ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS110

# ADVANCING ACCELERATOR SCIENCE THROUGH DATA-INTENSIVE RESEARCH AND TRAINING*

C. P. Welsch†, [1], Cockcroft Institute, Warrington, UK
[1]also at University of Liverpool, Liverpool, UK

*Abstract*

The Liverpool Centre for Doctoral Training in Innovation in Data Intensive Science (LIV.INNO) has made significant progress in applying data-intensive methods to accelerator research. This contribution presents research outcomes from the center with a focus on two key projects.

The first focuses on optimizing 3D imaging for medical and industrial applications, integrating Monte Carlo simulations and advanced collimation techniques to enhance low-dose, portable X-ray systems, with implications for wider accelerator diagnostics. The second lever-ages deep learning models to reconstruct transverse beam distributions at CERN, addressing challenges in image distortion from multimode optical fibers under high-radiation conditions. The results are connected with wider progress made in machine learning and artificial intelligence for particle accelerators. Furthermore, the paper summarizes the outcomes of several key LIV.INNO events: the STFC Summer School on Data Intensive Science, the LIV.INNO 2024 Industry Showcase and the 2025 AI for Innovation Summit.

## INTRODUCTION

The Liverpool Centre for Doctoral Training for Innovation in Data Intensive Science (LIV.INNO) was established in 2022 as a strategic response to the evolving landscape of data-driven research. Funded by the UK Science and Technology Facilities Council (STFC), LIV.INNO is jointly hosted by the University of Liverpool and Liverpool John Moores University, forming a collaborative and inclusive academic environment in the North West of England. The center is structured to deliver interdisciplinary training to three cohorts of PhD students between 2022 and 2024, each comprising approximately twelve researchers per year [1].

LIV.INNO addresses the data challenges emerging across a broad spectrum of STFC-supported domains, including nuclear and particle physics, accelerator science, astrophysics, mathematics, and computer science. As a hub for training the next generation of scientific leaders, the center integrates cutting-edge research with a comprehensive academic program in data intensive science. This includes foundational and advanced training in Monte Carlo simulations, high-performance computing, artificial intelligence, machine learning, and advanced data analysis methodologies. Students are embedded in interdisciplinary projects designed to harness the large, complex datasets that define modern scientific inquiry.

LIV.INNO builds upon the success of its predecessor, the LIV.DAT Centre for Doctoral Training, which operated between 2017 and 2023, and successfully trained approximately 40 PhD students [2]. Both centers leverage over a decade of experience from pan-European training networks, including several Marie Skłodowska-Curie Actions networks in accelerator science. These initiatives have trained over 100 early-career researchers and have significantly shaped postgraduate training methodologies in the UK and across Europe.

## RESEARCH

The rapid growth of data in science and engineering, driven by advances in sensors, mobile devices, biotechnology, and digital technologies, has created major challenges in managing, analyzing, and interpreting large, complex datasets. Despite the clear demand, targeted training in data intensive science remains limited.

LIV.INNO directly addresses this skills gap through a structured research programme built around three core scientific Work Packages (WPs):

- **WP1 – Monte Carlo Methods and High Performance Computing (HPC):** This WP trains students in advanced computational techniques essential for simulating complex systems, from cosmic evolution to subatomic interactions. Monte Carlo tools and HPC are applied across fields like astrophysics, particle physics, and engineering.
- **WP2 – Artificial Intelligence (AI) and Machine Learning (ML):** Students develop and apply ML algorithms, including deep learning methods on GPUs, to extract insights from high-dimensional data. These tools support cutting-edge research in fields such as detector physics, medical imaging, and galaxy modeling.
- **WP3 – Data Analysis:** This WP focuses on building efficient and scalable analysis pipelines to process multimodal datasets. Projects often integrate outputs from simulation (WP1) and AI models (WP2), with applications spanning from core physics to healthcare and industry.

