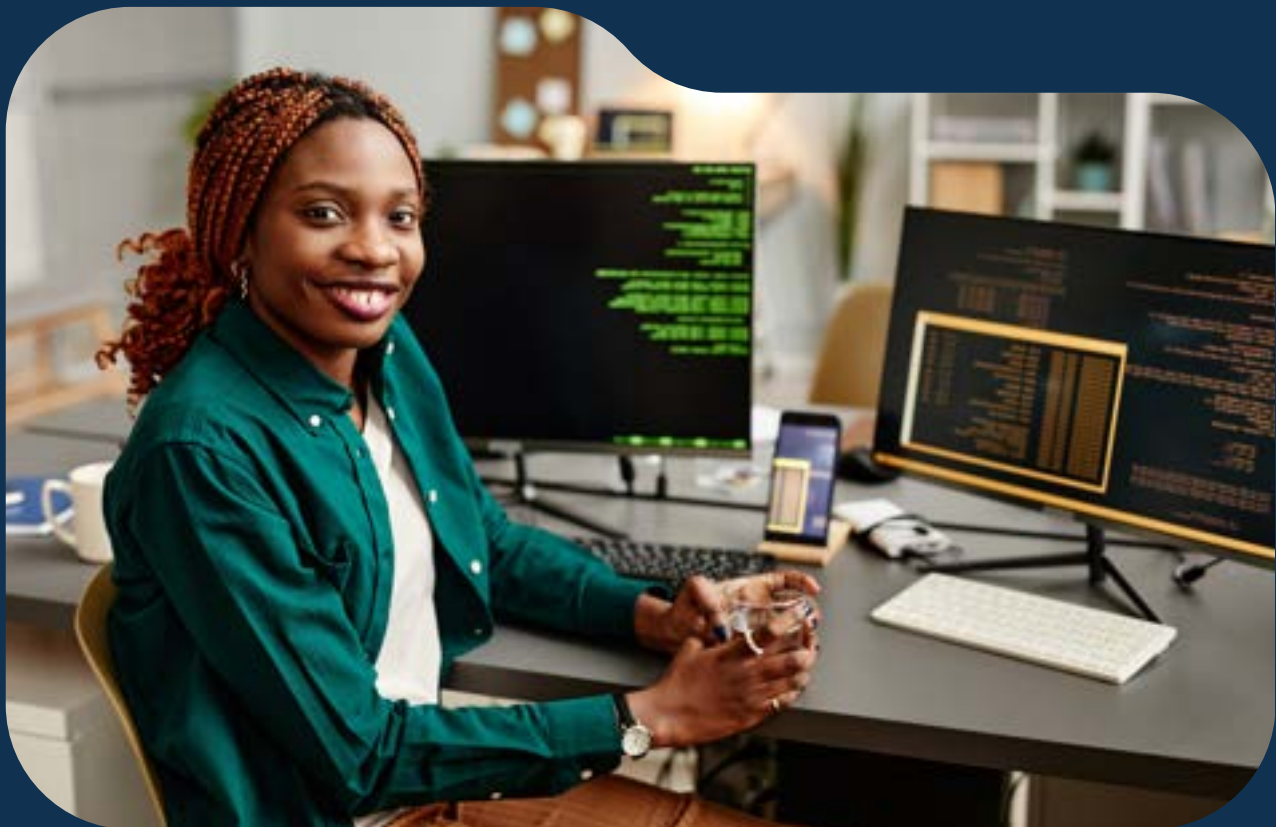
Strategic Aims	Short term objectives (1 year)	Medium term objectives (2-3 years)	Long term objectives (4-5 years)
<p>1. Developing, supporting & retaining a diverse and inclusive professional services staff community</p>	<p>A) Ensuring that LIDA staff engagement and the effective, transparent and mutual flow of information and ideas throughout the team, is at the centre of our culture.</p> <p>B) Ensuring all staff have at least 20% of their time allocated to training and development in their role, supported by an appropriate training budget and access to opportunities.</p>	<p>C) Professional data scientist and Technical Staff career development is supported through the provision of clear University career pathways, development and training and opportunities for career progression.</p> <p>D) Mapping and regularly reviewing EDI data and creating a forum for sharing expertise that will underpin and promote equity, diversity and inclusion in LIDA's Team and its Data Science and AI communities.</p>	<p>E) LIDA develops, grows and retains a diverse workforce with a wide skills mix in both academic, technical business and managerial data science and AI specialisms.</p>
<p>2. LIDA's facilities provide a best-in-class collaborative work environment</p>	<p>A) Develop the LIDA environment business management support capability to enhance the working environment and research and student experience.</p> <p>B) Enhance provision of highly secure, restricted working environments, through the SafePod network and open plan Safe Space to ensure the appropriate balance between data security and wellbeing .</p>	<p>C) Optimise the access to physical space that meets all of LIDA's communities' requirements, by developing an appropriate blend of independent, shared and collaborative work spaces supported by easy-to-use booking technology.</p>	<p>D) Provide new digital technologies, environments and resources in LIDA that will encourage and enhance on-campus work, remote/hybrid collaborative research capabilities and student learning and training experience.</p>

