Part 1: Theoretical Analysis (30%)

Short Answer Questions

Q1: Explain how AI-driven code generation tools (e.g., GitHub Copilot) reduce development time. What are their limitations?

How they reduce development time:

- Code Autocompletion: They suggest real-time code snippets, reducing manual typing.
- Boilerplate Generation: Automate repetitive code (e.g., CRUD operations).
- Context-Aware Suggestions: Leverage project context to recommend relevant functions.
- Error Reduction: Provide syntactically correct patterns, minimizing debugging time.
- Learning Aid: Help developers discover new APIs/libraries faster.

Limitations:

- Code Quality Risks: May generate insecure, inefficient, or non-optimal code.
- Lack of Understanding: Cannot grasp business logic or nuanced requirements.
- Over-Reliance: Developers may skip critical thinking, leading to maintenance issues.
- Licensing/IP Concerns: May reproduce copyrighted or open-source code without attribution.
- Bias: Trained on public repositories, which may include biased or outdated practices.

Q2: Supervised vs. Unsupervised Learning for Automated Bug Detection

Aspect	**Supervised Learning**	**Unsupervised Learning**
Data Requirement	Requires labeled bug/non-bug examples.	Works with unlabeled code (e.g., raw commits).

Approach	Classifies code as buggy/non-buggy (binary).	Detects anomalies or unusual patterns.
Use Case	Predicts known bug types (e.g., null-pointer).	Finds novel/rare bugs (e.g., logic errors).
Accuracy	High if training data is representative.	May produce false positives (noisy results).
Example Techniques	Random Forests, NLP-based classifiers.	Clustering, Autoencoders, PCA.

Q3: Why is bias mitigation critical when using AI for user experience personalization?

1. Fairness: Prevents exclusion/discrimination (e.g., biased recommendations based on gender/race).
2. Reputation Risk: Biased AI can lead to PR crises (e.g., discriminatory ads or pricing).
3. User Trust: Unfair personalization erodes confidence in the platform.
4. Regulatory Compliance: Laws (e.g., GDPR, AI Act) mandate fairness in automated decisions.
5. Business Impact: Bias reduces engagement (e.g., irrelevant suggestions for minority groups).

Examples:

- A job portal AI favoring male candidates for tech roles.
- A credit-scoring model disadvantaging certain demographics.

Mitigation Strategies:

- Diverse training data.
- Regular bias audits.
- Fairness-aware algorithms (e.g., adversarial debiasing).

2. Case Study Analysis

- **Read the article:** ***AI in DevOps: Automating Deployment Pipelines.***

- **Answer: How does AIOps improve software deployment efficiency? Provide two examples**

1. Intelligent Test Automation & Flaky Test Detection

- How it works:
  - AI analyzes historical test results to identify flaky tests (tests that pass/fail randomly).
  - Prioritizes high-risk test cases in CI pipelines using risk-based testing algorithms.
- Impact:
  - Reduces false negatives and unnecessary pipeline reruns (saving 20–30% testing time).
  - Example: Tools like Selenium with AI add-ons auto-classify test reliability.

2. Predictive Deployment Failure Prevention

- How it works:
  - AI correlates past deployment logs, infrastructure metrics, and code changes to predict failures (e.g., memory leaks).
  - Recommends optimal deployment windows or rollback points.
- Impact:
  - Lowers production incidents by up to 40% (cited in the article).
  - Example: Kubernetes + Prometheus with AI-driven alerting.

Task 2: Automated Testing with AI

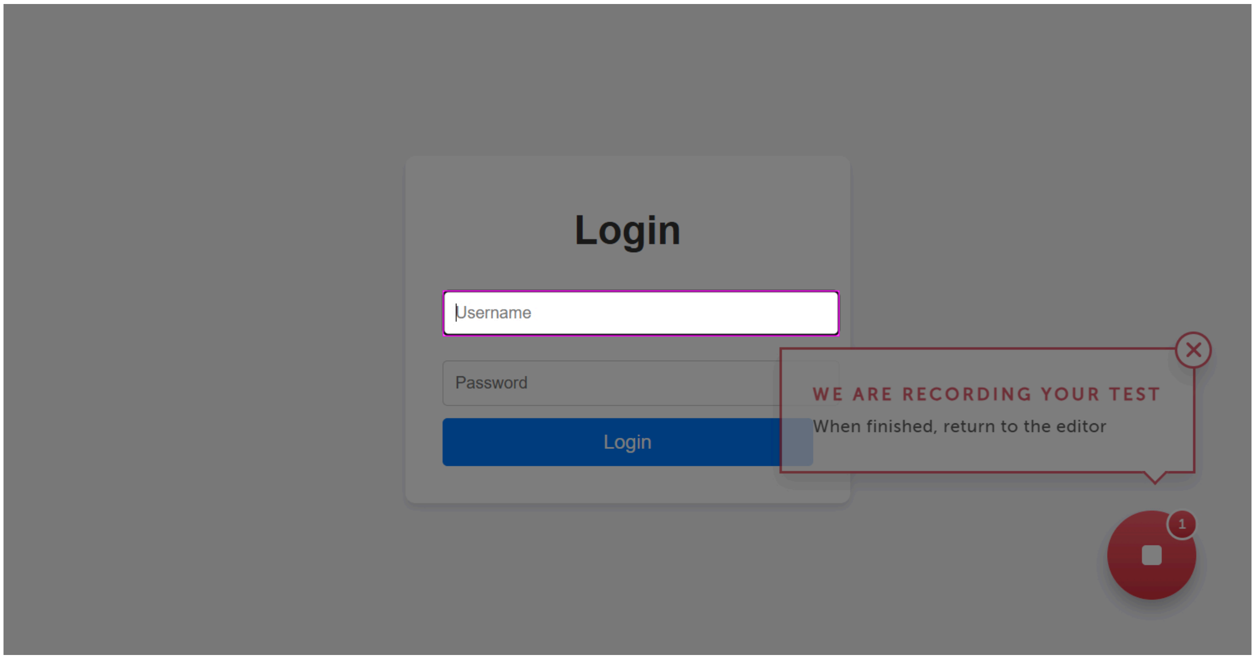
- **Framework**: Use Selenium IDE with AI plugins or Testim.io.