These areas are highly relevant to accelerator science, where improvements in beam diagnostics, control, and dynamics rely heavily on sophisticated data techniques. LIV.INNO researchers contribute to key initiatives at CERN and beyond, applying data science to advance both experimental precision and innovation. Two examples from current research projects are given here.

* This work was supported by the Science and Technology Facilities Council (STFC) under grant agreement ST/W006766/1.
† carsten.welsch@cockcroft.ac.uk

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS110



### *3D X-ray Scanning Techniques for Application in Medical Imaging*

The project of Lauryn Eley explores the development of a portable 3D imaging modality for chest diagnostics using digital tomosynthesis (DT). DT is a cost-effective, lower-dose alternative to computed tomography (CT), relying on the acquisition of multiple projection images over a limited angular range to reconstruct volumetric data. While DT has lower depth resolution than CT, it offers faster acquisition and significantly reduced radiation exposure, making it suitable for use in settings where conventional CT is unavailable or impractical.

This research is carried out with Adaptix Ltd., a UK-based multi-award-winning medical imaging company that has developed a novel flat-panel field emission source array [3]. The technology eliminates the need for a rotating anode x-ray tube, enabling compact, low-power, and portable systems. Adaptix has successfully commercialized this technology for use in orthopedic and veterinary imaging. However, its potential for clinical human applications, particularly chest imaging, remains an open challenge [4].

The primary aim of this research is to support the technical upscaling of Adaptix' existing platform to develop a clinically viable 3D x-ray system capable of imaging the human thorax. To this end, detailed Monte Carlo simulations using the Geant4 particle transport toolkit are employed to create a high-fidelity digital twin of the imaging system.

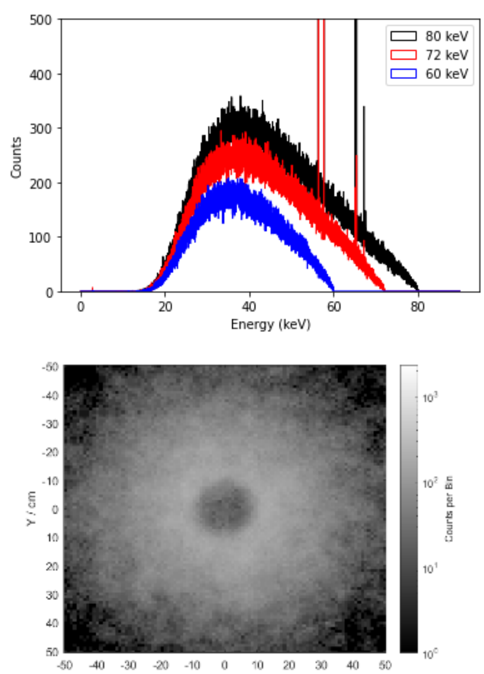


Figure 1: Examples from the OptiX simulation environment: x-ray spectra (top) and simulated object shadows, analogous to x-ray images (bottom).

This virtual environment enables systematic investigation of critical engineering and imaging parameters such as anode target thickness, spectral shaping via filtration, and spatial characteristics of the emitted x-ray beam - all without the time and expense of physical prototyping. In the current phase of the project, the simulation results are being benchmarked against experimental data. Some simulation results are shown in Fig. 1. A proof-of-concept experimental platform has been assembled, integrating a robotic positioning system that enables precise control of the source-detector geometry. This allows for the emulation of motion necessary to collect the angular projections required for tomosynthesis image reconstruction over a field-of-view large enough for human chest imaging.

This work helps Adaptix in refining their existing hardware and addresses broader research questions in computational medical imaging and x-ray system optimization. The ultimate goal is to demonstrate that DT, implemented using a flat-panel field emission source and enabled by advanced modeling, can serve as a low-cost, portable alternative to CT in a variety of clinical contexts, especially in point-of-care or resource-limited environments.

