

Skills gap for AI-enabled research: What is available and what must be done?

eResearch Australasia 2024

Acknowledgement of Country



*I begin by acknowledging the Traditional Owners of Country throughout Australia and pay my respects to their Elders past and present.
I extend that respect to Aboriginal and Torres Strait Islander people here today.*

Recap



Past engagements

- eResearch Australia and NZ conferences in Oct 2023 and February 2024
- Supercomputing Asia conference in March 2024

Common comments

- Community needs hands-on training on basics of Machine Learning, including PyTorch, TensorFlow, and other ML Libraries.
- Difficulty in finding the available training resources

Agenda



- Update from ML4AU stakeholders
 - What's currently happening
 - Existing skills development resources
 - 6 to 12 months plan
- Audience Participation and Panel Discussion
 - Identify the knowledge gaps: Skills to equip researchers to tackle ever-growing AI capabilities and requirements.
- Wrap up

Audience Participation



slido.com

#2174434

What do you think are the skills and knowledge gaps in AI-enabled research?

Updates from the Stakeholders

What's currently happening

Existing skills development resources

6 to 12 months plan

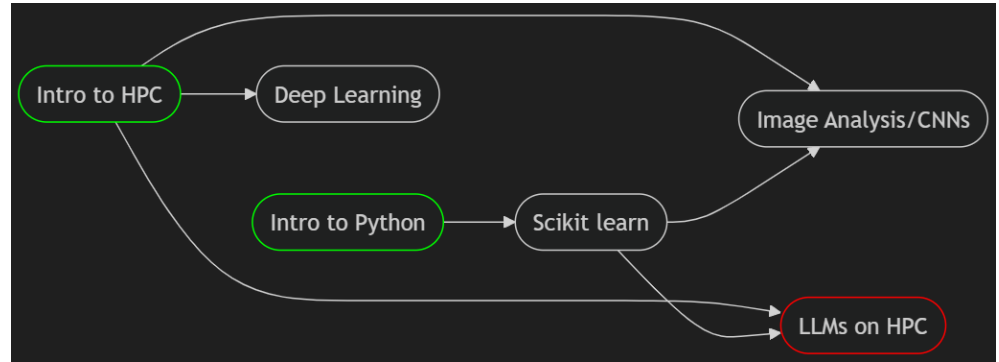
New training manager (Soon™)

Algorithm uplift, Alphafold etc

Quantum simulation and shenanigans



- A learning pathway:



- Improving our scope:

- Understanding AI and biases, FAIR / CARE, ethics, beyond the technical
- The “I” in FAIR bringing in relevant, interoperable datasets, be more engaging

- Next step:

- Machine Learning 103: Getting Started with Large Language Models on NeSI

Intersect's journey to ML/AI training

Intersect
launched
the Digital
Research
Skills
Training
Program

Released the
series of 3
Python and 3
R Machine
Learning
courses

Release of
AI
Bootcamps

Exploring
more ...

Jan
2012

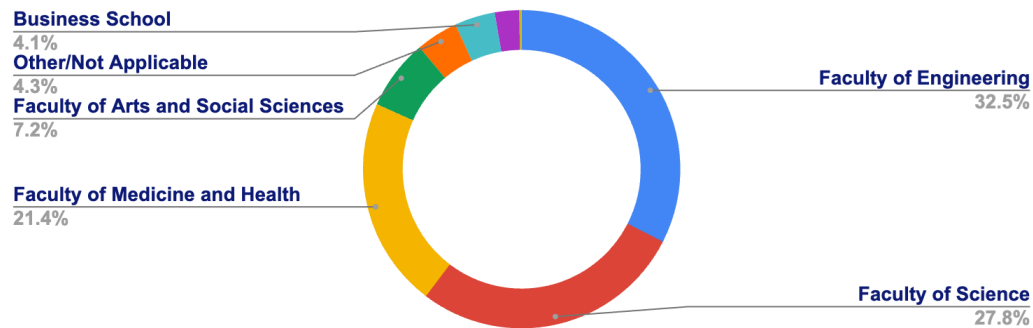
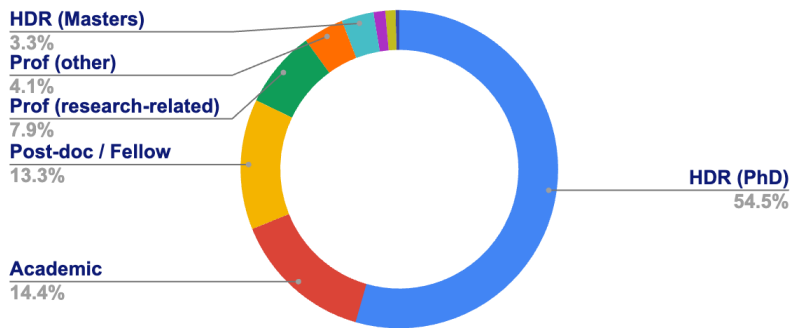
Nov
2020

