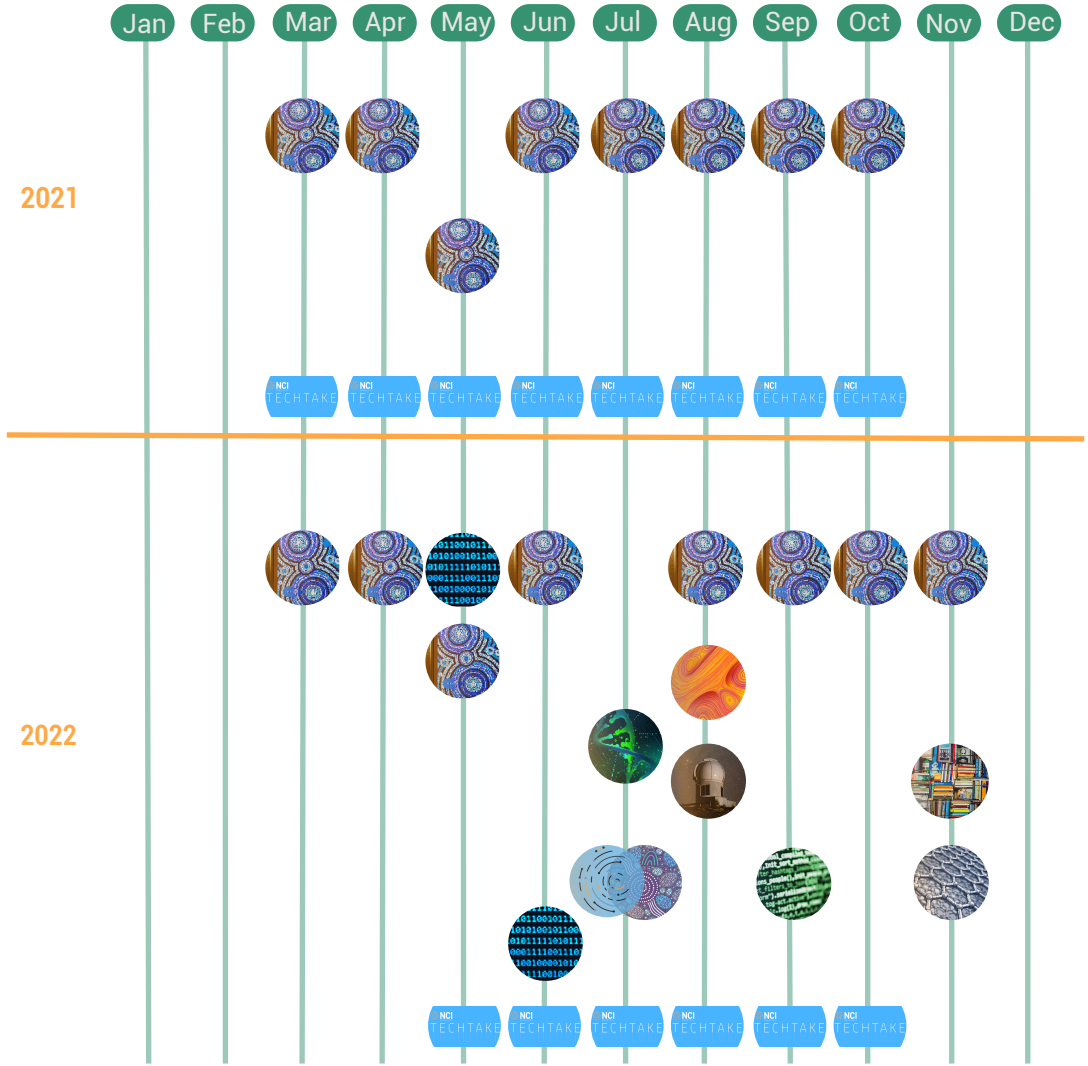
# NCI AUSTRALIA Training Program

## Course Calendar




To address the overwhelming need for fundamental command line and programming skills, we formulated a partnership with Intersect to offer a series of introductory-level courses. The development of the content is well on track in terms of when the HPC, data science, cloud computing and AI/ML applications on Gadi curriculum will be defined and launched.

We offer a variety of training courses and workshops on a regular basis. NCI also has a long-standing relationship with our vendors who co-host some of our training events by offering their expertise to address skills and knowledge as a complementary component.

### Introduction Course

Introduction to Gadi
Introduction to ARE
Introduction to Nirin

### HPC Toolkit

Introduction to OpenMP
Introduction to MPI
Parallel Python

### AI/ML Applications on Gadi

Astronomy
Climate and Weather
Computational Biology
Earth Observation
Natural Language Processing
Material Science

Scan the QR code to access course schedules and registration portals at NCI events page

If you require specific training topics or dedicated training sessions for your group, please contact training.nci@anu.edu.au

# Introduction To Gadi

Introduction Course

Gadi is Australia's most powerful supercomputer, a highly parallel cluster comprising more than 150,000 processor cores on ten different types of compute nodes. Gadi accommodates a wide range of tasks, from running climate models to genome sequencing, from designing molecules to astrophysical modelling.