Strategic Aims & Objectives

Strategic Aims	Short term objectives (1 year)	Medium term objectives (2-3 years)	Long term objectives (4-5 years)
3. LIDA is a trusted international collaborator	A) LIDA has a sector leading reputation, internally and externally, for the operational delivery and security (ISO 27001 and DSPT) of its data science and AI IT infrastructure and data.	B) Implement a LIDA data science communications and engagement strategy across all LIDA portfolios that builds a clear brand, academic network and public and business partnership model, that fosters a sense of belonging amongst staff and all of our communities and generates strong partnerships with external partners.	C) In partnership with RIS, embed new interdisciplinary and collaborative external partnerships that support LIDA's academic goals resulting in new and diverse areas of income generating data science and AI research activity.
4. LIDA is adaptive, effective and financially and environmentally sustainable	A) LIDA has access to insightful and integrated management information across academic (KRISTAL, SIMPLECTIC, IRIS), finance, IT and KPI performance to inform decision making.	B) Continuous improvement of transparent and accountable governance structures, particularly in relation to EDI, for both LIDA and its Communities, making the LIDA Executive responsible for achieving change.	C) LIDA invests its annual surplus, whilst maintaining a balanced budget, over the period of the 5 year IPE in strategic priorities identified by Faculties and the LIDA Community.

10

Equity, Diversity & Inclusion



Challenges & Opportunities

Research finds that certain groups (such as women, those from minority ethnic backgrounds and people with disabilities) are underrepresented in the data workforce. A lack of workforce diversity has the potential to amplify existing inequalities and prejudices.

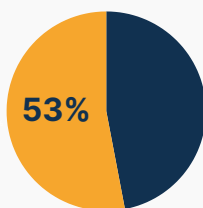
Data about the demographics of the data science workforce is limited, but studies indicate that there are still significant disparities.

Gender



24% of employees in AI roles are women

53% of organisations have no female employees in AI roles



21%

of students taking a GCSE in computing are girls

Of higher education students, women comprise

23%

on computing courses

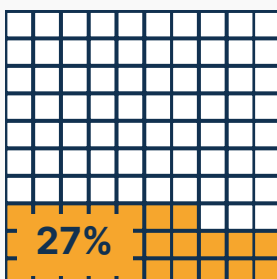
20%

on engineering courses



Women in AI and data roles are more likely to occupy a job with lower status and pay, despite having higher formal education levels than men.

Ethnicity



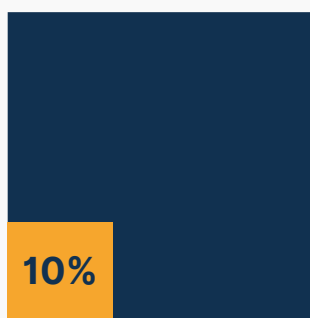
Minority ethnic employees make up 27% of the AI workforce

40%

of firms have no ethnic minority employees in their AI teams

Disability

There is a lack of data about people with disabilities in the data science workforce, but evidence exists for IT specialists.



Of the 158,000 IT specialists in the UK, 10% have a disability. This is lower than the proportion of the UK workforce with disabilities, which was approximately 14%.

Strategic aims & objectives

LIDA will contribute to the University of Leeds' strategy to eliminate gaps between groups with protected characteristics, and without, by focussing on the following aims and objectives:



Development, progression, pay and retention rates across protected characteristics (e.g. race, gender, disability, sexual orientation, age)

- We will aim to improve gender balance in grade 9 and 10 professional services and academic roles through proven mechanisms such as positive action recruitment.
- Increase appointment rate between ethnic minority interviewed candidates and white interviewed candidates, for LIDA managerial roles and programmes, including the DSDP and OfS Artificial Intelligence and Data Science Postgraduate Conversion Course Scholarship

Increased engagement of LIDA managerial and academic research, education and training community staff with leadership development, to drive forward LIDA's strategy, culture and values



- We will aim to increase the leadership development of both managerial staff and our academic community by improving the:
 - % of all Black, Asian and Minority Ethnic LIDA staff actively engaged in recent leadership development programmes
 - % of LIDA staff from grade 6-10 actively engaged in recent leadership development activities
 - % of all LIDA staff who have access to the academic data science and AI learning and development required to do their job well.