Nov
2024

Feb
2025

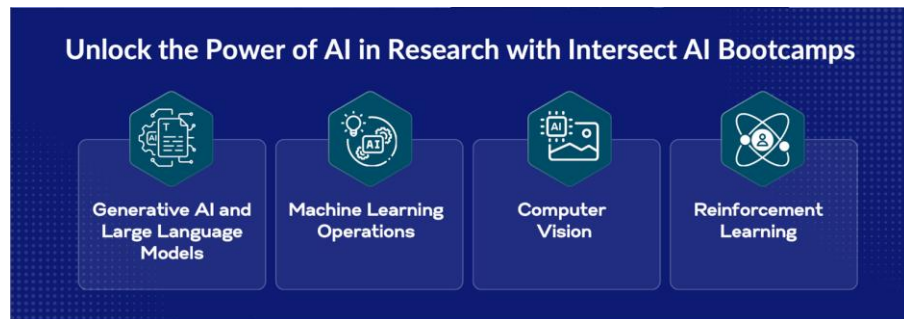
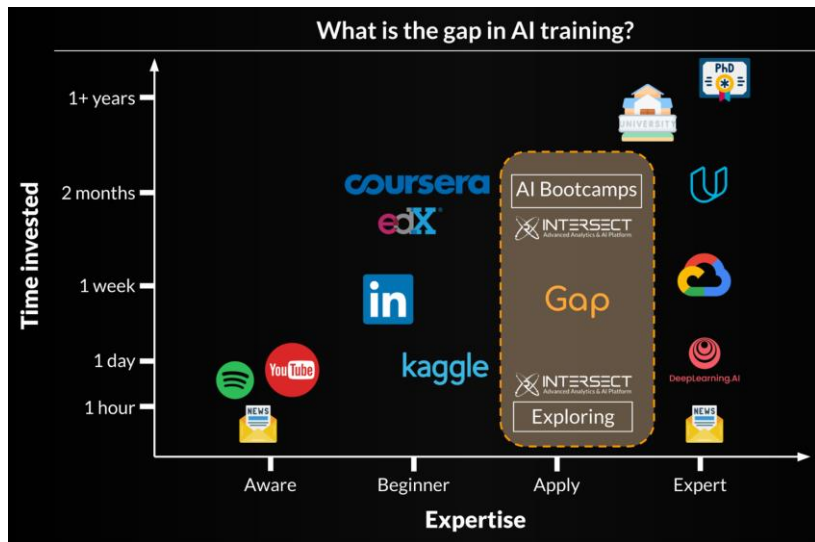
Intersect Python and R ML training courses

135 Courses delivered since 2020
2,639 Participants attended these courses
1,412 Unique Attendees



Official Launch of AI Bootcamps

Led by experienced research data scientists, our bootcamps provide hands-on training and direct application of cutting-edge AI techniques.