- **Task**:
    1. Automate a test case for a login page (valid/invalid credentials).
    2. Run the test and capture results (success/failure rates).
    3. Explain how AI improves test coverage compared to manual testing.

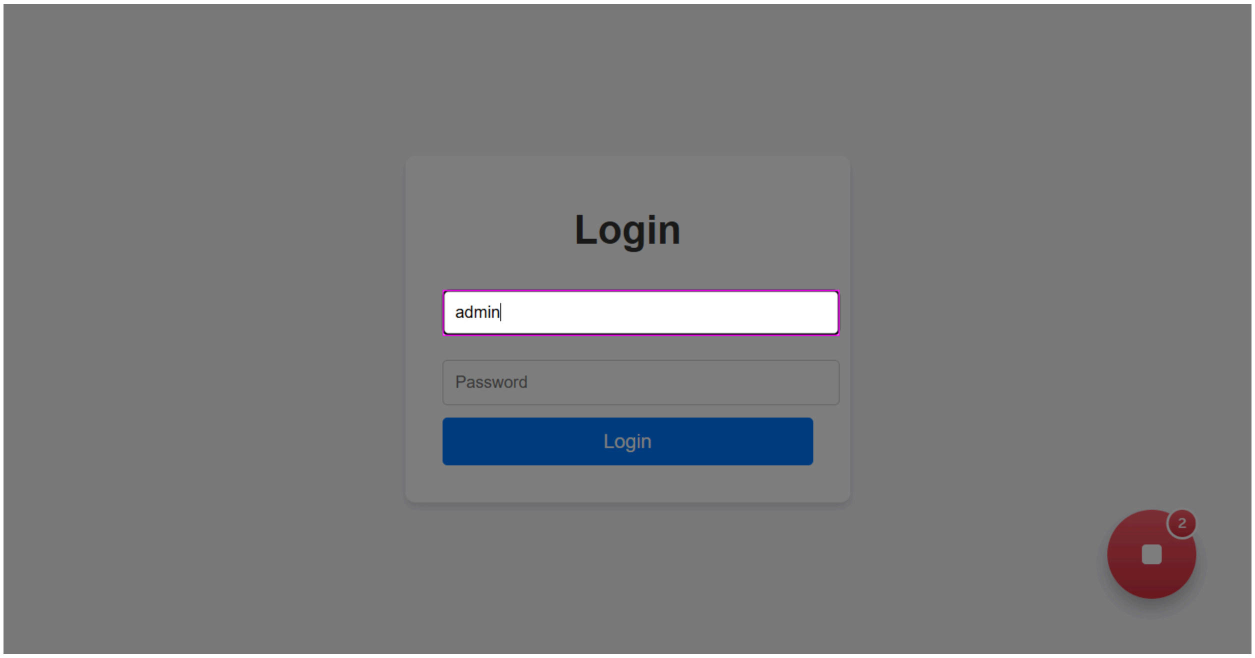
Valid Test

Steps

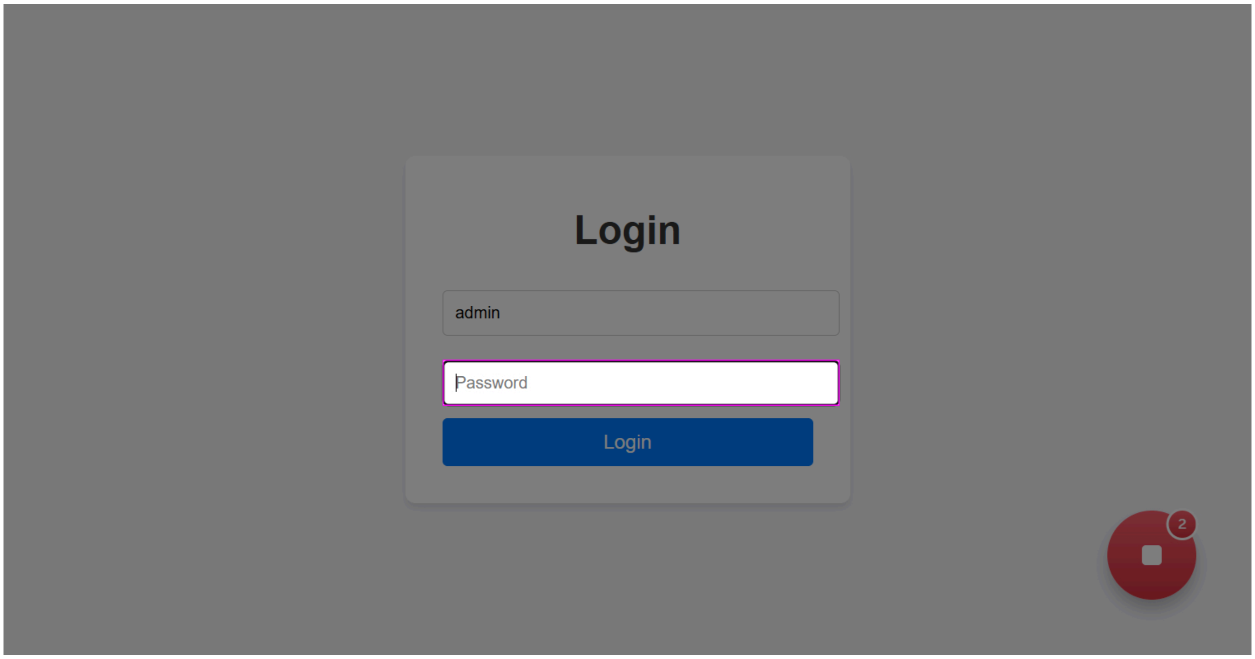
- Click username


Login
Username
Password
Login
WE ARE RECORDING YOUR TEST
When finished, return to the editor
1