## Goals

- Login to Gadi
- Transfer data on and off Gadi
- Run module commands to customise user environment and configure software applications
- Submit jobs
- Check and maintain compute, storage, and job status
- Estimate job cost
- Request resource adequate for your jobs
- Monitor job status/progress and its resource utilisation
- Understand common reasons why jobs finish with errors

## Targeted Audience

Introduction to Gadi is designed for new users, or users that want a refresher on the basics of Gadi.



# Introduction to ARE

Introduction Course

# Introduction to Nirin

Nirin is NCI's Cloud Service. ARE is a web-based graphical interface for performing your computational research at NCI. It combines the familiarity of your regular desktop/laptop with the power of NCI's world-class research HPC capabilities. ARE gives you access to NCI's Gadi supercomputer, Nirin cloud and Data collections, all from an easy to use, graphical interface.

## Goals

- Understand the functions that ARE and Nirin offer
- Submit computation jobs on ARE
- Set job configurations using ARE
- Launch Jupyter Lab for interactive jobs using ARE

## Targeted Audience

New or regular Gadi users who would like to access Gadi, Nirin and Data from a graphical interface.

# HPC TOOLKIT
# Introduction to OpenMP



OpenMP is a directive-based API (application programming interface) for writing parallel programs on a shared-memory system. The implementation renders parallelism for programs by running concurrent multi threads. The typical usage case is to accelerate nested loops by sharing workloads between multiple threads, which had been its main delivery before OpenMP 3.0. The workshop is designed to introduce some most common yet powerful OpenMP practices to scientists to turn a serial iterative C code into parallel quickly.

## Goals

- Know when to use OpenMP,
- Create Parallel Construct,
- Create a team of threads,
- Identify potential data race conditions,
- Distinguish data storage attributes,
- Understand how to split loop iterations to improve efficiency,
- Understand the limitations of multithreaded programming,
- Feel confident to advance to next-level parallel programming.

## Targeted Audience

Only basic experience with C/C++ is required. Knowledge about C preprocessor directives, functions, and pointer array is sufficient.



# HPC TOOLKIT
# Introduction to MPI



The Message Passing Interface (MPI) is arguably the primary programming model used for applications' internode parallelism. Developed by the NCI Training Team, the Introduction to MPI workshop demonstrates MPI procedures based on the latest MPI Standard - 4.0, with the hands-on finite difference exercise seen in the Introduction to OpenMP workshop.

In this workshop participants will be shown various MPI communication mechanisms to achieve exchanging boundary information of parallel finite difference method. Additionally, MPI-IO and MPI profiling will also be discussed.

## Goals

- Know when to use MPI for parallelization,
- How to prepare the buffer for communication calls,
- Distinguish and use blocking and non-blocking communications,
- Understand different communication modes,
- Overlap the communication and computation,
- Perform basic one-sided communications
- Output data in parallel with MPI-IO,
- Profile MPI applications,
- Feel confident for more advanced parallel programming.

## Targeted Audience

The workshop demonstrates examples in C. Only basic experience with C/C++ is required. Knowledge about C functions, pointers and memory management is sufficient.

# HPC TOOLKIT
## Parallel Python



Python is one of the most used programming languages worldwide with applications in almost every data-oriented application domain. The Python data science ecosystem is a rich platform for scaling up workflows, enhancing scientific research and improving insight. However, Python can be performance limited when large datasets or challenging computations are required. Parallel computing and efficient data handling can overcome this barrier, enhancing research throughput.

### Goals

- How use vectorized computation using NumPy
- How to load, annotate and work with data using xarray
- Serialise large datasets to file using xarray
- Load data from cloud using OpenDap and xarray
- Parallelise common workflows and arbitrary code using Dask
- Combine Dask and xarray for big data processing
- Combine Dask and GPUs for maximum data throughput
- Feel confident in your data science skills to tackle your own problems

### Targeted Audience

This Parallel Python workshop is designed to teach cutting edge techniques to work with big data and process data in parallel using Python and is suitable for all participants who want to enhance their data science capabilities



# NCI PRESENTS TECHTAKE



NCI Presents: TechTake is an exciting opportunity for international computational and data science leaders to discuss and demonstrate how technology supports research. Taking place on the last Tuesday of each month, this event will run online in order to reach diverse audiences across the globe and from all fields.

TechTake is designed to prompt engaging and in-depth conversations about both the current state and potential futures of technology to broaden and deepen understanding.