- ✓ Run across 7-8 weeks
- ✓ 5-day hands-on workshops
- ✓ Project work, exercises, quizzes
- ✓ Additional reading material

analytics@intersect.org.au | anastasios@intersect.org.au

- **What's currently happening (current challenges)**
 - Difficulty in sourcing expertise at **application of AI** in a researcher's primary field
 - Challenges in **staying up-to-date** in a field where change occurs at near light speed
 - Non-technical skills required to apply AI such as **awareness of bias, ethics, trust,** interpretability might be overlooked if too much focus is placed on technical skills
 - Steep learning curve; **people have to learn coding** (R, Python) first which takes quite some time, before they can get to ML/AI
 - Lack of understanding what appropriate data looks like, e.g., just because you have a lot of data doesn't mean it is appropriate for ML/AI - **rubbish in - rubbish out**

- **Existing skills development resources**
 - New AI training in development, for example: [Image Classification with Convolutional Neural Networks](#) in the Carpentries Incubator
- **6 to 12 month plan**



Tier-2

One-to-one consultation for members and post-workshop participants



Tier-1

Workshops and QCIF member advice provided by our eResearch Analysts



Tier-0

Self-help resources (open access, post-workshop access, members only)



Non-award courses with our member institutions for non-higher-ed sectors

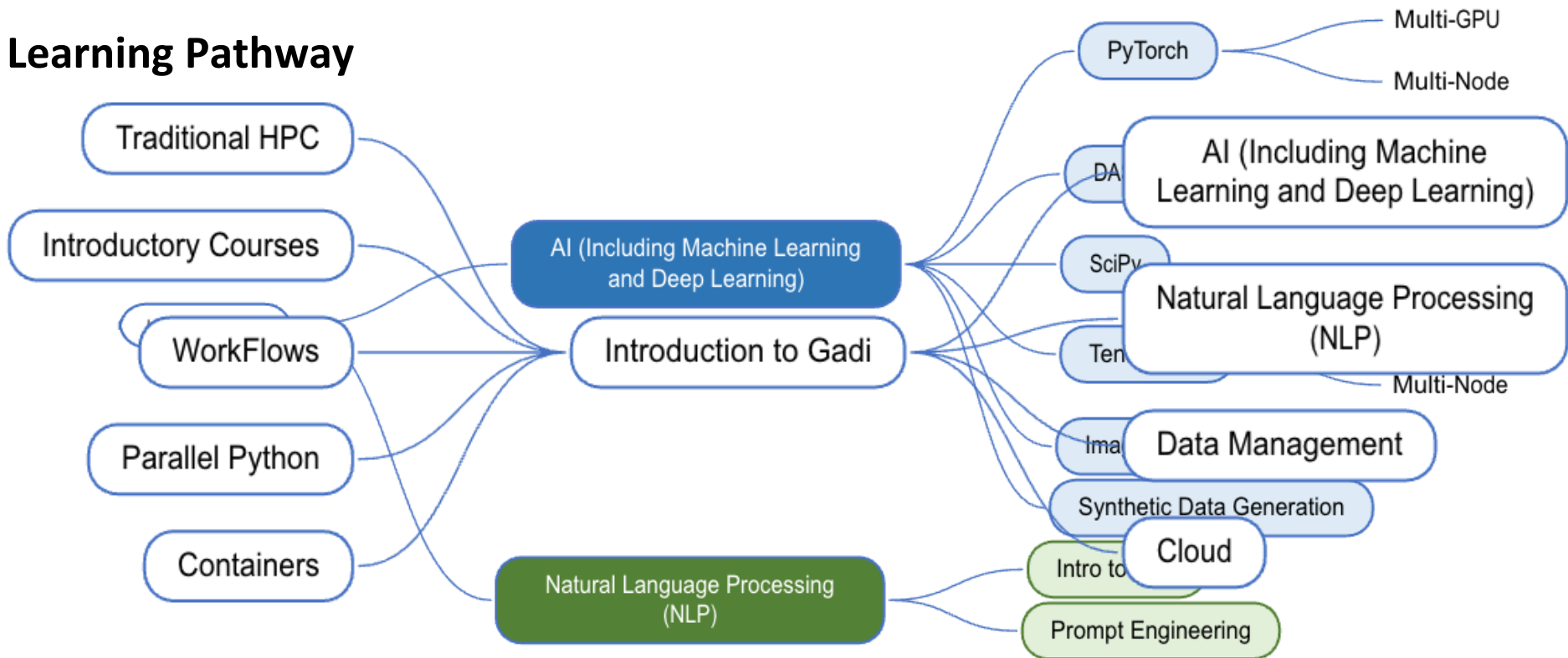
Machine Learning eResearch Platform (MLeRP):

