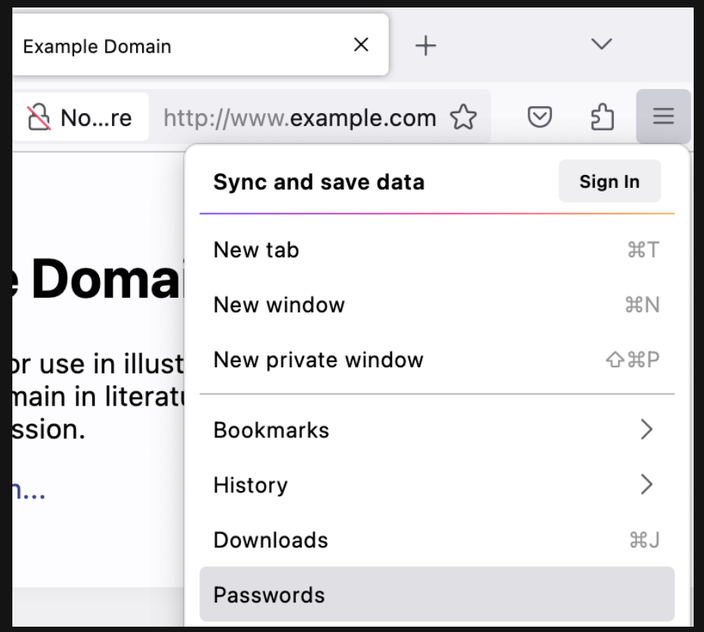
RippleGUItester: Change-Aware Exploratory Testing

Yanqi Su*, Michael Pradel#, Chunyang Chen*

* TU Munich, # CISPA

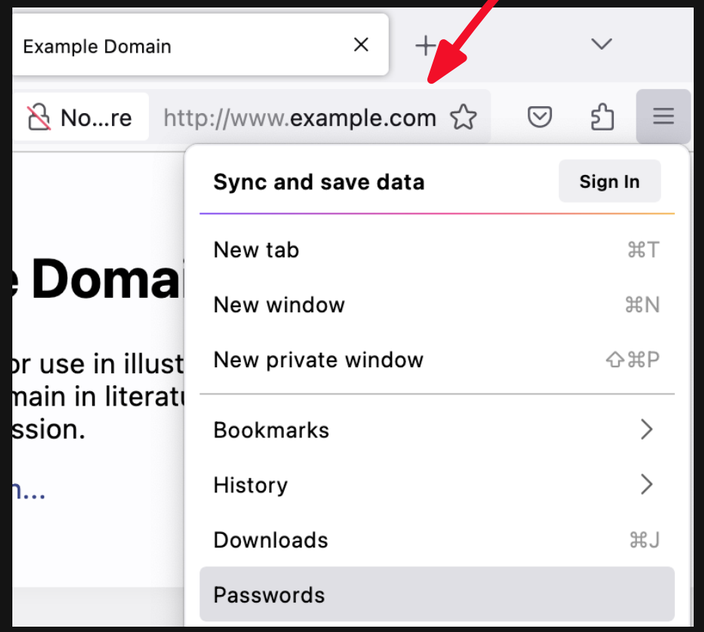
Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL



Motivating Example (Firefox)

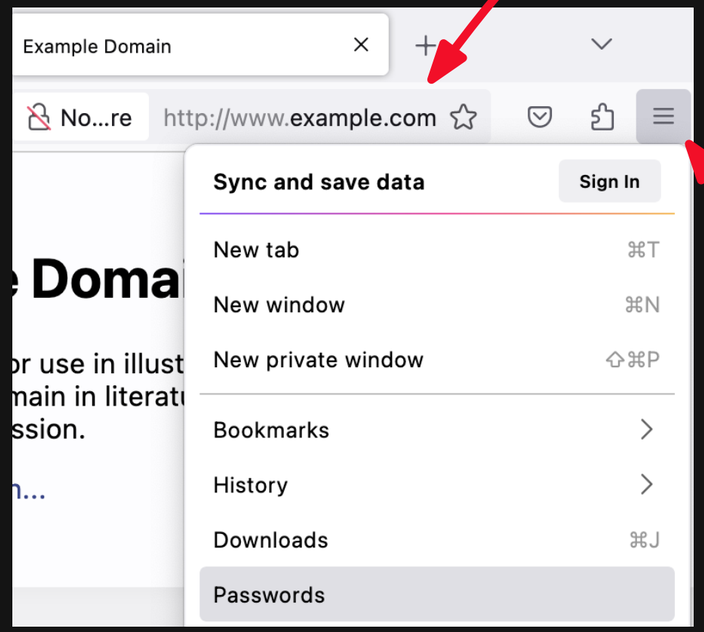
Issue 1858633: Password page opens with incorrect URL