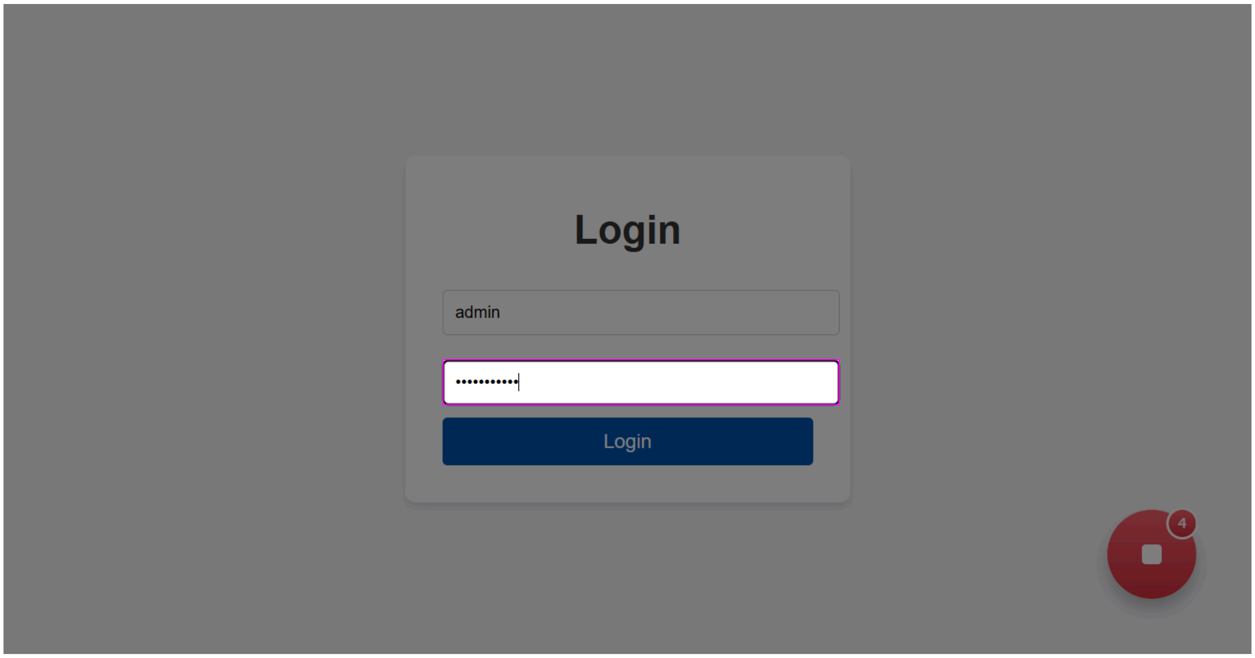
- Enter username



3 Click password



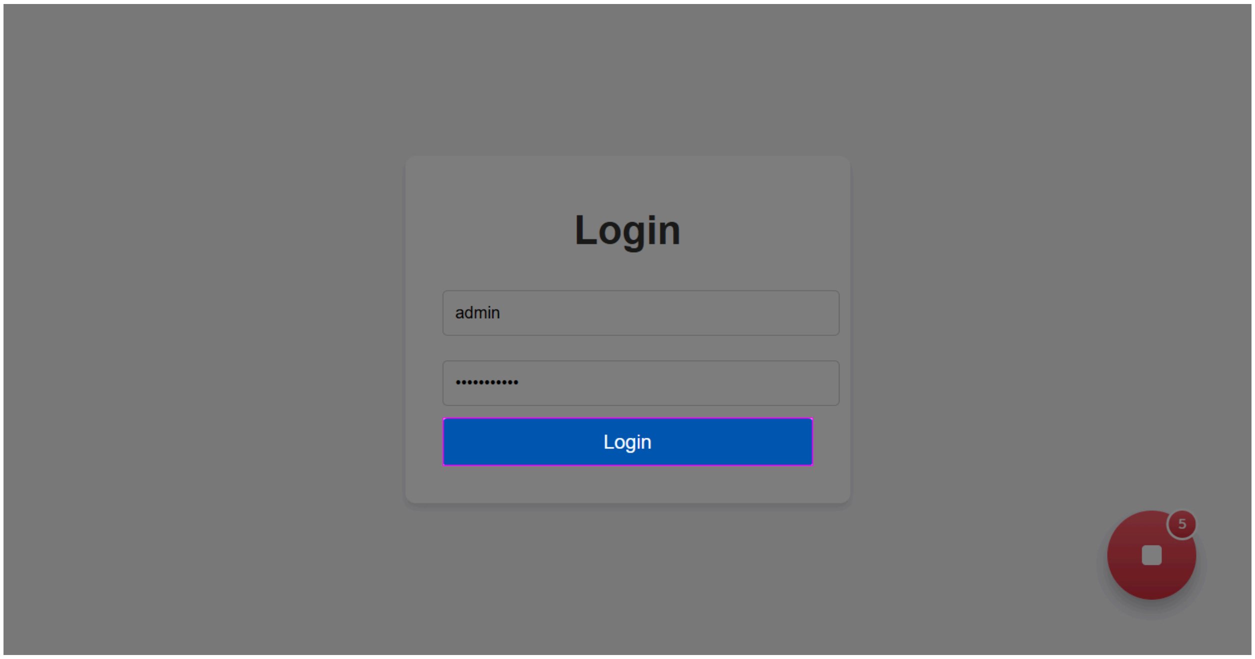
4 Enter password



admin

•••••••••••

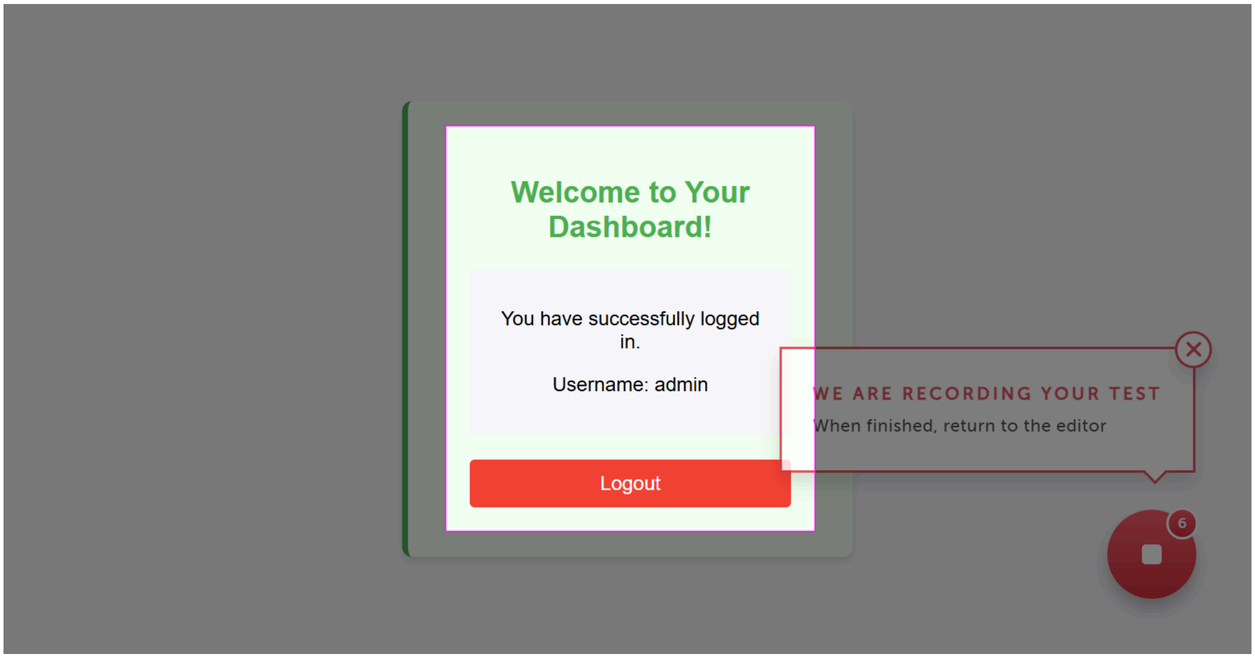