- Platform is optimised for interactivity
 - Jupyter
 - Terminal
 - Batch Job
- Lowering the barriers to machine learning development
- Providing an environment where they can take on skills when ready

NCI Skills Development

- Learning Pathway
- Rejigged our training strategy
- New Courses
- Community Events

Learning Pathway

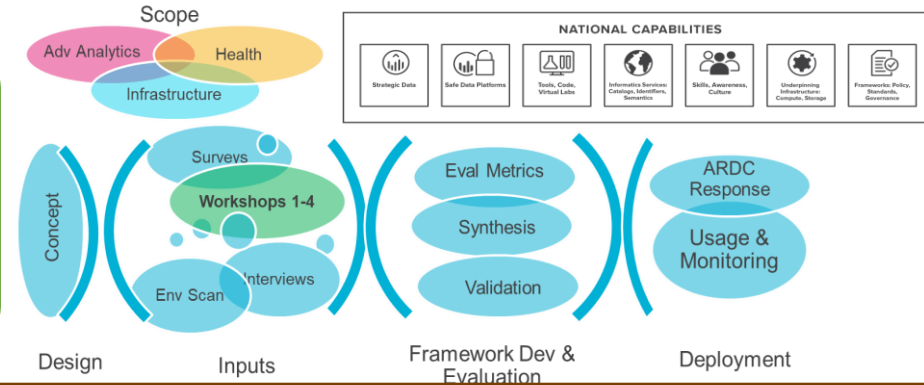


- Current
 - Introduction to NLP
 - Introduction to neural Network and Pytorch
 - Introduction to Deep Learning Development in Weather and Climate Studies
 - Parallel Programming, Python for HPC, C for Python users
- Upcoming
 - NCI QCIF Training Partnership
 - AI CodeFest in collaboration with Nvidia and National AI Centre
 - More Machine Learning and Deep Learning courses

People RDC Advanced Analytics Framework and Pathfinder Projects

Framework & Pathfinder

We co-developed a Framework Project (broadbrush) and a Pathfinder (drilldown) Project to explore one of the key infrastructure challenges



Rationale: To support infrastructure by: (i) capturing researcher inputs, (ii) providing opportunities for the research community to partner and participate in the journey, (iii) developing subsequent service and partnerships, and (iv) enabling adoption and usage

Framework Project: Breadth, sweeping the sector for challenges, needs, aspirations

Report: Armstrong, N. J., Bharathy, G. K., Buchhorn, K., Mehta, D., Mengersen, K., Papaioannou, A., Perrin, D., & Tanaka, E. (2024). **People RDC framework for advanced analytics for national healthcare research infrastructure.** Zenodo.

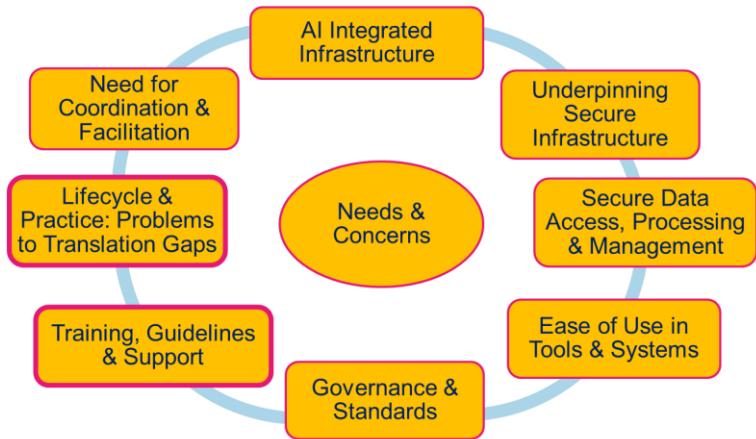
<https://doi.org/10.5281/zenodo.13831386>

