



UniTs - University of Trieste

Faculty of Data Science and Artificial Intelligence/Scientific Computing
Department of mathematics informatics and geosciences

Generative AI and NLP

Author:
Christian Faccio

March 9, 2025

This document is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike](https://creativecommons.org/licenses/by-nc-sa/4.0/) (CC BY-NC-SA) license. You may share and adapt this material, provided you give appropriate credit, do not use it for commercial purposes, and distribute your contributions under the same license.

Contents

1	Git	1
1.1	About Git	1
1.1.1	About version control and Git	1
1.1.2	Repositories	1
1.1.3	GitHub	2
1.1.4	Command Line Interface	2
1.1.5	Models for collaborative development	4
1.2	Pushing commits to a remote repository	5
1.2.1	About <code>git push</code>	5
1.2.2	Renaming branches	5
1.2.3	Dealing with "non-fast-forward" errors	6
1.2.4	Pushing tags	6
1.2.5	Deleting a remote branch or tag	6
1.2.6	Remotes and forks	6
1.3	Getting changes from a remote repository	7
1.3.1	Options for getting changes	7
1.3.2	Cloning a repository	7
1.3.3	Fetching changes from a remote repository	8
1.3.4	Merging changes into your local branch	8
1.3.5	Pulling changes from a remote repository	8
1.4	Dealing with non-fast-forward errors	9
1.5	Splitting a subfolder out into a new repository	9
1.6	About Git subtree merges	11
1.6.1	Setting up the empty repository for a subtree merge	11
1.6.2	Adding a new repository as a subtree	11
1.6.3	Synchronizing with updates and changes	12

1.1 About Git

1.1.1 About version control and Git

Definition: *Version control*

A version control system, or VCS, tracks the history of changes as people and teams collaborate on projects together. As developers make changes to the project, any earlier version of the project can be recovered at any time.

Developers can review project history to find out:

- Which changes were made?
- Who made the changes?
- When were the changes made?
- Why were changes needed?

In a **distributed version control system (DVCSs)**, every developer has a full copy of the project and project history. They don't need a constant connection to a central repository. **Git** is the most popular distributed version control system. It is commonly used for both open source and commercial software development, with significant benefits for individuals, teams and businesses.

Tip:

- **Git lens** is used to see the entire timeline of the project, including decisions, changes and progressions. From the moment they access the history of a project, the developer has all the context they need to understand it and start contributing.
- Developers work in every time zone. With a DVCS like Git, collaboration can happen any time while maintaining source code integrity. Using **branches**, developers can safely propose changes to production code.

1.1.2 Repositories

A repository, or Git project, encompasses the entire collection of files and folders associated with a project, along with each file's revision history. The file history appears as **snapshots** in time called commits. The commits can be organized into multiple lines of development called branches. Because Git is a DVCS, repositories are self-contained units and anyone who has a copy of the repository can access the entire codebase and its history. Using the command line or other ease-of-use interfaces, a Git repository also allows for: interaction with the history, cloning the repository, creating branches, committing, merging, comparing changes across versions of code, and more. A repository, or Git project, encompasses the entire collection of files and folders associated with a project, along with each file's revision history. The file history appears as snapshots in time

called commits. The commits can be organized into multiple lines of development called branches. Because Git is a DVCS, repositories are self-contained units and anyone who has a copy of the repository can access the entire codebase and its history. Using the command line or other ease-of-use interfaces, a Git repository also allows for: interaction with the history, cloning the repository, creating branches, committing, merging, comparing changes across versions of code, and more.

1.1.3 GitHub

GitHub hosts Git repositories and provides developers with tools to ship better code through command line features, issues (threaded discussions), pull requests, code review, or the use of a collection of free and for-purchase apps in the GitHub Marketplace. With collaboration layers like the GitHub flow, a community of 100 million developers, and an ecosystem with hundreds of integrations, GitHub changes the way software is built.

