

## SAGE Chatbot – Git Workflow (TEAM STANDARD)

### ◆ PHASE 1: First-time setup (DO THIS ONCE)

#### 1 Clone the repository (create local copy)

```
git clone https://github.com/<username>/<repo>.git
```

→ Creates a local folder with the full project.

---

#### 2 Enter the project folder

```
cd <repo>
```

---

#### 3 Open VS Code in this folder

```
code .
```

---

#### 4 Set Git identity (ONLY ONCE PER SYSTEM)

```
git config --global user.name "Your Name"
```

```
git config --global user.email "your_email@example.com"
```

---

#### 5 Create & activate virtual environment

```
python -m venv .venv
```

Activate:

- Windows:

```
.venv\Scripts\activate
```

- macOS/Linux:

```
source .venv/bin/activate
```

---

#### 6 Install required packages

```
pip install -r requirements.txt
```

---

### ◆ PHASE 2: BEFORE YOU START CODING (EVERY DAY)

#### 7 Check current status

```
git status
```

---

### **8 Pull latest updates from main branch**

git pull origin main

- **MANDATORY** before coding
  - Prevents conflicts and missing core files
- 

### **◆ PHASE 3: CODING PHASE**

#### **✓ Edit **only** your assigned file**

src/data\_extraction/extract\_<your\_pdf>.py

**✗** Do NOT touch:

- extract\_base.py
  - clean\_text.py
  - run\_extraction.py
- 

### **◆ PHASE 4: VERIFY YOUR WORK**

#### **9 Check what files you modified**

git status

---

#### **10 Stage **only** your extractor file**

git add src/data\_extraction/extract\_<your\_pdf>.py

**!** Do NOT use git add .

---

#### **1 1 Commit your changes**

git commit -m "Add extractor for <PDF name>"

Good commit message examples:

- Add extractor for AI Act PDF
  - Add extraction rules for syllabus document
- 

### **◆ PHASE 5: UPDATE ORIGINAL (PUSH)**

#### **1 2 Pull again (SAFETY STEP)**

```
git pull origin main
```

→ If conflicts appear:

- Fix them
- Then:

```
git add <file>
```

```
git commit -m "Resolve merge conflict"
```

---

### 1 3 Push to GitHub

```
git push origin main
```

- Username → GitHub username
- Password → **Classic Personal Access Token**