Pathfinder Project: Diving deeper into a cardinal challenge

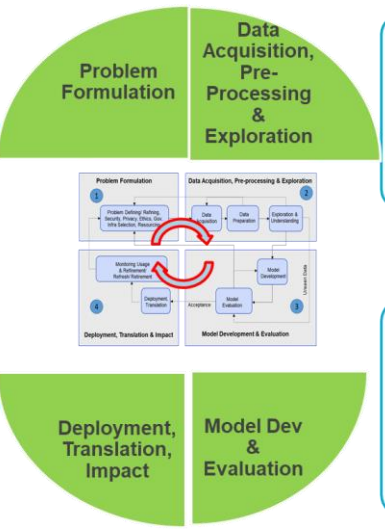
Report: Holloway, L., Anees, A., Al Mouiee, D., Uddin, A., Vafae, F., Haidar, A., Gorse, D., Sullivan, R., Wang, D., & Bharathy, G. K. (2024). **People RDC national pathfinder project: exploring federated learning tools, opportunities and resource requirements.** Zenodo.

<https://doi.org/10.5281/zenodo.13831454>

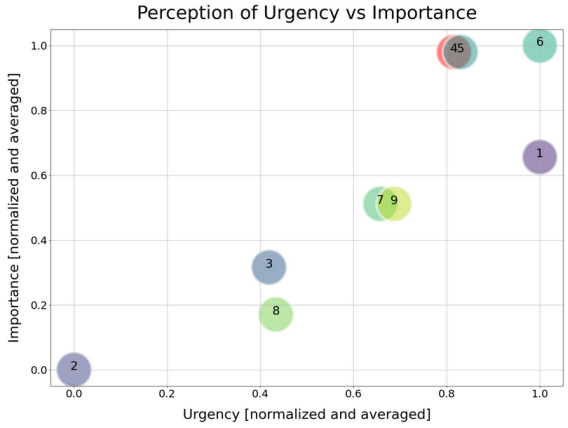




- Lack of Skills in Problem Formulation, Responsible AI
 - Need for Training & Socio-Tech Assets (e.g. Guidelines)
 - Most challenging and ignored step
-
- Gap between deployment/translation
 - Concern around inappropriate usage of advanced analytics



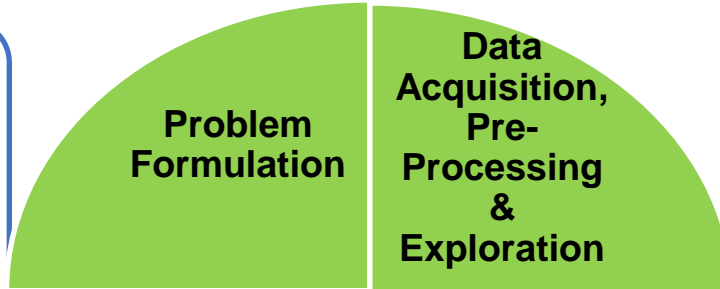
- **Shortfall in skills**
 - Managing sensitive data, data-problem gap
 - Interest in synthetic data
-
- Lack of supporting infrastructure for tool validation, despite available resources for development
 - Federated learning infrastructure challenges Socio-Tech



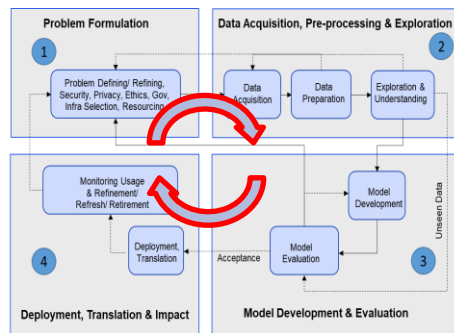
- [4] Data Assets - **Data Curation/ Pre-Processing Help,**
- [5] Socio-Tech Assets - **Training and Capacity Development - AI/ML, Coding, Version Control etc,**
- [6] Socio-Tech Assets - **Risk Management - Responsible AI, Guidelines, CheckLists,**
- [7] Tools/Platforms/ Environments - **Ease of Analysis with Virtual Labs, No-Code or Low Code Tools,**



- **Lack of Skills in Problem Formulation, Responsible AI**
- **Need for Training & Socio-Tech Assets (e.g. Guidelines)**
- Most challenging and ignored step



- **Shortfall in skills**
- Managing sensitive data, data-problem gap
- Interest in synthetic data



