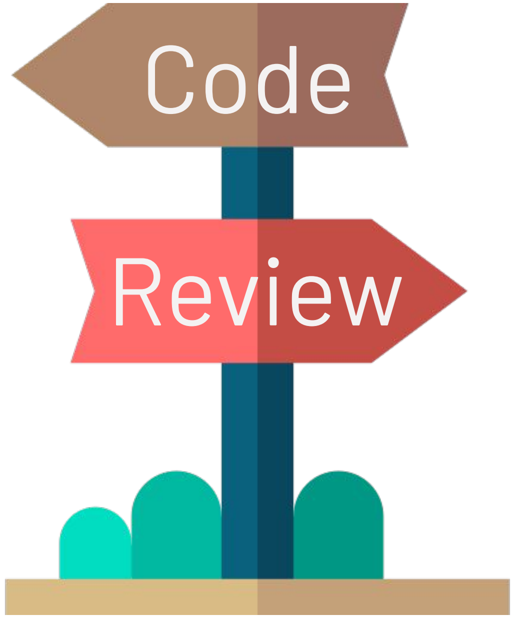
scieneers
DRIVEN BY DATA

A Data Scientist's Guide to Code Reviews

PyCon/PyData 2022



Doing code reviews

Pros

- They'll improve the code clarity
- They might uncover errors
- I'll probably learn something while doing it
- Literally everyone says that I should

Cons

- I don't want to

⇒ I will **not** do code reviews

Doing code reviews on data science work

Pros

- They'll improve the code clarity
- They might uncover errors
- I'll probably learn something while doing it

Cons

- Hardly anyone says that I should
- The code will likely not end up in a live system as-is
- The work is a one-off thing

⇒ I will **not** do code reviews

However, having someone else review your work is as important in data science as in software engineering.

scieneers
DRIVEN BY DATA

What are code reviews for?

- → Verifying that the specified goal is achieved
- → Uncovering errors and misunderstandings
- → Knowledge transfer
- → Feedback for architectural or design decisions
- → Improving your code & coding practice

The traditional code review practice is not applicable to "typical" data science work.

Different focus

Software Engineering

- Is the artifact functional?
- Are there bugs?
- Are coding guidelines & quality standards met?
- Can someone else than the author work on the artifact?

⇒ Code Review

Data Science

- Is the chosen approach comprehensible & clear?
- Have data peculiarities been taken into account?
- Are the results plausible?
- Can someone else than the author explain the concept?

⇒ Peer Review

What are code reviews for in data science?

→ Verifying that the specified goal is achieved ✅

→ Uncovering errors and misunderstandings

→ Knowledge transfer

→ Feedback for architectural or design decisions

→ Improving your code & coding practice

What are code reviews for in data science?

- → Verifying that the specified goal is achieved ✅
- → Uncovering **logical** errors and misunderstandings ✅
- → Knowledge transfer ✅
- → Feedback for architectural or design decisions
- → Improving your code & coding practice

What are code reviews for in data science?