GitHub builds collaboration directly into the development process. Work is organized into repositories where developers can outline requirements or direction and set expectations for team members. Then, using the GitHub flow, developers simply create a branch to work on updates, commit changes to save them, open a pull request to propose and discuss changes, and merge pull requests once everyone is on the same page.

1.1.4 Command Line Interface

To use Git, developers use specific commands to copy, create, change, and combine code. These commands can be executed directly from the command line or by using an application like GitHub Desktop. Here are some common commands for using Git:

- **git init**: Initializes a new Git repository. Until you run this command inside a repository or directory, it's just a regular folder. Only after you input this does it accept further Git commands.
- **git config**: Short for "configure," this is most useful when you're setting up Git for the first time.
- **git help**: Forgot a command? Type this into the command line to bring up the 21 most common git commands.
- **git status**: We all get a little nervous the first time we commit our changes. This command shows you the status of changes as untracked, modified, or staged.
- **git add**: This does not add new files to your repository. Instead, it brings new files to Git's attention. After you add files, they're included in Git's "snapshots" of the repository.
- **git commit**: Git's most important command. After you make any sort of change, you input this in order to take a "snapshot" of the repository. Usually it goes **git commit -m "Message here"**.
- **git branch**: Working on multiple features at once? Use this command to list, create, or delete branches.
- **git checkout**: Literally allows you to "check out" a repository that you are not currently inside. This is a navigational command that lets you move to the repository you want to check.
- **git merge**: When you're done working on a branch, you can merge your changes back to the master branch, which is visible to all collaborators.
- **git push**: If you're working on your local computer, and want your commits to be visible online on GitHub as well, you "push" the changes up to GitHub with this command.
- **git pull**: If you're working on your local computer and want your commits to be visible online on GitHub as well, you "push" the changes up to GitHub with this command.

- **git clone** : If you're starting fresh and want to clone an existing repository from GitHub to your local computer, you can use this command to copy the repository to your computer.
- **git remote** : When you clone a repository from GitHub, it automatically creates a connection or "remote" that you can also push changes to.

Example: Contribute to an existing repository

```

1 # download a repository on GitHub to our machine
2 # Replace 'owner/repo' with the owner and name of the repository to
  clone
3 git clone https://github.com/owner/repo.git
4
5 # change into the 'repo' directory
6 cd repo
7
8 # create a new branch to store any new changes
9 git branch my-branch
10
11 # switch to that branch (line of development)
12 git checkout my-branch
13
14 # make changes, for example, edit 'file1.md' and 'file2.md' using the
  text editor
15
16 # stage the changed files
17 git add file1.md file2.md
18
19 # take a snapshot of the staging area (anything that's been added)
20 git commit -m "my snapshot"
21
22 # push changes to github
23 git push --set-upstream origin my-branch

```

Example: Start a new repository and publish on GitHub

```

1 # create a new directory, and initialize it with git-specific functions
2 git init my-repo
3
4 # change into the 'my-repo' directory
5 cd my-repo
6
7 # create the first file in the project
8 touch README.md
9
10 # git isn't aware of the file, stage it
11 git add README.md
12
13 # take a snapshot of the staging area
14 git commit -m "add README to initial commit"
15
16 # provide the path for the repository you created on github
17 git remote add origin https://github.com/YOUR-USERNAME/YOUR-REPOSITORY-
  NAME.git
18

```

```
19 # push changes to github
20 git push --set-upstream origin main
```

Example: Contribute to an existing branch on GitHub

```
1 # change into the `repo` directory
2 cd repo
3
4 # update all remote tracking branches, and the currently checked out
  branch
5 git pull
6
7 # change into the existing branch called `feature-a`
8 git checkout feature-a
9
10 # make changes, for example, edit `file1.md` using the text editor
11
12 # stage the changed file
13 git add file1.md
14
15 # take a snapshot of the staging area
16 git commit -m "edit file1"
17
18 # push changes to github
19 git push
```

1.1.5 Models for collaborative development

There are two primary ways people collaborate on GitHub:

- **Fork and pull model:** A project owner creates a master repository, and contributors fork that repository to their accounts. They clone the repository to their local machine, make changes, commit them
- **Shared repository model:** In this model, all collaborators have push access to the same repository. This is more common for small teams and organizations.

