





# Open Research Software

Best practices and tools for reliable, maintainable software

Liam Keegan, SSC



## Scientific Software Center

- Team of Research Software Engineers (currently 6)
- Offer researchers at Heidelberg University
  - Software development as a service
  - Consultation and advice
  - Teaching and training
- Our website / github page also offers
  - Coding guidelines
  - Template repositories
  - Slides and code samples from our courses



ssc.uni-heidelberg.de



<u>aithub.com/ssciwr</u>



### Research Software

- Is an increasingly vital part of scientific research
- Is an intrinsic part of reproducible science
- Is not only code written by "real programmers"
  - Your Python data analysis script is also research software!

For people to trust your research, they need to trust your software

- Needs to be open
- Needs to be reliable
- Needs to be maintainable



## Best practices for reliable, maintainable software

- Open source development
- Version control
- Testing
- Documentation
- Continuous integration
- Community involvement

### For each of these I will

- Describe what it is and what the benefits are
- Make some concrete recommendations
- Show an example of this from an open source library (pybind11)



## Tools for writing reliable, maintainable software

- Integrated Development Environments (e.g. vscode, pycharm)
- Al code completion (e.g. copilot, intellicode)
- Generative AI (e.g. chatgpt, claude, gemini)

### For each of these I will

- Describe what it is and what the benefits are
- Make some concrete recommendations
- Show an example of use



# Open source development



## Open source development

- Making your source code publicly available
  - o e.g. GitHub, GitLab, Software Heritage, Zenodo
- With a suitable open source license
- Advantages
  - Makes it easier for people to reproduce your results
  - People can find mistakes and bugs
  - People can fix mistakes and bugs
  - People can offer suggestions, improvements
  - People can cite and use your work
  - Gives others confidence in the value of your code



## Which platform to use?

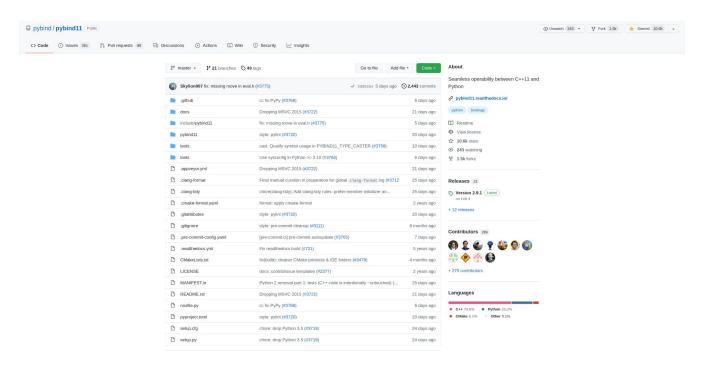
- GitHub.com / GitLab.com / etc
  - Commercial git hosting with a (substantial) free tier of services
- Self-hosted gitlab / forgejo / etc
  - Your institute may offer self-hosted gitlab or other code hosting services
- Software Heritage
  - Public software archive, provides a SWHID for your code
- Zenodo
  - Public data archive, provides a DOI for your code

#### Recommendation:

Some form of git hosting + Software Heritage + Zenodo



# Open source development example





# Version control



### Version control

- Use a tool to track changes to your software
  - e.g. git, subversion, mercurial

### Advantages

- Easily keep track of changes to the code
- What changed, who changed it, when and why it changed
- Easy to refer to specific commit, tag or version for reproducibility
- Easy to undo or revert changes
- Easy for multiple people to collaborate on the same code
- Gives others confidence in the history of your code



## Which version control system to use?

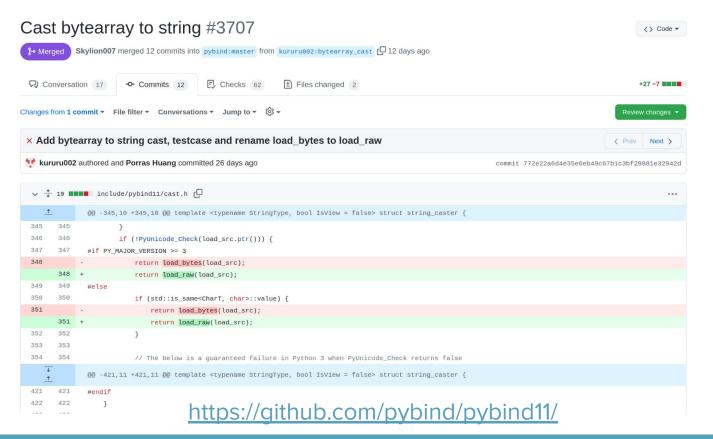
- Git
  - The de-facto standard, now used by the vast majority of open source projects
- Workflows
  - There are many ways to use git, known as workflows
  - Centralized workflow, Feature branching, Forking workflow
  - o git-flow, gitlab-flow, github-flow, ...