### Goals

- Learning about technology applications in different fields each session
- Learning about other researchers' projects and their methods
- Getting updated with new NCI technologies
- Direct conversation with hosts other fellow researchers

### Targeted Audience

This series of sessions are perfect for all researchers using computational methods in their projects who looks for an inspiring exchange. The session topics can be found on the NCI event page prior to each event date.

# AI/ML Applications on Gadi
## Computational Biology

In computational biology, we will talk about how to use linear regression to predict the total vaccinated population; and how to build a Support Vector Classifier to distinguish fake news; how to build a simple CNN model to predict Transcription Factor's binding site in single nucleotide resolution.

## Goals

- Understanding the basics of AI/ML
- Understanding the strength and weaknesses of different methods
- Formulating problems in Computational Biology (such as medical image segmentation, nanopore sequencing and protein binding sites prediction)
- Understanding how to apply basic ML/DL to biology problems
- Learning how to use Gadi

## Targeted Audience

Having basic programming experience with Python is required. Knowledge about using python packages like NumPy, pandas and scikit learn is advantageous. Know basic theory of Machine Learning and Deep Learning and have intentions to use AI/ML and supercomputers to boost their research.





# AI/ML Applications on Gadi
## Astronomy

Are you a Bayes fan but your model is just not up to the task because with great power comes great computing power consumption as well? Look no further; in this course, we will showcase how one can easily construct a neural network to rapidly emulate summary statistics from almost any complex model. We will also demonstrate how to write a simple Bayesian framework to perform inference with the emulator.

## Goals

- Models in astronomy (CMB, 21-cm, IGM neural fraction, Galaxy population)
- AI/ML in astronomy
- Basics of neural network
- Data preparation (using the recent 21-cm inference work as a demonstration, https://arxiv.org/abs/2108.07282)
- Write and train a neural network step by step
- Write and perform a Bayesian framework step by step

## Targeted Audience

Astronomy and Bayes fans. Only basic experience with Python is required. Knowledge about Machine Learning and Bayesian inference is sufficient.

NCI AUSTRALIA

# AI/ML Applications on Gadi
## Natural Language Processing

Natural Language Processing (NLP), the deciphering of text and data by machines, has revolutionised data analytics across all industries. It is the artificial intelligence-driven process of making human input language decipherable to software. In this course, we will showcase NLP use cases that are relevant to the HPC and STEM communities. NLP will help not only to speed up this process but also provide a much more comprehensive overview by extracting all the relevant papers from a global scholarly connected database.

### Goals

- Understanding the basics of NLP
- Understanding the pre-processing steps of text data
- Understanding how to apply basic NLP techniques
- Understanding how to run NLP applications on Gadi

### Targeted Audience

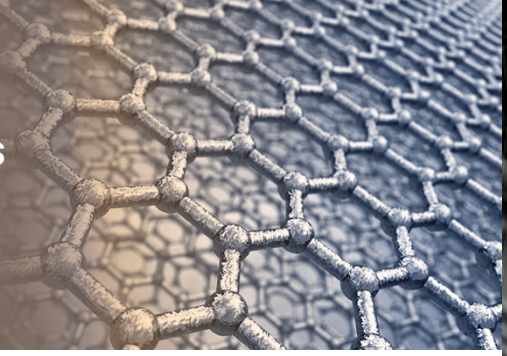
Having basic programming experience with Python is highly recommended. Knowledge about using text processing python packages like NLTK is advantageous. Attendees will ideally know some basic theory of Machine Learning and Deep Learning, and have intentions of using AI/ML and supercomputers to boost their research.





NCI AUSTRALIA

# AI/ML Applications on Gadi
## Materials Science



In this AI/ML in Materials Science course, we will discuss how ML is poised to have a similar transformative impact in materials science. Applied on large data sets, ML techniques can be used to discover novel technological materials, to model complex systems at an accuracy beyond the reach of traditional computational techniques, and to enhance the accuracy and speed of interpreting characterisation data. A key focus of this course will be on the heterogeneity and scarcity of materials data, the challenges these characteristics present for ML, and the potential approaches to overcome them.

### Goals

- Understanding the basic of AI/ML
- Understanding the strength and weakness among different methods
- Understanding how to apply basic ML/DL techniques
- Understanding how to use Gadi

### Targeted Audience

Having basic programming experience with Python is required. Knowledge about using python packages like NumPy, pandas and scikit learn is advantageous.
Attendees will ideally know some basic theory of Machine Learning and Deep Learning, and have intentions of using AI/ML and supercomputers to boost their research.

NCI AUSTRALIA

## General Enquires

02 6125 9800

enquiries@nci.org.au

## Media Enquires

02 6125 43

## Location

National Computational Infrastructure
The Australian National University
143 Ward Road Acton2601ACT

## Help desk

If you are an existing user, submit a help request at help.nci.org.au

help@nci.org.au

## Training team

training.nci@nci.org.au

## License