With a shared repository, individuals and teams are explicitly designated as contributors with read, write, or administrator access. This simple permission structure, combined with features like protected branches, helps teams progress quickly when they adopt GitHub.

For an open source project, or for projects to which anyone can contribute, managing individual permissions can be challenging, but a fork and pull model allows anyone who can view the project to contribute. A fork is a copy of a project under a developer's personal account. Every developer has full control of their fork and is free to implement a fix or a new feature. Work completed in forks is either kept separate, or is surfaced back to the original project via a pull request. There, maintainers can review the suggested changes before they're merged.

1.2 Pushing commits to a remote repository

1.2.1 About `git push`

The `git push` command takes two arguments.

- A remote name, for example, `origin`
- A branch name, for example, `main`

For example, the command `git push origin main` pushes the commits in the local `main` branch to the remote repository named `origin`.

Example: *Pushing changes to a remote repository*

```
1 # push changes to the remote repository
2 git push origin main
```

1.2.2 Renaming branches

To rename a branch, you'd use the same `git push` command, but you would add one more argument: the name of the new branch. For example:

Example: *Renaming a branch*

```
1 # rename the local branch to new-name
2 git push origin main:new-name
```

This pushes the `main` branch to the remote repository, but the branch is renamed to `new-name`.

Tip:

- If you want to delete a branch, you can use the `--delete` flag with the `git push` command. For example, `git push origin --delete new-name` will delete the `new-name` branch from the remote repository.
- If you want to rename the branch you're currently on, you can use the `-u` flag to set the upstream branch. For example, `git push -u origin main:new-name` will push the `main` branch to the remote repository, renaming it to `new-name`, and set the upstream branch to `new-name`.

1.2.3 Dealing with "non-fast-forward" errors

If your local copy of a repository is out of sync with, or "behind", the upstream repository you're pushing to, you'll get a message saying `non-fast-forward updates were rejected`. This means that you must retrieve, or "fetch," the upstream changes, before you are able to push your local changes.

💡 Tip: *Fetching*

Fetching means retrieving recent commits from a remote repository without merging them into your local branch. This lets you view and compare new changes before deciding how to incorporate them into your work.

To fetch the changes from the remote repository, you can use the `git fetch` command. This command retrieves the changes from the remote repository, but it doesn't merge them into your local branch. After fetching the changes, you can merge them into your local branch using the `git merge` command.

1.2.4 Pushing tags

By default, and without additional parameters, `git push` sends all matching branches that have the same names as remote branches.

To push a single tag, you can issue the same command as pushing a branch:

```
1 git push origin tag-name
```

To push all your tags, you can type the command:

```
1 git push origin --tags
```

1.2.5 Deleting a remote branch or tag

The syntax to delete a branch is a bit more arcane at first glance:

```
1 git push origin --delete branch-name
```

Note that there is a space before the colon. The command resembles the same steps you'd take to rename a branch. However, here, you're telling Git to push **nothing** into `branch-name`, effectively deleting it. Because of this, `git push` deletes the branch on the remote repository.

1.2.6 Remotes and forks

You can **fork a repository** on GitHub.

When you clone a repository you own, you provide it with a remote URL that tells Git where to fetch and push updates. If you want to collaborate with the original repository, you'd add a new remote URL, typically called `upstream`, to your local Git clone:

```
1 git remote add upstream REMOTE_URL
```

Now you can fetch updates and branches from their fork:


```

1 git fetch upstream
2 # Grab the upstream remote's branches
3 > remote: Counting objects: 75, done.
4 > remote: Compressing objects: 100% (53/53), done.
5 > remote: Total 62 (delta 27), reused 44 (delta 9)
6 > Unpacking objects: 100% (62/62), done.
7 > From https://github.com/OCTOCAT/REPO
8 > * [new branch]      main      -> upstream/main

```

When you're done making local changes, you can push your local branch to GitHub and initiate a pull request.