### Recommendation:

- Git with a main branch
- New code is developed on a new branch and then merged into main



# Version control example





# Testing



# **Testing**

- Write automated tests that check the software is working correctly
- Advantages
  - Ensure correctness of your code
  - Maintain correctness of your code
  - Find bugs earlier and more easily
  - Make changes or refactor code without fear
  - Easier for new contributors to make positive changes
  - Complement the documentation as examples of use
  - Gives others confidence in the correctness of your code



## Types of tests

- Unit tests
  - Test a small, isolated part of code
- Integration / system tests
  - Test larger, connected parts of code
- Regression tests
  - Test for a bug that was fixed to ensure it doesn't come back
- Approval tests
  - Retro-fitting tests before making changes to legacy code

#### Recommendation:

- Write unit tests for new projects or new code in legacy projects
- Write approval tests for legacy code which doesn't have any tests



## Testing example

```
platform linux -- Python 3.9.10, pytest-7.0.0, pluggy-1.0.0
18 rootdir: /home/runner/work/pybind11/pybind11/tests, configfile: pytest.ini
   plugins: timeout-2.1.0, github-actions-annotate-failures-0.1.6
   timeout: 300.0s
  timeout method: signal
   timeout func_only: False
   collected 528 items
   test_async.py ...
   test_buffers.py .....
   test_builtin_casters.py ....s.....
   test_call_policies.py .....
  test_callbacks.py .....
   test_chrono.py .....
   test class.pv .....
   test_const_name.py .....
   test_constants_and_functions.py .....
   test_copy_move.py ....s..
   test_custom_type_casters.py ...
   test_custom_type_setup.py ...
   test_docstring_options.py .
   test eigen.py ......
39 test_enum.py .....
   test_eval.py ....
41 test_exceptions.py .....
```



# Documentation



## Documentation

- Document how your code works and how to use it
- Advantages
  - Helps users understand how to use the code
  - Helps developers understand how to modify the code
  - Encourages people to learn about your code
  - Gives others confidence in the usability of your code
  - By writing it you can identify hard-to-use code that could be improved



## Types of Documentation

- Source code
  - Target audience is other humans, not the computer!
- Comments
  - For you and other developers
- API Documentation
  - Technical documentation for developers / power users
- User documentation
  - Documentation written for users
- Examples
  - Very helpful

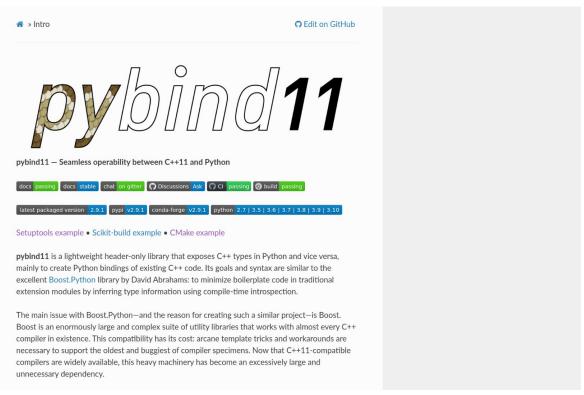
#### Recommendation:

Include your documentation in your git repository and update it alongside code changes



## Documentation example







# Continuous integration



## Continuous integration

- Automatic checks before code changes are accepted
- Advantages
  - Ensure all tests pass before code is changed
  - Can automatically apply uniform formatting of the code
  - Can automatically do static analysis to identify code smells or bugs
  - Can require that new code is covered by tests
  - Test the code on multiple platforms (e.g. Windows, Mac, Linux)
  - Can automatically deploy new releases of software
  - Helps others improve the quality of their proposed code changes



## Types of continuous integration

- Integrated into git hosting service
  - o GitHub Actions, GitLab CI/CD, ...
- External services
  - o Travis Cl, Circle Cl, ...
- Self hosted
  - Jenkins, ...

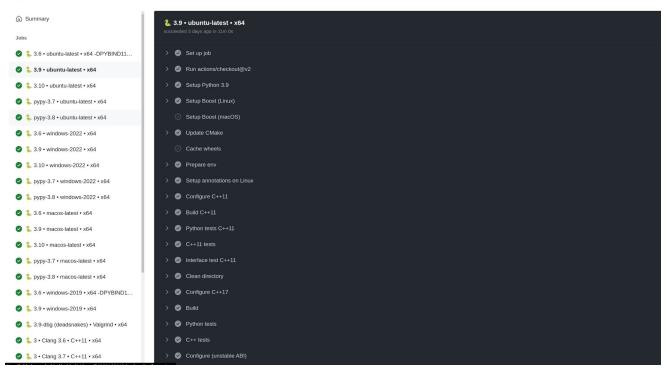
#### Recommendation:

- Typically easiest to use the CI provided by your git hosting service
- E.g. for code on GitHub use GitHub Actions



## Continuous integration example

Docs: No Strip in Debug CI #4327



https://github.com/pybind/pybind11/



# Community involvement



## Community involvement

- Enable people to contribute bug reports, feature requests and code
- Advantages
  - People can find mistakes and bugs
  - People can fix mistakes and bugs
  - People can improve the documentation
  - People can offer suggestions, improvements
  - People can help each other to use your code
  - More contributors can make a project more sustainable
  - Helps others to use and contribute to your work



### Communication channels

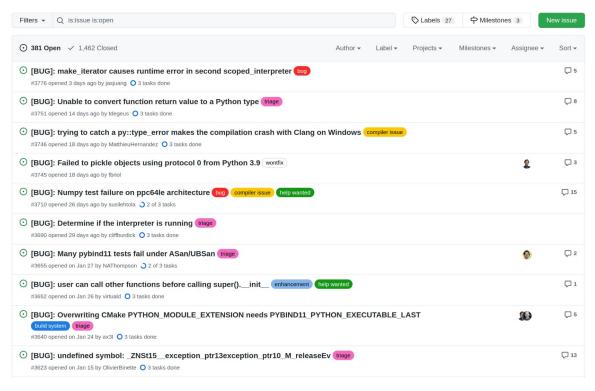
- Issue trackers on git hosting service
- Public chat room / discussion board
- Mailing list
- Contact email for support / questions
- Wiki pages

### Recommendation:

• Use public issue trackers for all feedback / discussions / bugs / features



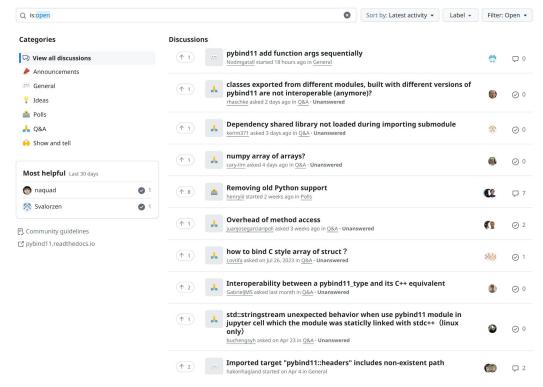
# Community involvement example



https://github.com/pybind/pybind11/issues



# Community involvement example



https://github.com/pybind/pybind11/discussions



# Integrated Development Environment



# Integrated Development Environment (IDE)

The IDE is where you type your code, and can provide

- Syntax highlighting
- Code formatting
- Code completion
- Code refactoring
- Compiling, running and debugging
- Warnings about errors and issues
- Suggested improvements based on best practices



### Which IDE to use?

- Jetbrains (PyCharm / CLion / etc)
  - Commercial, but students can get a free license
  - There is also a free community edition of PyCharm for Python coding
- Visual Studio Code
  - o Free, open-source, widely used
- Many others available
  - Eclipse, Visual Studio (commercial), SublimeText, notepad++, vim

#### Recommendation:

VS Code or JetBrains



# IDE example (Python code in PyCharm)

```
import numpy as np
class MyClass:
    def __init__(self, x=[1.0, 0.5]):
         self.x = x
                                            Default argument value is mutable
         self.rotation = np.cos(x)
         self.translation = np.sin(x Replace mutable default argument Alt+Shift+Enter More actions... Alt+Enter
         self.result = self.rotation + self.translation
class MyClass:
                                                                             Default argument value is mutable
                                                                             Reports a problem when a mutable value as a list or dictionary is detected in a default value for an argument.
def __init__(self, x=None):
                                                                             Default argument values are evaluated only once at function definition time, which means that modifying the default value
                                                                             of the argument will affect all subsequent calls of that function.
         if x is None:
                                                                             Example:
             x = [1.0, 0.5]
         self.x = x
                                                                             When the quick-fix is applied, the code changes to:
         self.rotation = np.cos(x)
         self.translation = np.sin(x)
                                                                                 if cache is None:
         self.result = self.rotation + self.translation
                                                                                 cache[s] = None
```



# Al code completion



## Al code completion

- Autocomplete (on steroids) for your IDE
- As you type it makes pretty good suggestions with surprisingly little input
- You can start writing a function and it will suggest the rest
- It can also generate documentation for a function
- It can write tests for a function
- It can also simplify or explain code
- Works especially well with commonly used or "boilerplate" code



## Which AI code completion to use?

- IntelliCode
  - o Free, included in VS Code
- GitHub Copilot
  - Commercial
  - But students can get a free GitHub PRO account which includes copilot
- Many others available
  - But the two above are probably the most widely used