Improve the socio – economic diversity of our research and education communities



- We will improve the socio-economic diversity of our research community through the recruitment of early career researchers in to LIDA research programmes
- We will increase community outreach and engagement activity to reduce early barriers to equitable data science recruitment for local young people, such as low aspiration, socio-economic drivers and stereotypes of what a data scientist looks like/ought to be.
- Increase opportunities for post-graduates to undertake data science and AI masters through providing scholarships for underrepresented groups.

LIDA Case Studies

Diverse recruitment of Data Scientists

Black people are underrepresented in data science and have made up only 4% of data scientists on the DSDP between 2016 and 2020.

In 2021, LIDA worked with HDR UK to introduce positive action recruitment to the DSDP, by reserving a post exclusively for black data scientists.



The result was a 33% conversion of black applicants to data scientists in 2021. Equitable recruitment has continued since and increased numbers of Global Majority (GM) appointees to 44% in 2021-22 and 50% in 2022-23. Targeting one underrepresented group also stimulates applications from other underrepresented groups e.g. the number of women also rose from 44% in 2021-22 to 50% in 2022-23.

Taking data science to schools in disadvantaged areas

The LODSS widening participation initiative sees LIDA data scientists working with young people in local schools which identify strongly for deprivation markers such as low aspiration, food insecurity and lack of positive role models.

The project looks to inspire, engage and develop their digital skills e.g. via code clubs in schools.



Following the pilot, pupils were shown to be inspired and engaged. Survey responses included, "Data science can improve the world"; 70% of pupils said they were inspired by the events.

LIDA Case Studies

MSc scholarships for underrepresented groups

In 2023, we awarded 19 scholarships for underrepresented groups to study data science-related Masters courses.

The new scheme aims to encourage more women, black students, disabled students and those from lower socioeconomic backgrounds into data science and AI industries.



Results of first year of applications - 6 of the 9 underrepresented groups identified by OfS were represented by our applicants at shortlisting stage and 5 of the 9 groups were represented at award stage. In 2024, we are repeating the process with 23 anticipated scholarships to award.

11

Environmental Sustainability



Strategic Aims & Objectives

The University's Climate Plan represents the single biggest investment the University has ever made: £174 million over the next decade, more than £150 million of which will be put towards achieving our goal of net zero greenhouse gas emissions by 2030.

How LIDA will contribute to the University's **Net Zero** targets by 2030:



Reduce the primary energy demand of our building and estate.

Building retrofit (including fabric measures), energy efficient lighting, enhanced building control and estate rationalisation.



Reduce emissions from business travel, including:

- Continue to support and improve hybrid working
- Support high quality digital research, education and training
- Reduce our travel footprint and provide new online-only degrees
- Partner opportunities in HE sector for more off-setting projects



Decarbonise energy usage related to LIDA's use of gas and electricity in our building and use of compute infrastructures

How LIDA will contribute to the University's **Net Zero Plus** targets by 2030:



Supplier contracts: communicate our intention to become net zero by 2030 to our suppliers and ask them to join us on our journey.



IT hardware: Global technology companies are increasingly setting science-based emission targets, including those product-based. We will work with UoL IT and Procurement to ensure minimum standards for IT hardware that tracks industry best practice.



Food and catering: A large proportion of catering emissions are related to meat and dairy consumption. We will ensure careful menu planning across delivered food, conference and event catering to help to reduce this and unnecessarily wasted food.

LIDA Case Studies

“

Sustainability will be key to all our activities, though we acknowledge the challenges in terms of the global reach of our work and the technical infrastructure needed to support it.

At the heart of the Leeds Student Education Strategy is the promotion of “environmental and social sustainability through our portfolio, our actions and student opportunities”.

Through LIDA’s E&T strategy we will ensure that ethics and ethical AI are central to all of our activities, ensuring we train all future data scientists to work in an environmentally sustainable and socially responsible way.

Professor Paul D. Baxter
Director of LIDA Education & Training

”

LIDA office estate reduced by 1/3

Following the covid-19 pandemic and to support colleagues’ preference for hybrid working, LIDA reduced its office space, equipment and energy footprint by a third. LIDA continues to monitor use of its office space on a weekly basis to ensure space occupancy and resource utilisation remain optimal and sustainable for data science.





Leeds Institute for Data Analytics
University of Leeds

LIDA