1.3 Getting changes from a remote repository

1.3.1 Options for getting changes

These commands are very useful when interacting with a remote repository. `clone` and `fetch` download remote code from a repository's remote URL to your local computer, `merge` is used to merge different people's work together with yours, and `pull` is a combination of `fetch` and `merge`.

1.3.2 Cloning a repository

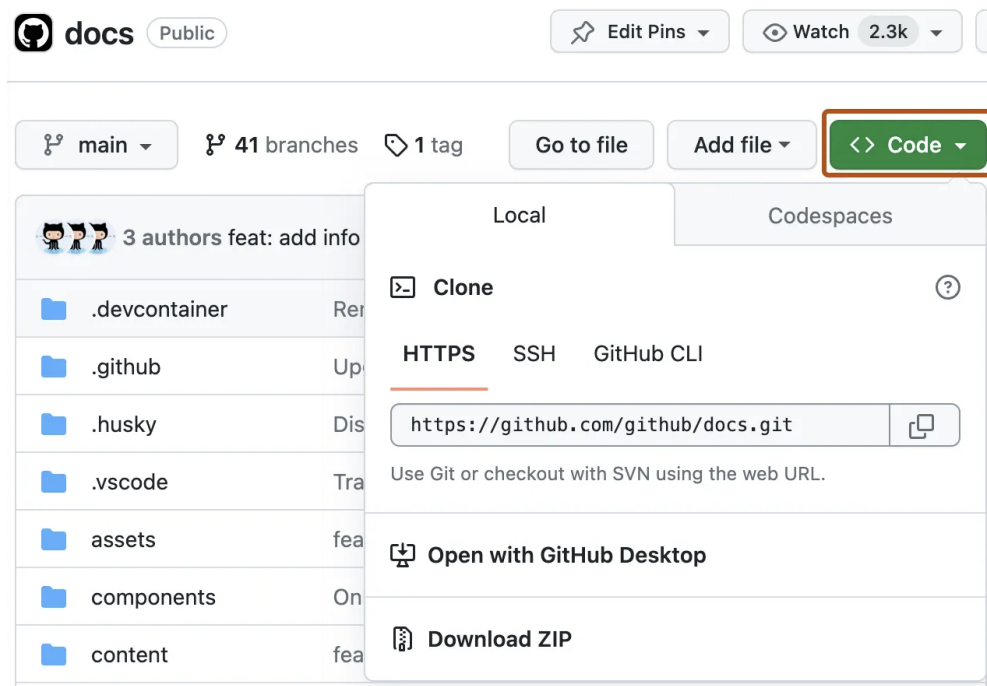
To grab a complete copy of another user's repository, use `git clone` like this:

```

1 git clone https://github.com/USERNAME/REPOSITORY.git
2 # Clones a repository to your computer

```

You can choose from different URLs when cloning a repository. While logged in to GitHub, there are URLs available on the main page of the repository when you click `<> Code`.



When you run `git clone`, the following actions occur:

- A new folder called `repo` is made
- It is initialized as a Git repository
- A remote named `origin` is created, pointing to the URL you cloned from
- All of the repository's files and commits are downloaded there
- The default branch is checked out

For every branch `foo`, a corresponding remote-tracking branch `refs/remotes/origin/foo` is created in your local repository. You can usually abbreviate such remote-tracking branch names to `origin/foo`.

1.3.3 Fetching changes from a remote repository

Use `git fetch` to retrieve new work done by other people. Fetching from a repository grabs all the new remote-tracking branches and tags without merging those changes into your own branches. If you already have a local repository with a remote URL set up for the desired project, you can grab all the new information by using `git fetch *remotename*` in the terminal.

```
1 git fetch REMOTE-NAME
2 # Fetches updates made to a remote repository
```

1.3.4 Merging changes into your local branch

Merging combines your local changes with changes made by others.