### *Reconstruction of Transverse Beam Distribution using Machine Learning*

Transverse beam profile measurements are critical for the operation and optimization of modern particle accelerators. Traditionally, such diagnostics are performed by inserting a scintillating screen into the beam path and imaging the resulting light distribution using a camera. However, in high-radiation environments such as those found at CERN, prolonged exposure degrades camera sensors, leading to reliability issues and hardware failure. Until recently, radiation-hardened tube cameras addressed this challenge, but their global discontinuation has necessitated the development of alternative imaging solutions.

The project of Qiyuan Xu investigates a novel beam diagnostic approach that relocates vulnerable imaging electronics to radiation-safe areas. The proposed method uses a multimode optical fiber (MMF) to transmit the light signal produced at the scintillating screen to a remote CMOS camera [5, 6]. While this setup protects sensitive components, it introduces complex imaging challenges. Specifically, intermodal dispersion and mode coupling within the fiber transform the original beam profile into a scrambled speckle pattern, complicating the task of image reconstruction.

To overcome this, a deep learning-based approach has been developed that leverages encoder-decoder neural network architectures to reconstruct the beam profile from the fiber output. Unlike conventional inversion techniques such as phase conjugation or transmission matrices, this method employs fully data-driven models, trained entirely on synthetic datasets generated via Stochastic Gaussian Mixture simulation. A digital micromirror device (DMD),

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS110

illuminated by a laser, is used to emulate the scintillating screen and produce both training and test patterns.

Crucially, the model exhibits strong generalization to real-world beam samples, despite having seen only simulated data during training. This robustness is illustrated in Fig. 2, which shows a reconstructed transverse beam distribution based on the speckle output at the fiber end. The result highlights the potential of this method for non-invasive, radiation-resilient beam diagnostics.

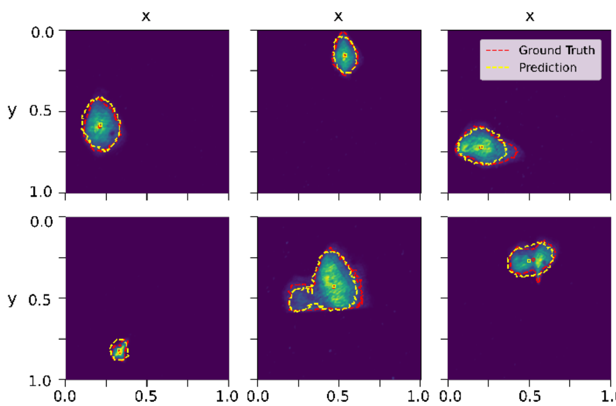

Figure 2: Reconstructed transverse beam profile from multimode fiber output using a machine learning model trained on synthetic data.

To ensure operational viability, the system's sensitivity to environmental perturbations—such as temperature shifts, mechanical vibrations, and radiation-induced changes in fiber refractive index—is being studied. These factors can alter the transmission characteristics of the fiber, impacting the model's reconstruction accuracy. Accordingly, ongoing research involves building a more comprehensive dataset that captures these variations. Experimental setups are being developed at both the Cockcroft Institute and CERN's CLEAR facility to evaluate performance under realistic operational conditions. Future work will focus on integrating more representative scintillation light sources, improving robustness through transfer learning techniques, and optimizing the system for long-term deployment in operational beamlines. The ultimate aim is to deliver a reliable, AI-enhanced diagnostic tool that extends the reach of precision beam imaging into environments previously inaccessible due to radiation constraints.

## TRAINING

LIV.INNO delivers a comprehensive, research-led training program designed to prepare the next generation of leaders in data intensive science. Central to its mission is the integration of academic excellence with real-world impact, achieved through a unique blend of interdisciplinary education, industrial engagement, and international collaboration.