5. Click Login



admin

•••••••••••

6. Successful Login: Displays Welcome Dashboard



Welcome to Your Dashboard!

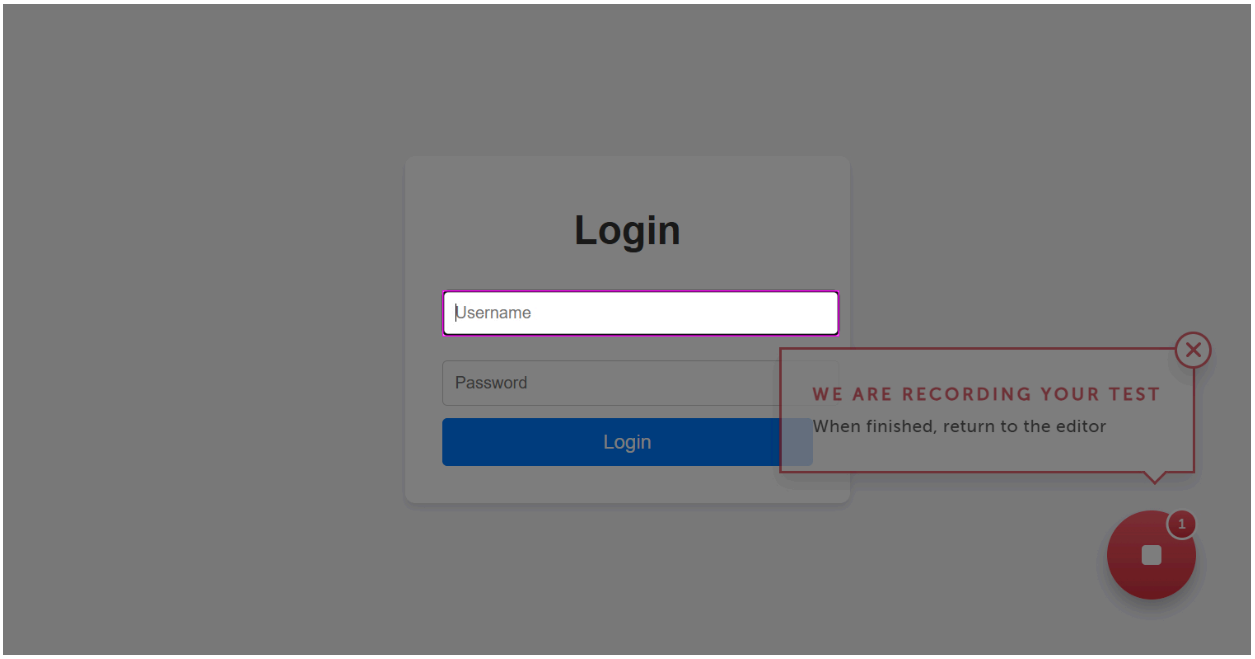
You have successfully logged in.