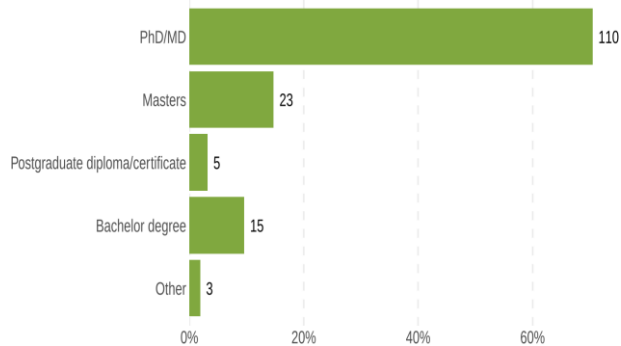
- Gap between deployment/ translation
- **Concern around inappropriate usage of advanced analytics**

- Lack of supporting infrastructure for tool validation, despite available resources for development
- Federated learning infrastructure challenges Socio-Tech

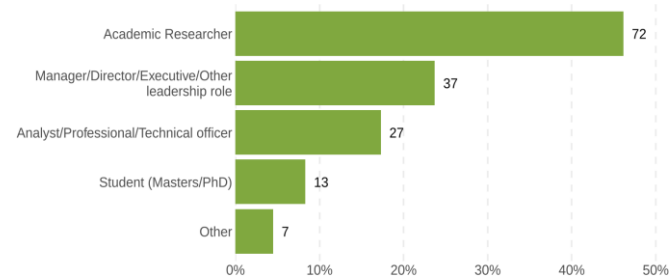
Skills – From National Survey (ADSN)

- Results today based on 118 valid responses
- Respondents were mainly
 - PhD graduates
 - Health/biomedical field
 - Academics (University or MRI)

What is your highest qualification?



What is your current primary position?

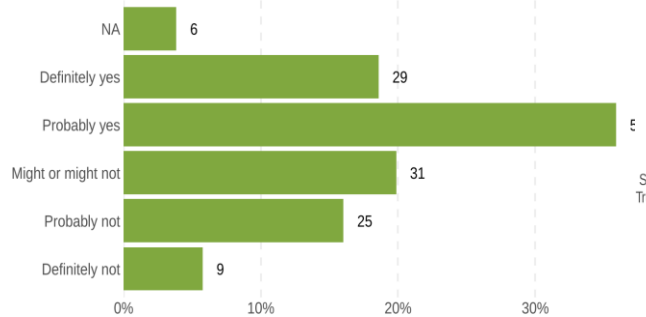


Skills

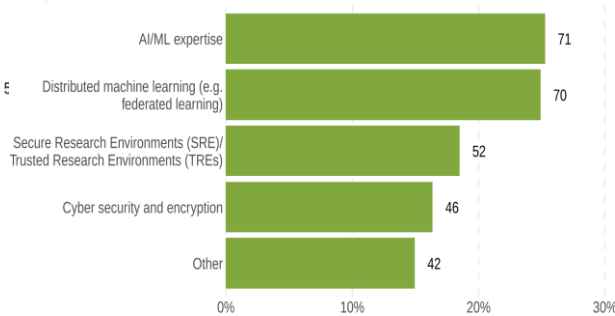
- Maths/stats: 90/118 say good or better
- AI/ML: 36/118 say good or better
- R/Python programming: 60/118 say good or better
- HPC/Cloud: 55/118 say good or better
- Databases: 46/118 say good or better

Skills – From National Survey (ADSN)

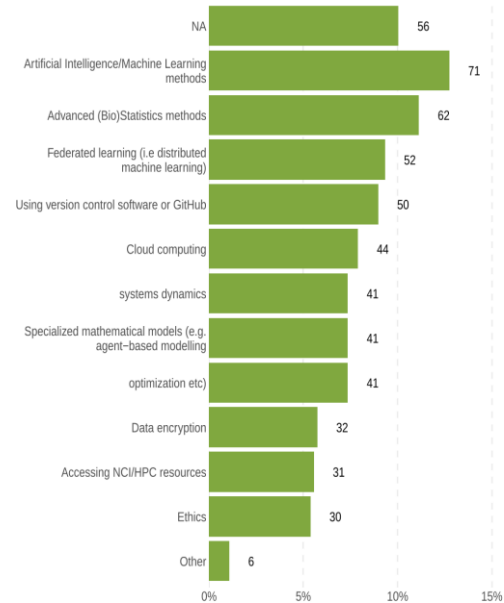
24. Does your organisation have sufficient advanced analytics capabilities in-house?