→ Verifying that the specified goal is achieved ✅

→ Uncovering **logical** errors and misunderstandings ✅

→ Knowledge transfer ✅

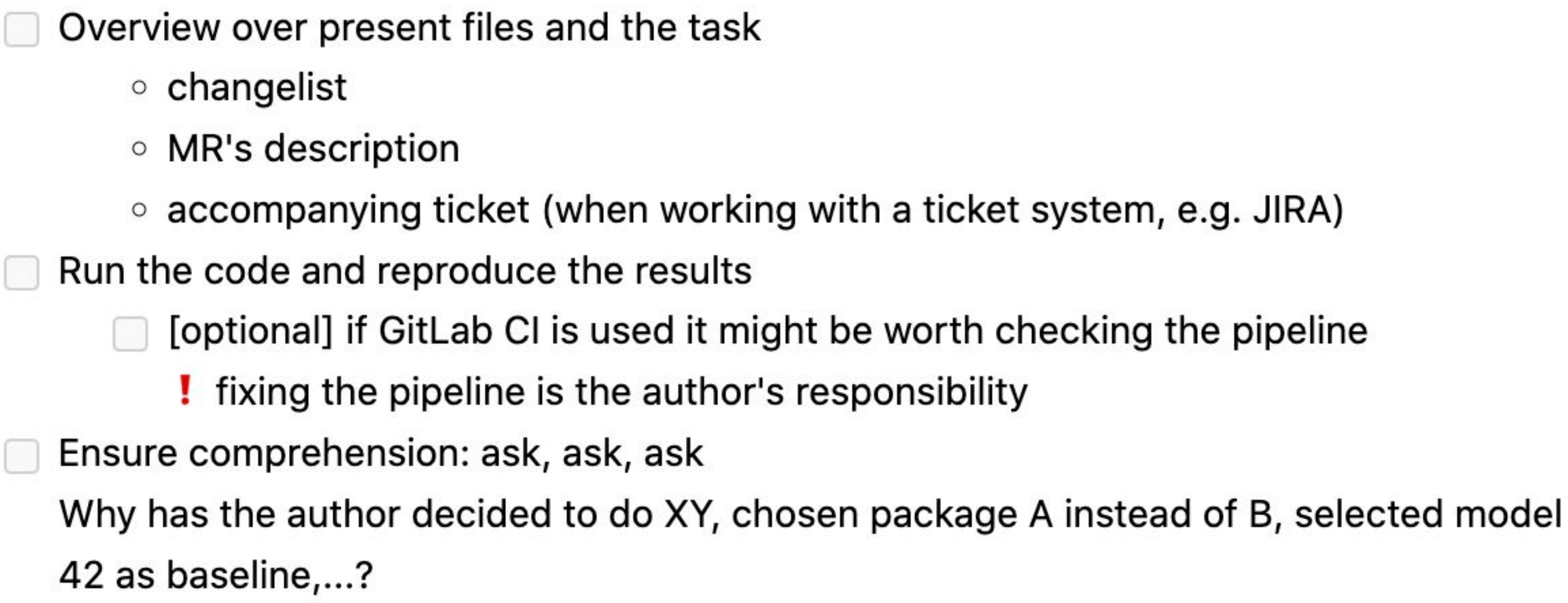
→ Feedback for ~~architectural or design decisions~~ **approach** ✅

→ Improving your code & coding practice

What are code reviews for in data science?

- Verifying that the specified goal is achieved ✅
- Uncovering **logical** errors and misunderstandings ✅
- Knowledge transfer ✅
- Feedback for ~~architectural or design decisions~~ **approach** ✅
- ~~Improving your code & coding practice~~ **Reproducibility** 🔁

Code review checklist



- □ Overview over present files and the task
    - changelist
    - MR's description
    - accompanying ticket (when working with a ticket system, e.g. JIRA)
- □ Run the code and reproduce the results
    - □ [optional] if GitLab CI is used it might be worth checking the pipeline
        - ❗ fixing the pipeline is the author's responsibility
- □ Ensure comprehension: ask, ask, ask

  Why has the author decided to do XY, chosen package A instead of B, selected model 42 as baseline,...?

github.com/awoerner92/talks

How to do code reviews?