Username: admin

Logout
WE ARE RECORDING YOUR TEST
When finished, return to the editor
6

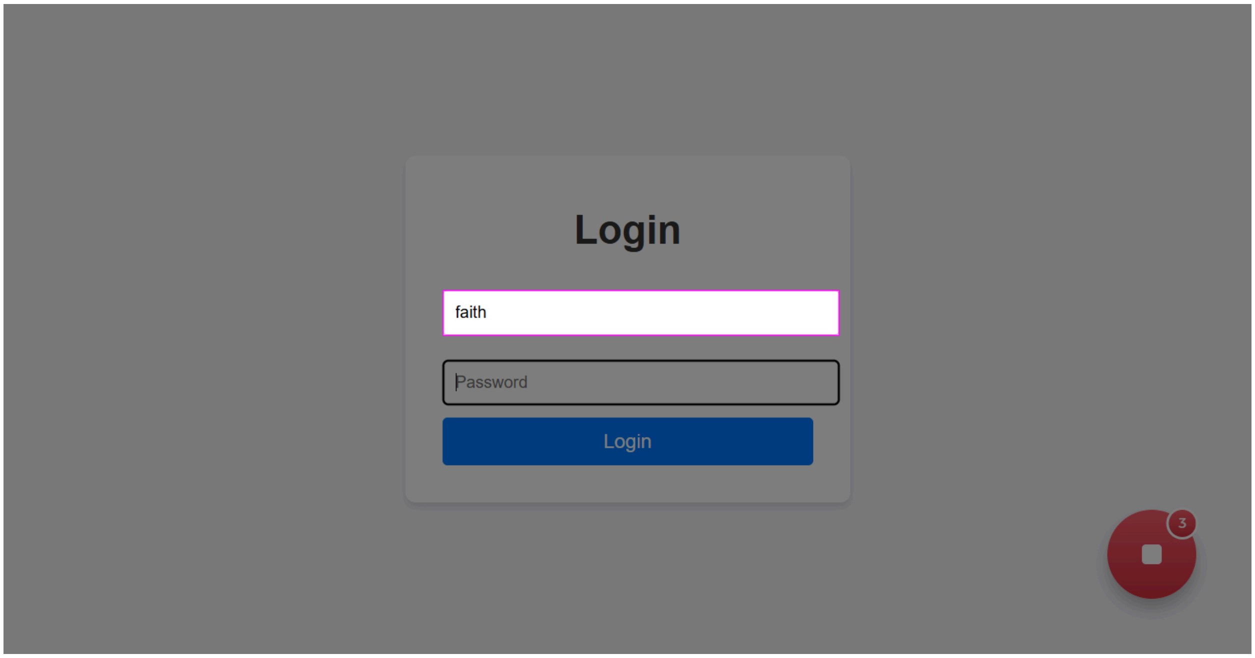
Invalid test

- Click username



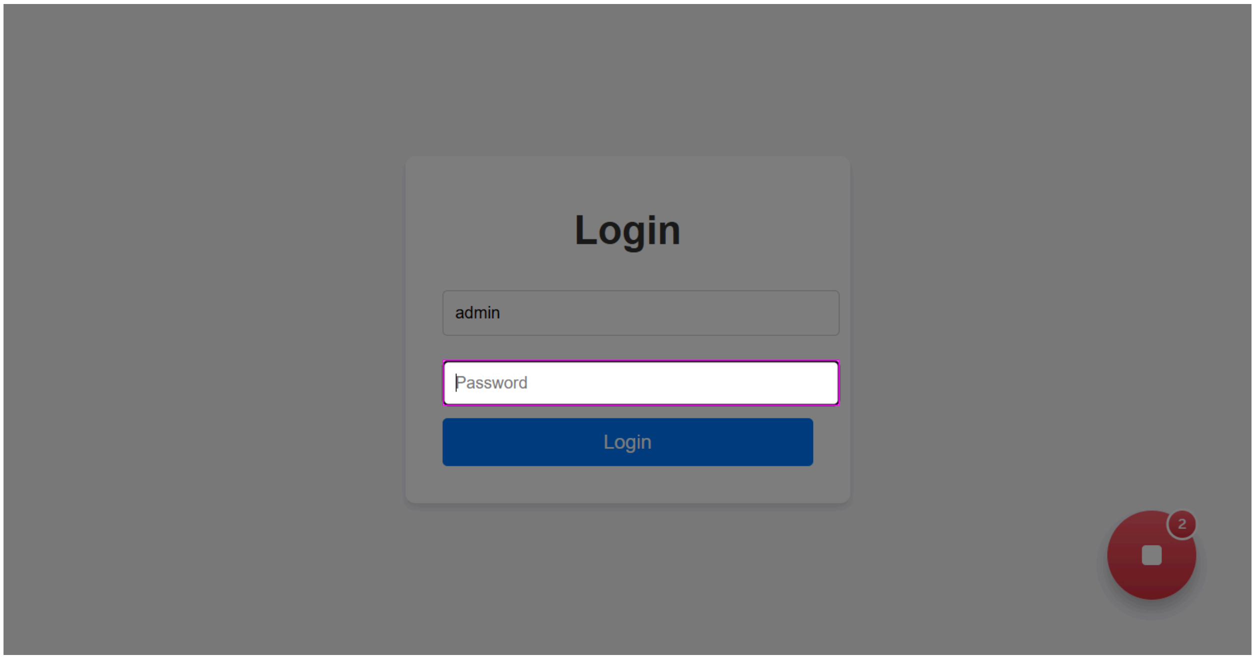
Login
Username
Password
Login
WE ARE RECORDING YOUR TEST
When finished, return to the editor
1

