Code Quality, Security & Ethical Governance

AI Masters Capstone Project - Presentation 5

Jonathan Agustin

November 2024

Agenda

- Code Quality Assurance
- Security Scanning & Attestation
- Automating Documentation & Reporting
- Ethical Software Development: Accountability & Justice

Code Quality: Concept

- Style guidelines & linters to maintain consistency
- Code reviews as ethical checkpoints
- Clear comments & docs to expose questionable logic

Code Quality: Linting & Review Integration

- Automated linting with flake8 or Black
- Fail pipeline if code smells appear, prompting deeper review
- Combine with PR templates encouraging ethical considerations

Code Quality: Linter Workflow

```yaml
name: Lint Code

on: [push, pull_request]

jobs:
  lint:
    runs-on: ubuntu-latest
    steps:
      - uses: actions/checkout@v3
      - name: Set up Python
        uses: actions/setup-python@v4
        with:
          python-version: '3.11'
      - name: Install flake8
        run: pip install flake8
      - name: Run flake8
        run: flake8 src/ --count --select=E9,F63,F7,F82 \
             --show-source --statistics
```

Security Scanning & Attestation: Concept

- Snyk or CodeQL for real-time vulnerability checks in CI
- Sigstore Cosign to sign & verify container images
- Prevent malicious tampering and ensure traceability

Security Scanning (Snyk) - Part 1

```yaml
name: Security Scan

on: [push]

jobs:
  snyk-check:
    runs-on: ubuntu-latest
    steps:
      - uses: actions/checkout@v3
      - name: Snyk Auth
        run: snyk auth ${{ secrets.SNYK_TOKEN }}
      - name: Snyk Test
        run: snyk test --severity-threshold=high
```

Security Scanning (CodeQL) - Part 2

```yaml
name: CodeQL Analysis

on: [push, pull_request]

jobs:
  codeql:
    runs-on: ubuntu-latest
    steps:
      - uses: actions/checkout@v3
      - uses: github/codeql-action/init@v2
        with:
          languages: python
      - uses: github/codeql-action/analyze@v2
```

Attestation: Sigstore Cosign - Concept

- Sign container images with Cosign
- Verify signatures in CI before deployment
- Ensure every artifact is traceable to a verified source

Attestation (Cosign) Code Example

```yaml
name: Attest Image

on: [push]

jobs:
  attest-image:
    runs-on: ubuntu-latest
    steps:
      - uses: actions/checkout@v3
      - name: Install cosign
        run: curl -sSfL https://raw.githubusercontent.com/sigstore/cosign/main/install.sh | sh -
      - name: Sign image
        run: ./cosign sign --key ${{ secrets.COSIGN_KEY }} registry/model-service:latest
      - name: Verify image
        run: ./cosign verify registry/model-service:latest
```

Documentation & Reporting: Concept

- Model cards detailing performance, bias, limitations
- Audit logs for changes, decisions, approvals
- Automatic updates on every model iteration

Model Card Generation Code

```python
import json

def generate_model_card(model_name, performance, fairness, dataset_info):
    card = {
        "model_name": model_name,
        "performance": performance,
        "fairness": fairness,
        "dataset_info": dataset_info,
        "intended_use": "Classifier for domain X",
        "limitations": "Limited testing on low-resource languages",
        "ethical_considerations": "Potential bias in underrepresented groups"
    }
    return card

# Example usage
perf={"accuracy":0.93,"f1":0.90}
fair={"max_disparity":0.1}
data={"source":"internal_v4","size":60000}

card=generate_model_card("my_model_v3",perf,fair,data)
with open("model_card.json","w") as f:
    json.dump(card,f,indent=2)
```

Model Card CI Integration

```yaml
name: Model Card Update

on: [push]

jobs:
  model-card:
    runs-on: ubuntu-latest
    steps:
      - uses: actions/checkout@v3
      - name: Generate model card
        run: python scripts/generate_model_card.py
      - name: Upload artifact
        uses: actions/upload-artifact@v2
        with:
          name: model_card
          path: model_card.json
```

Ethical Software Development: Concept

- Ethics checklists in PR templates
- Education on historical abuses (e.g., disinformation campaigns)
- A culture where “just following orders” is unacceptable

Ethical PR Template Example

```
<!-- Pull Request Template -->

**High-Level Description of Changes:**
- Introduces a new classification threshold feature.

**Ethical Considerations:**
- Could this feature disadvantage a protected group? If yes, how are we mitigating?
- Does it align with our fairness & compliance policies?

**Security & Privacy:**
- Any new data access? If yes, is it secured and necessary?
- Did we run Snyk/CodeQL checks? Any findings?

**Testing & Validation:**
- Mention test results and model card updates.

**Approvals Needed:**
- Ethics reviewer sign-off required.
```

A Complete Ethical ML Ecosystem

- Unified approach: quality, security, transparency, ethics
- Continuous vigilance and improvement
- A practice that acknowledges the gravity of ML's impact on society

Final Thoughts

- Responsibility: Own your code's impact
- Courage: Question unethical directives
- Integrity: Build systems aligned with humanity's highest values