Typically, you'd merge a remote-tracking branch (i.e., a branch fetched from a remote repository) with your local branch:

```
1 git merge REMOTE-NAME/BRANCH-NAME
2 # Merges updates made online with your local work
```

1.3.5 Pulling changes from a remote repository

`git pull` is a convenient shortcut for completing both `git fetch` and `git merge` in the same command:

```
1 git pull REMOTE-NAME BRANCH-NAME
2 # Grabs online updates and merges them with your local work
```

Because `pull` performs a merge on the retrieved changes, you should ensure that your local work is committed before running the `pull` command. If you run into a **merge conflict** you cannot resolve, or if you decide to quit the merge, you can use `git merge --abort` to take the branch back to where it was in before you pulled.

1.4 Dealing with non-fast-forward errors

Sometimes, Git can't make your change to a remote repository without losing commits. When this happens, your push is refused.

If another person has pushed to the same branch as you, Git won't be able to push your changes:

Example:

```
1 git push origin main
2 > To https://github.com/USERNAME/REPOSITORY.git
3 > ! [rejected]          main -> main (non-fast-forward)
4 > error: failed to push some refs to 'https://github.com/USERNAME/
  REPOSITORY.git'
5 > To prevent you from losing history, non-fast-forward updates were
  rejected
6 > Merge the remote changes (e.g. 'git pull') before pushing again.
  See the
7 > 'Note about fast-forwards' section of 'git push --help' for
  details.
```

You can fix this by fetching and merging the changes made on the remote branch with the changes that you have made locally:

```
1 git fetch origin
2 # Fetches updates made to an online repository
3
4 git merge origin YOUR_BRANCH_NAME
5 # Merges updates made online with your local work
```

Or simply use `git pull` to perform both commands at once:

```
1 git pull origin YOUR_BRANCH_NAME
2 # Grabs online updates and merges them with your local work
```

1.5 Splitting a subfolder out into a new repository

You can turn a folder within a Git repository into a brand new repository.

Warning:

You need Git version 2.22.0 or later to follow these instructions, otherwise `git filter-repo` will not work.

If you create a new clone of the repository, you won't lose any of your Git history or changes when you split a folder into a separate repository. However, note that the new repository won't have the branches and tags of the original repository.

Steps:

1. Open Terminal.
2. Change the current working directory to the location where you want to create your new repository.
3. Clone the repository that contains the subfolder.

```
1 git clone https://github.com/USERNAME/REPOSITORY-NAME
```

4. Change the current working directory to your cloned repository.

```
1 cd REPOSITORY-NAME
```

5. To filter out the subfolder from the rest of the files in the repository, install `git-filter-repo`, then run `git filter-repo` with the following arguments.

- **FOLDER-NAME**: The folder within your project where you'd like to create a separate repository.

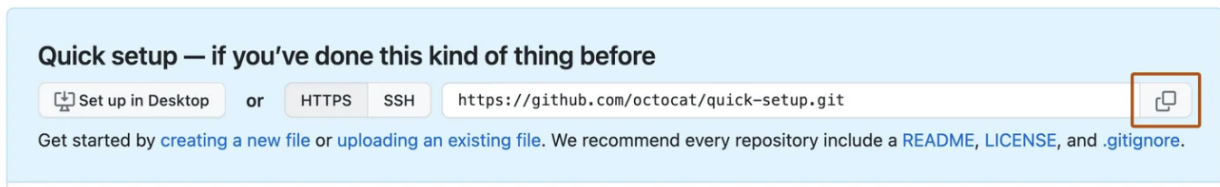
```
1 git filter-repo --path FOLDER-NAME/  
2 # Filter the specified branch in your directory and remove empty  
   commits
```

The repository should now only contain the files that were in your subfolder(s).

If you want one specific subfolder to be the new root folder of the new repository, you can use the following command:

```
1 git filter-repo --subdirectory-filter FOLDER-NAME/  
2 # Filter the specified branch in your directory and remove empty  
   commits
```

6. Create a repository on GitHub.
7. At the top of your new repository on GitHub's Quick Setup page, click ☐ to copy the remote repository URL.