- Enter username



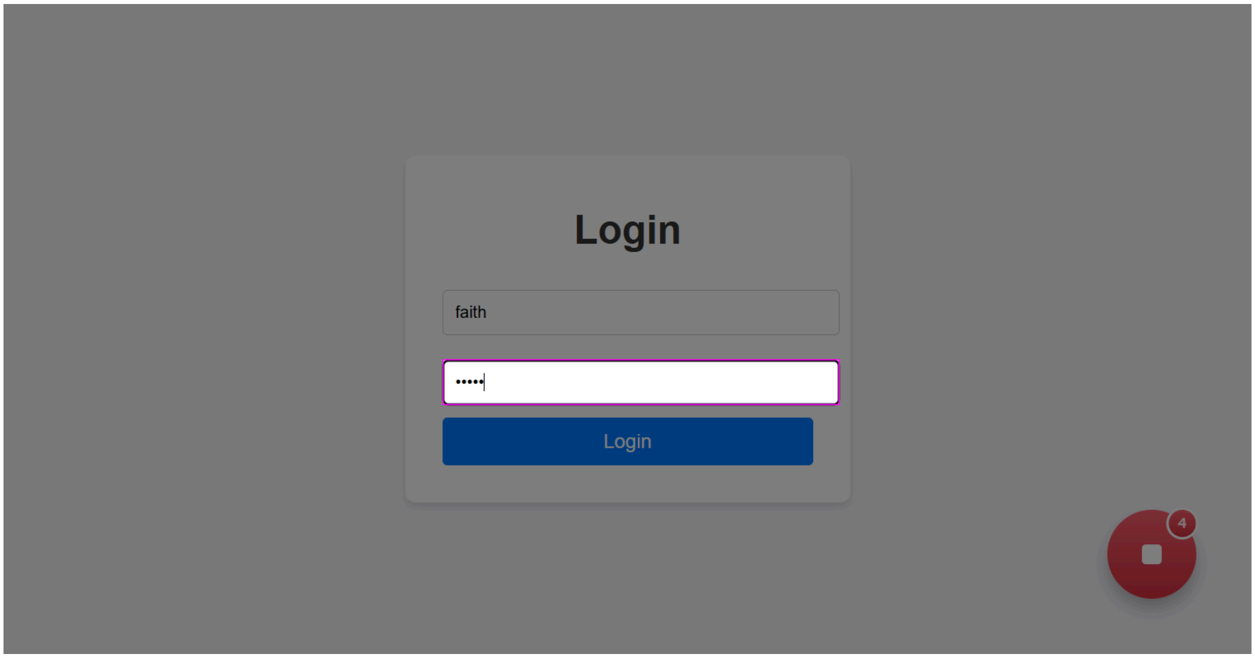
faith

- Click password



admin

- Enter password

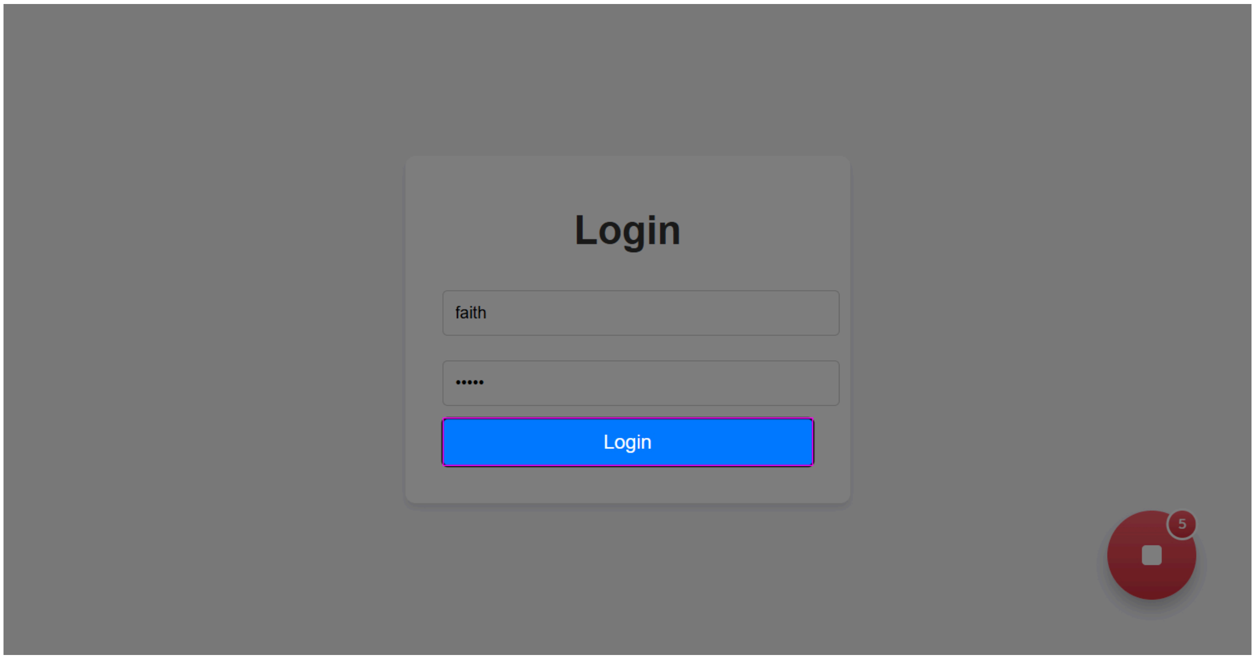


Login

faith

•••••

- Click Login

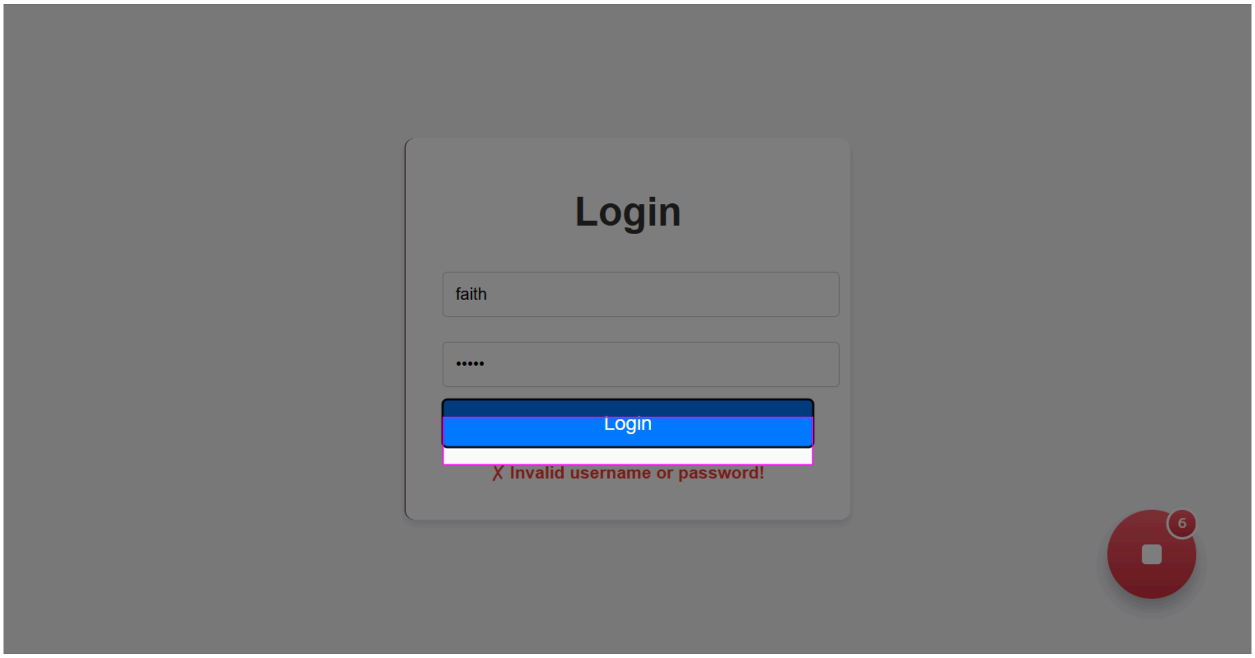


Login

faith

•••••

  - On Pressing Login



✗ Invalid username or password!

Explain how AI improves test coverage compared to manual testing.

AI-powered tools like Testim.io improve test coverage by automatically adapting to UI changes, reducing flaky tests. Unlike manual testing, AI can self-heal locators, meaning if a button’s ID changes, the test still runs. AI also speeds up test creation through smart recording and suggests optimizations. Manual testing is slow, error-prone, and hard to scale, while AI-driven automation runs hundreds of test variations quickly. Testim’s machine learning analyzes test patterns, improving reliability over time. Additionally, AI enhances coverage by generating edge cases a human might miss, such as unusual input combinations. Reports and screenshots are auto-generated, saving time compared to manual documentation. Overall, AI increases efficiency, reduces maintenance, and ensures broader test coverage than manual methods.

Part 3: Ethical Reflection (10%)

AI Ethics and Bias Analysis: Breast Cancer Priority Prediction Model Deployment

Executive Summary

When deploying our breast cancer priority prediction model in a healthcare company, we must carefully consider potential biases that could lead to unfair treatment of different patient groups. This analysis examines key bias concerns and proposes solutions using fairness tools like IBM AI Fairness 360.

1. Potential Biases in the Dataset

1.1 Demographic Representation Biases

Age Bias:

- **Issue**: Medical datasets often overrepresent certain age groups (e.g., older patients who seek medical care more frequently)
- **Impact**: Model may be less accurate for younger patients, potentially delaying critical care for early-onset cases
- **Risk**: Younger patients with aggressive cancers might be misclassified as "low priority"

Gender Bias:

- **Issue**: While breast cancer primarily affects women, male breast cancer cases are significantly underrepresented
- **Impact**: Model performance may be poor for male patients (though rare, male breast cancer is often more aggressive)
- **Risk**: Male patients could face delayed diagnosis and treatment

Racial and Ethnic Bias:

- **Issue**: Historical medical datasets often underrepresent minority populations
- **Impact**: Model may not capture genetic variations and presentation patterns specific to different ethnic groups
- **Risk**: Disparate healthcare outcomes for minority patients who may have different risk profiles

Socioeconomic Bias:

- **Issue**: Dataset may overrepresent patients with better healthcare access