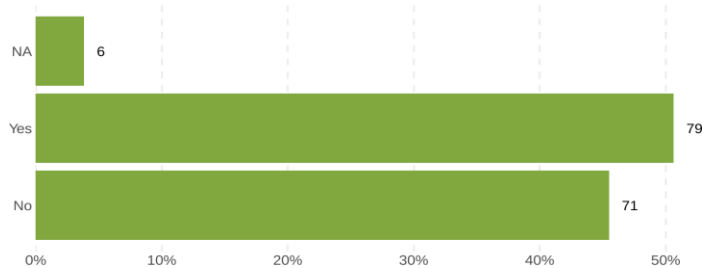
24.1 If no, what area(s) of advanced analytics are needed?



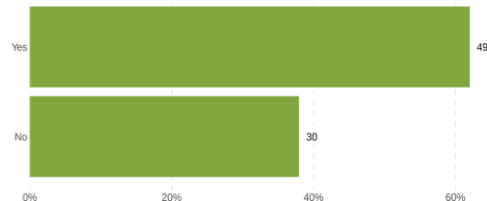
26. In which areas would you like training to be available?



25. Does your organisation provide access to formal training in advanced analytics?

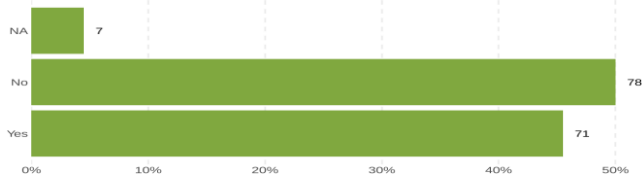


25.1 If yes, is this training sufficient for your needs?

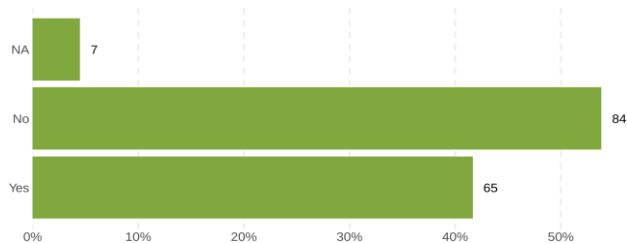


CoPs & Training – From National Survey (ADSN)

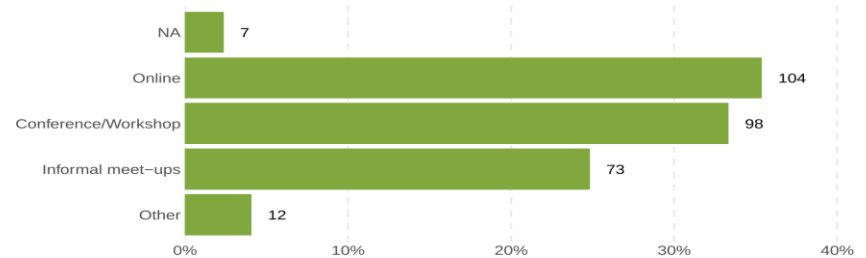
27. Do you have access to and/or utilise other informal training opportunities such as a slack channel, meetup or hacky hour?



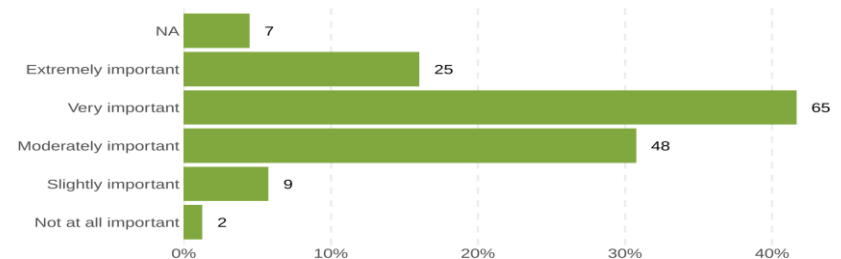
29. Do you participate in any community of practice for learning and development for advanced analytics in general and in health in particular?