Tip:

For information on the difference between HTTPS and SSH URLs, see [About remote repositories](#).

8. Add a new remote name with the URL you copied for your repository. For example, `origin` or `upstream` are two common choices.

```
1 git remote add NEW-REMOTE-NAME URL
```

9. Verify that the remote URL was added with your new repository name.

```
1 git remote -v  
2 # Verify new remote URL  
3 > origin https://github.com/USERNAME/NEW-REPOSITORY-NAME.git (  
   fetch)
```

```
4 > origin https://github.com/USERNAME/NEW-REPOSITORY-NAME.git (push
  )
```

10. Push your changes to the new repository on GitHub.

```
1 git push -u origin BRANCH-NAME
```

1.6 About Git subtree merges

Typically, a subtree merge is used to contain a repository within a repository. The "subrepository" is stored in a folder of the main repository.

The best way to explain subtree merges is to show by example. We will:

- Make an empty repository called `test` that represents our project.
- Merge another repository into it as a subtree called `Spoon-Knife`.
- The `test` project will use that subproject as if it were part of the same repository.
- Fetch updates from `Spoon-Knife` into our `test` project.

1.6.1 Setting up the empty repository for a subtree merge

1. Open Terminal.
2. Create a new directory and navigate to it.

```
1 mkdir test
2 cd test
```

3. Initialize the directory as a Git repository.

```
1 git init
2 > Initialized empty Git repository in /Users/octocat/tmp/test/.git/
```

4. Create and commit a new file.

```
1 touch .gitignore
2 git add .gitignore
3 git commit -m "initial commit"
4 > [main (root-commit) 3146c2a] initial commit
5 > 0 files changed, 0 insertions(+), 0 deletions(-)
6 > create mode 100644 .gitignore
```

1.6.2 Adding a new repository as a subtree

1. Add a new remote URL pointing to the separate project that we're interested in.

```
1 git remote add -f spoon-knife https://github.com/octocat/Spoon-
  Knife.git
2 > Updating spoon-knife
3 > warning: no common commits
4 > remote: Counting objects: 1732, done.
5 > remote: Compressing objects: 100% (750/750), done.
6 > remote: Total 1732 (delta 1086), reused 1558 (delta 967)
```

```
7 > Receiving objects: 100% (1732/1732), 528.19 KiB | 621 KiB/s, done
8 > Resolving deltas: 100% (1086/1086), done.
9 > From https://github.com/octocat/Spoon-Knife
10 > * [new branch]      main      -> Spoon-Knife/main
```

2. Merge the **Spoon-Knife** project into the local Git project. This does not change any of your files locally, but it does prepare Git for the next step.

If you're using Git 2.9 or above:

```
1 git merge -s ours --no-commit --allow-unrelated-histories spoon-
  knife/main
2 > Automatic merge went well; stopped before committing as requested
```

If you're using Git 2.8 or below:

```
1 git merge -s ours --no-commit spoon-knife/main
2 > Automatic merge went well; stopped before committing as requested
```

3. Create a new directory called **spoon-knife**, and copy the Git history of the **Spoon-Knife** project into it.

```
1 git read-tree --prefix=spoon-knife/ -u spoon-knife/main
2 > fatal: refusing to merge unrelated histories
```

4. Commit the changes to keep them safe.

```
1 git commit -m "Subtree merged in spoon-knife"
2 > [main fe0ca25] Subtree merged in spoon-knife
```

Although we've only added one subproject, any number of subprojects can be incorporated into a Git repository.

Tip:

If you create a fresh clone of the repository in the future, the remotes you've added will not be created for you. You will have to add them again using the `git remote add` command.

1.6.3 Synchronizing with updates and changes

When a subproject is added, it is not automatically kept in sync with the upstream changes. You will need to update the subproject with the following command:

```
1 git pull -s subtree REMOTE-NAME BRANCH-NAME
```

For the example above, this would be:

```
1 git pull -s subtree spoon-knife main
```

1.7 About Git rebase