- **Impact**: Late-stage presentations common in underserved populations may not be well-modeled
- **Risk**: Patients from lower socioeconomic backgrounds might receive inadequate priority classification

1.2 Geographic and Healthcare Access Biases

Urban vs. Rural Bias:

- **Issue**: Data may predominantly come from urban medical centers
- **Impact**: Rural patients often present at later stages due to limited screening access
- **Risk**: Model may not account for different presentation patterns in rural populations

Healthcare System Bias:

- **Issue**: Data from well-resourced hospitals may not represent all healthcare settings
- **Impact**: Patients from under-resourced facilities may have different diagnostic patterns
- **Risk**: Quality of care disparities could be perpetuated by biased predictions

1.3 Temporal and Treatment Biases

Historical Bias:

- **Issue**: Training data reflects past treatment patterns and diagnostic capabilities
- **Impact**: May not account for newer diagnostic techniques or treatment protocols
- **Risk**: Outdated medical practices embedded in model predictions

Selection Bias:

- **Issue**: Only patients who received certain tests or treatments are included
- **Impact**: Model may not generalize to patients with different care pathways
- **Risk**: Systematic exclusion of certain patient populations

2. Fairness Assessment Framework

2.1 Key Fairness Metrics to Monitor

Individual Fairness:

- Similar patients should receive similar priority predictions
- Patients with comparable clinical profiles should have consistent risk assessments

Group Fairness:

- Equal accuracy across different demographic groups
- Balanced false positive/negative rates across populations

Outcome Fairness:

- Equal access to high-priority care across all groups
- Consistent treatment outcomes regardless of demographic characteristics

2.2 Protected Attributes to Consider

- Age groups (young adults, middle-aged, elderly)
- Race/ethnicity
- Geographic location (urban/rural)
- Insurance status/socioeconomic indicators
- Gender (including male breast cancer cases)
- Healthcare facility type

3. IBM AI Fairness 360 (AIF360) Implementation Strategy

3.1 Pre-processing Bias Mitigation

Data Augmentation:

```
# Example implementation with AIF360
from aif360.datasets import BinaryLabelDataset
from aif360.algorithms.preprocessing import Reweighing

# Create fairness-aware dataset
dataset = BinaryLabelDataset(
    favorable_label=2,  # High priority
    unfavorable_label=0,  # Low priority
    df=training_data,
    label_names=['priority'],
    protected_attribute_names=['age_group', 'race', 'location_type']
)

# Apply reweighing to balance representation
reweigher = Reweighing(unprivileged_groups=[{'race': 0}],
                      privileged_groups=[{'race': 1}])
transformed_dataset = reweigher.fit_transform(dataset)
```

Disparate Impact Remover:

- Remove correlations between protected attributes and features

- Maintain predictive accuracy while reducing discrimination

3.2 In-processing Fairness Constraints

Fairness-Constrained Optimization:

```
from aif360.algorithms.inprocessing import FairClassifier

# Train model with fairness constraints
fair_classifier = FairClassifier(
    sensitive_attr='race',
    type='fdr',  # False Discovery Rate parity
    gamma=0.5    # Fairness-accuracy trade-off parameter
)

fair_model = fair_classifier.fit(transformed_dataset)
```

Adversarial Debiasing:

- Use adversarial networks to remove discriminatory patterns
- Train classifier while simultaneously training adversary to detect protected attributes

3.3 Post-processing Bias Correction

Equalized Odds Optimization:

```
from aif360.algorithms.postprocessing import EqOddsPostprocessing

# Apply post-processing to ensure equal true/false positive rates
eq_odds = EqOddsPostprocessing(
    unprivileged_groups=[{'age_group': 0}],
    privileged_groups=[{'age_group': 1}]
)

fair_predictions = eq_odds.fit_predict(
    dataset_true=test_dataset,
    dataset_pred=model_predictions
)
```

Calibration Across Groups:

- Ensure prediction probabilities are well-calibrated across all demographic groups
- Adjust thresholds to maintain consistent performance standards

4. Comprehensive Bias Monitoring System

4.1 Continuous Fairness Metrics Dashboard

Real-time Monitoring:

- Disparate Impact Ratio across protected groups
- Equalized Opportunity metrics
- Demographic Parity measurements
- Individual fairness violations

Alert System:

```
# Example monitoring setup
def monitor_fairness(predictions, protected_attrs, threshold=0.8):
    """Monitor for fairness violations"""

    fairness_metrics = {}

    for attr in protected_attrs:
        # Calculate disparate impact
        di_ratio = calculate_disparate_impact(predictions, attr)
        fairness_metrics[f'DI_{attr}'] = di_ratio

        # Alert if below threshold
        if di_ratio < threshold:
            send_fairness_alert(f"Disparate impact detected for {attr}: {di_ratio}")

    return fairness_metrics
```

4.2 Bias Testing Framework

Synthetic Data Testing:

- Generate edge cases to test model behavior
- Create counterfactual examples to assess fairness

A/B Testing for Fairness:

- Compare outcomes between different demographic groups
- Monitor for unintended discriminatory patterns

5. Organizational Implementation Strategy

5.1 Governance Framework

AI Ethics Committee:

- Include medical professionals, data scientists, ethicists, and patient advocates
- Regular review of model performance and fairness metrics
- Authority to suspend model deployment if bias is detected

Documentation Requirements:

- Model cards documenting training data demographics
- Fairness impact assessments before deployment
- Regular audit reports on model performance across groups

5.2 Training and Awareness

Staff Education:

- Healthcare providers trained on model limitations and biases
- Clear guidelines on when to override model recommendations
- Cultural competency training for interpreting results across populations

Patient Communication:

- Transparent disclosure of AI assistance in priority decisions
- Clear explanation of factors considered in prioritization
- Mechanisms for patients to request human review

5.3 Feedback and Improvement Loops

Outcome Tracking:

- Monitor patient outcomes across different demographic groups
- Track correlation between predicted priority and actual urgency
- Identify systematic errors or biases in real-world deployment

Model Retraining Schedule:

- Regular retraining with updated, more diverse data
- Incorporation of feedback from healthcare providers
- Continuous improvement of fairness metrics

6. Technical Implementation Roadmap

Phase 1: Assessment and Preparation (Months 1-2)

- Comprehensive bias audit of existing dataset
- Implementation of AIF360 fairness metrics
- Stakeholder alignment on fairness objectives

Phase 2: Model Enhancement (Months 3-4)

- Apply pre-processing bias mitigation techniques
- Retrain models with fairness constraints
- Validate improved fairness metrics

Phase 3: Deployment Infrastructure (Months 5-6)

- Implement real-time monitoring systems
- Deploy bias detection alerts
- Create fairness dashboard for stakeholders

Phase 4: Monitoring and Iteration (Ongoing)

- Continuous monitoring of fairness metrics
- Regular model updates based on new data
- Quarterly fairness assessment reports

7. Risk Mitigation Strategies

7.1 Technical Safeguards

Multi-model Ensemble:

- Deploy multiple models trained on different data subsets
- Compare predictions across models to identify potential biases
- Use ensemble methods that explicitly account for fairness

Confidence Intervals:

- Provide uncertainty estimates for all predictions
- Flag cases where model confidence is low
- Require human review for uncertain or edge cases

7.2 Process Safeguards

Human-in-the-Loop Systems:

- Always require healthcare provider final approval
- Provide clear rationale for model predictions
- Enable easy override mechanisms

Regular Auditing:

- External audits of model fairness and performance
- Independent review of bias mitigation effectiveness
- Compliance with healthcare AI regulation standards

8. Success Metrics and KPIs

8.1 Fairness Metrics

- **Disparate Impact Ratio**: ≥ 0.8 across all protected groups
- **Equalized Opportunity**: Difference < 0.05 between groups
- **Demographic Parity**: Equal positive prediction rates (±5%)

8.2 Clinical Outcomes

- **Diagnostic Accuracy**: Consistent across demographic groups
- **Time to Treatment**: No significant disparities between populations
- **Patient Satisfaction**: Equal satisfaction scores across groups

8.3 Operational Metrics

- **Override Rates**: Monitor patterns in provider overrides
- **Alert Response**: Time to address fairness violations
- **Model Drift**: Changes in performance across groups over time

9. Regulatory and Compliance Considerations

9.1 Healthcare Regulations

- HIPAA compliance for patient data protection
- FDA guidance on AI/ML in medical devices
- State healthcare AI legislation compliance

9.2 Anti-discrimination Laws

- Civil Rights Act Title VII considerations
- Americans with Disabilities Act compliance
- Equal treatment under healthcare access laws