28. How do you prefer to engage with others about advanced analytics?



30. How do you rate the need for such a community of practice?



People RDC Framework

Framework Project: Breadth, sweeping the sector for challenges, needs, aspirations

Report: Armstrong, N. J., Bharathy, G. K., Buchhorn, K., Mehta, D., Mengersen, K., Papaioannou, A., Perrin, D., & Tanaka, E. (2024). **People RDC framework for advanced analytics for national healthcare research infrastructure.** Zenodo.
<https://doi.org/10.5281/zenodo.13831386>

Pathfinder Project: Diving deeper into a cardinal challenge

Report: Holloway, L., Anees, A., Al Mouiee, D., Uddin, A., Vafaei, F., Haidar, A., Gorse, D., Sullivan, R., Wang, D., & Bharathy, G. K. (2024). **People RDC national pathfinder project: exploring federated learning tools, opportunities and resource requirements.** Zenodo.
<https://doi.org/10.5281/zenodo.13831454>

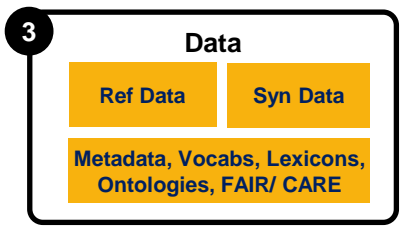
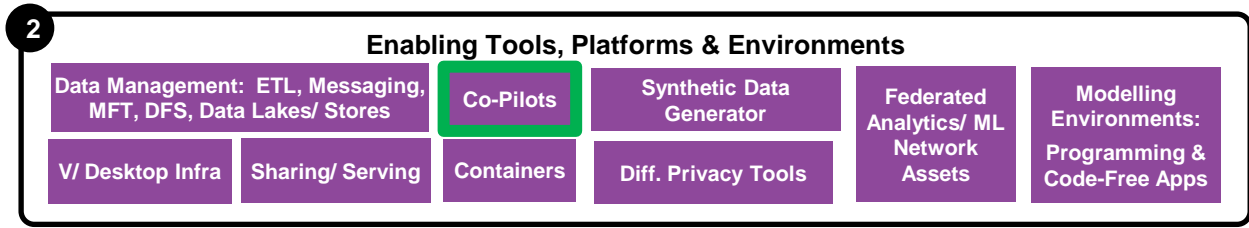
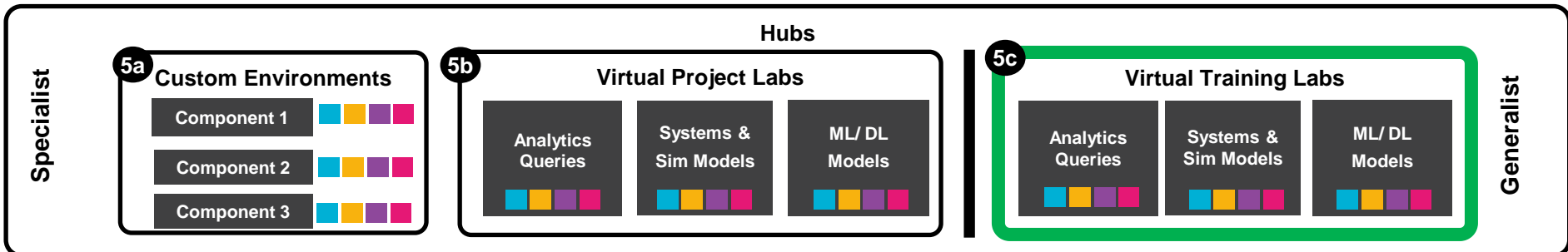


Reference Architecture & Roadmap



Program Design & Co-Investment Projects

AI & Advanced Analytics Reference Architecture



Panel Discussion & Audience Participation

Current skills gap and what's required to equip researchers to tackle ever-growing AI capabilities and requirements.

Audience Participation



slido.com

#2174434

**In there a gap between our
plan and what you expect
from the ML4AU
community?**

Conclusion



<https://ml4au.zulipchat.com/join/hlr3a7lqdvdlz4k7ri2otodr/>