### Recommendation:

IntelliCode or Copilot



## Al code completion examples (copilot/pycharm)

I started writing a function by typing "def rotate\_vector":

```
def rotate_vector (vector, angle):
    """Rotate a vector by a given angle in radians."""
    x, y = vector
    cos_angle = math.cos(angle)
    sin_angle = math.sin(angle)
    return (x * cos_angle - y * sin_angle, x * sin_angle + y * cos_angle)
```



## Al code completion examples (copilot/pycharm)

Then edited my text to "def rotate\_3d\_vector":

```
def rotate_3d_vector(v, angle, axis):
    """Rotate a 3D vector by a given angle around a given axis."""
    # Normalize the axis vector
    axis = axis / np.linalq.norm(axis)
    # Compute the rotation matrix
    R = rotation_matrix(angle, axis)
    # Rotate the vector
    return np.dot(R, v)
```



# Generative Al



## Generative Al

- You can simply ask a chatbot to write your code for you
- The results are not perfect but often pretty good!
- You can also iterate, asking for changes and improvements
- The kind of code you get depends on the kind of prompt you use
  - E.g. "Please write high quality code that follows best practices in software development."
- Can be a good way to brainstorm alternative solutions
- Also very helpful when you are not an expert with the language you are using
- Your role here changes from writing code to reviewing code



## Possible issues with generative Al

- Copyright issues
  - Was the training data copyrighted? Is your data copyrighted?
- Spaghetti code
  - Easy to quickly generate a lot of not-great code: makes your life harder later
- Subtly wrong code
  - They sometimes "hallucinate", which is a polite word for lying convincingly to you

### Recommendation:

Use with care!



## Generative AI examples

You
write a python function that rotates a 3d vector

#### 

To write a Python function that rotates a 3D vector, we need to use some linear algebra. Specifically, we'll use rotation matrices. A common way to represent 3D rotations is using Euler angles or quaternions, but for simplicity, we'll use rotation matrices for rotations around the principal axes (x, y, and z).

Here's the code to rotate a 3D vector using rotation matrices:

```
python

import numpy as np

def rotate_vector(vector, angle, axis):

"""

Rotate a 3D vector around a given axis by a specified angle.

Parameters:

vector (list or np.ndarray): The 3D vector to rotate.

angle (float): The rotation angle in radians.

axis (str): The axis to rotate around ('x', 'y', or 'z').

Returns:

np.ndarray: The rotated 3D vector.

"""
```

```
if axis not in ['x', 'y', 'z']:
    raise ValueError("Axis must be 'x', 'y' or 'z'")
if axis == 'x':
    rotation_matrix = np.array([
       [0, np.cos(angle), -np.sin(angle)],
       [0, np.sin(angle), np.cos(angle)]
    1)
elif axis == 'y':
    rotation_matrix = np.array([
       [np.cos(angle), 0, np.sin(angle)],
       [-np.sin(angle), 0, np.cos(angle)]
elif axis == 'z':
    rotation_matrix = np.array([
       [np.cos(angle), -np.sin(angle), 0],
       [np.sin(angle), np.cos(angle), 0],
rotated_vector = np.dot(rotation_matrix, vector)
return rotated_vector
```



## Generative AI examples

You

You are an expert python developer. Write a python function that rotates a 3d vector

#### ChatGPT

Certainly! Below is a Python function that rotates a 3D vector around a specified axis by a given angle. The function uses the Rodrigues' rotation formula, which is a common method to rotate vectors in 3D space.

```
ரி Copy code
import numpy as np
def rotate_vector(vector, axis, angle):
   numpy.array: The rotated vector.
```

```
# Normalize the axis vector
axis = axis / np.linalg.norm(axis)
# Compute the cross-product matrix of the axis
K = np.array([
    [0, -axis[2], axis[1]],
    [axis[2], 0, -axis[0]],
    [-axis[1], axis[0], 0]
])
# Compute the rotation matrix using Rodrigues' rotation formula
I = np.eye(3)
R = I + np.sin(angle) * K + (1 - np.cos(angle)) * np.dot(K, K)
# Rotate the vector
rotated_vector = np.dot(R, vector)
return rotated_vector
```



# Summary



## For reliable, maintainable software

### Follow best practices:

- Open source development
- Version control
- Testing
- Documentation
- Continuous integration
- Community involvement

### Use good tools:

- Integrated Development Environment
- Al-powered autocomplete
- Generative AI (with care)



## Getting started

- Start from a template repository
- Basic project ready to go
  - Open source development
  - Version control
  - Testing
  - Documentation
  - Continuous integration
  - Community involvement



Basic C++ Project Template

github.com/ssciwr/cpp-project-template



**Basic Python Project Template** 

github.com/ssciwr/python-project-template



Advanced C++ Project Template

github.com/ssciwr/cookiecutter-cpp-project