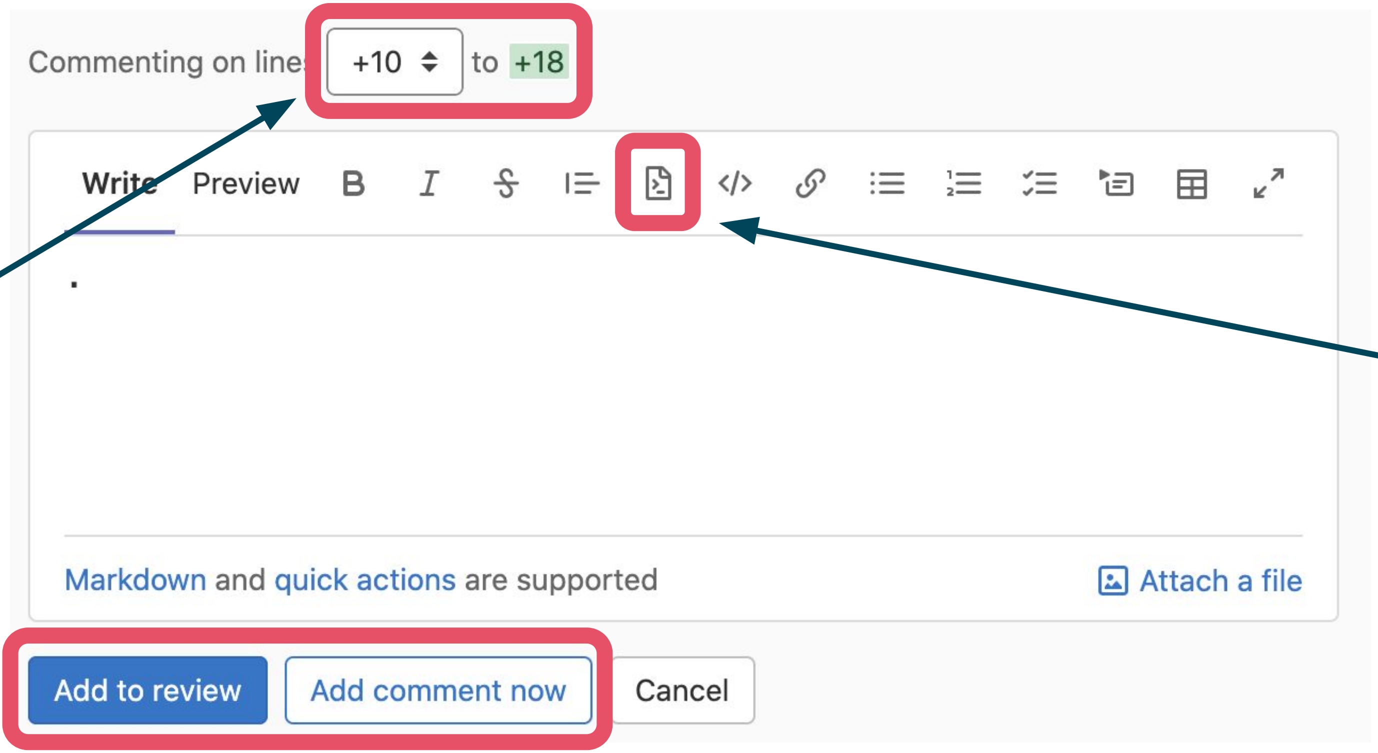
- How to Do Code Reviews Like a Human
  - https://mtlynch.io/human-code-reviews-1/
  - https://mtlynch.io/human-code-reviews-2/

- How to Make Your Code Reviewer Fall in Love With You
  - https://mtlynch.io/code-review-love/

Useful Git functionality: pre-commit hooks

- Runs pre-defined set of tools with every commit
- Tools:
  - Jupyter notebook conversion: nbconvert
  - Code formatter: black, isort
  - Linter: flake8

Useful GitLab functionalities: Comment field


Commenting on lines +10 to +18

comment & mark multiple lines

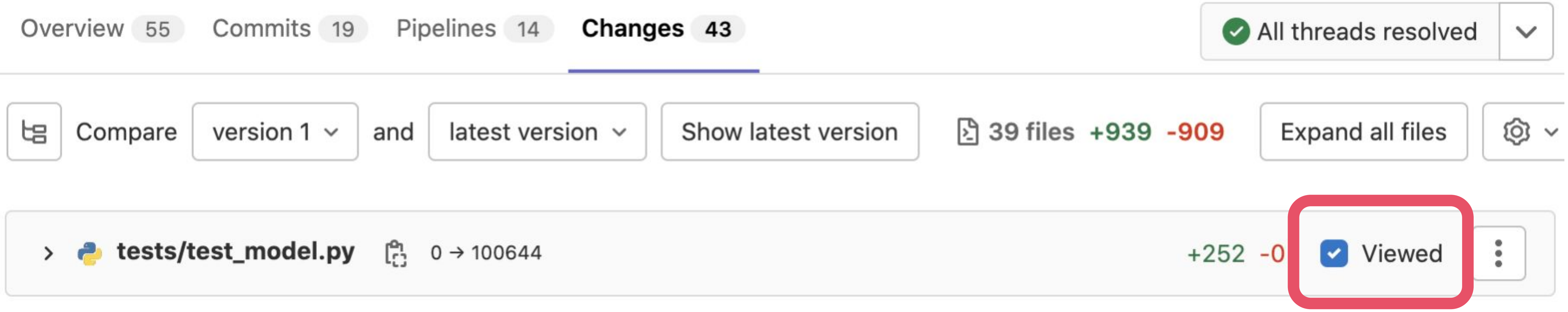
code suggestion

(GitHub:
~~~suggestion
```)

add comment to batch or comment immediately

GitLab documentation on merge requests: https://docs.gitlab.com/ee/user/project/merge_requests/
~~~

Useful GitLab functionalities: Mark viewed



→ Collapses the file

→ Helps to keep an overview

Thank you!

Alexandra Wörner

Data Scientist

alexandra.woerner@scieneers.de

@alex_woerner

Questions?