A key element of the LIV.INNO model is its mandatory 3- 6 month industry or public sector placement, where each PhD student applies their expertise to challenges beyond their core research project. These placements span diverse sectors, including healthcare, transportation, cybersecurity, and digital innovation, and expose students to real-world data problems. Past projects have involved AI-driven medical diagnostics, safety tools for rail operations, surrogate modeling for epidemiology, and commercial natural language processing applications

The center's structured training pathway begins with a first-year core program requiring 120 credits of coursework in data science, delivered jointly by the University of Liverpool and Liverpool John Moores University, with support from external industry experts. From years two to four, training becomes increasingly tailored through Development Needs Analysis (DNA) and a structured Career Development Plan (CDP), aligning further learning with individual research trajectories and career goals. Training is delivered through a rich mix of lectures, seminars, workshops, and schools, and is enhanced by international data science events and secondments. Notably, students benefit from resources such as the LIV.HUB, a multi-functional training and collaboration space, and are encouraged to participate in outreach and public engagement.

In addition to technical instruction, LIV.INNO places strong emphasis on transferable skills, including project management, scientific communication, presentation, and networking - core competencies for success in both academia and industry. The center has also partnered with the EuPRAXIA Doctoral Network to offer joint training activities, such as the introductory skills school [7].

LIV.INNO also runs an open data science seminar series that features prominent voices across academia and industry, and it hosted the STFC Summer School on Data Science in July 2024 [8], as well as an AI for Innovation online summit in 2025 [9]. A workshop on data science in healthcare and health technologies is planned on 25 June 2025 [10].

Through this multi-faceted training model, LIV.INNO equips its doctoral researchers with advanced scientific expertise, and with the agility, experience, and perspective to thrive across a wide range of data-intensive careers.

## SUMMARY AND OUTLOOK

LIV.INNO trains PhD researchers in data-intensive science with applications in accelerator science, medical imaging, nuclear and particle physics. Building on the LIV.DAT program, it combines a structured training program with interdisciplinary research. This paper presented results from two key projects to illustrate applications in accelerator science. The training integrates academic coursework, personalized development plans, and industry placements, while also hosting major events such as the 2024 STFC Summer School and the 2025 AI for Innovation Summit. Looking ahead, LIV.INNO will expand its impact through experimental validation, broader industry collaboration, and continued leadership in international scientific training.

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPS110

## REFERENCES

[1] LIV.INNO home page. http://www.livinno.org

[2] C. P. Welsch, "Using Data Intensive Science for Accelerator Optimization", in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 980-982. doi:10.18429/JACoW-IPAC2022-TUPOST051

[3] Adaptix, "home page". http://www.adaptix.com

[4] T. G. Primidis, *et al.*, "3D chest tomosynthesis using a stationary flat panel source array and a stationary detector: a Monte Carlo proof of concept", *Biomed. Phys. Eng. Express* 8, 2021. doi:10.1088/2057-1976/ac3880

[5] G. Trad and S. Burger, "Artificial Intelligence-Assisted Beam Distribution Imaging Using a Single Multimode Fiber at CERN", in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 339-342. doi:10.18429/JACoW-IPAC2022-MOPOPT041

[6] Q. Xu, A. Hill, H.D. Zhang, C.P. Welsch, F. Roncarolo, and G. Trad, "Estimation of Beam Transverse Parameters Through a Multimode Fiber Using Deep Learning," *in Proc. IBIC'24*, Beijing, China, Sep. 2024, pp. 170–173. doi:10.18429/JACoW-IBIC2024-TUP54

[7] LIV.INNO, "Introductory Skills School". https://indico.ph.liv.ac.uk/event/1839/

[8] LIV.INNO, "STFC School on Data Intensive Science 2024". https://indico.ph.liv.ac.uk/event/1639/

[9] LIV.INNO, "AI for Innovation Summit". https://bit.ly/ai-for-innovation-summit

[10] LIV.INNO, "Data Science in Healthcare and Health TechnologiesWorkshop". https://indico.ph.liv.ac.uk/event/1999/