Go to example.com

Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

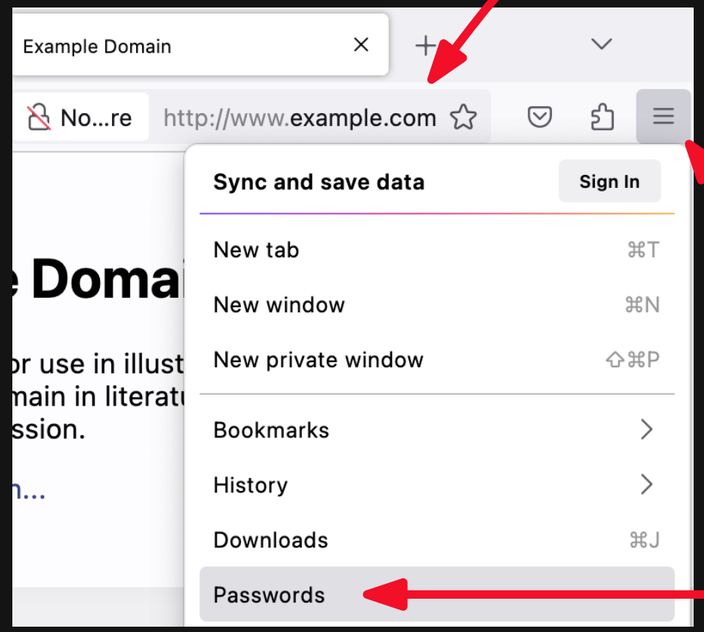


Go to example.com

Click on menu

Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL



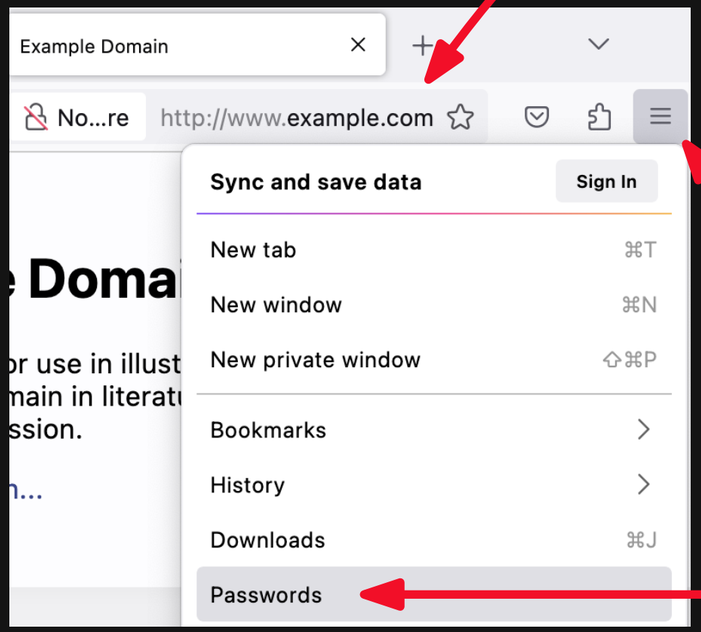
Go to example.com

Click on menu

Click on Passwords

Motivating Example (Firefox)

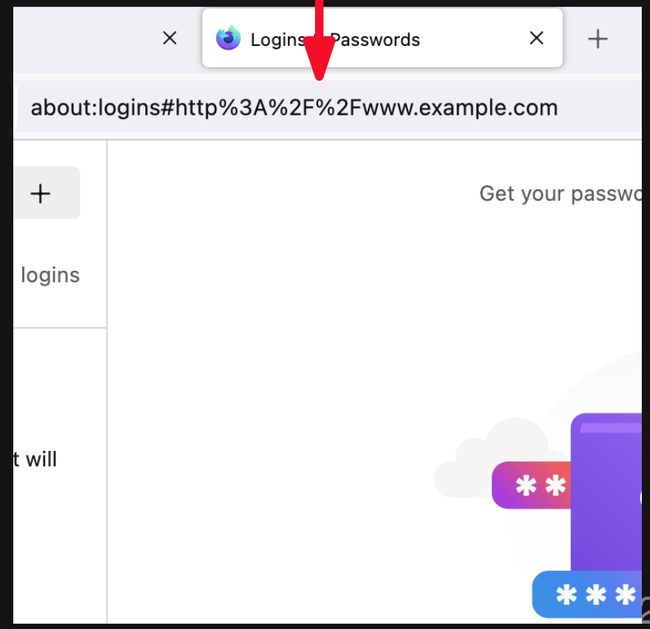
Issue 1858633: Password page opens with incorrect URL



Go to example.com

Click on menu

Click on Passwords



Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

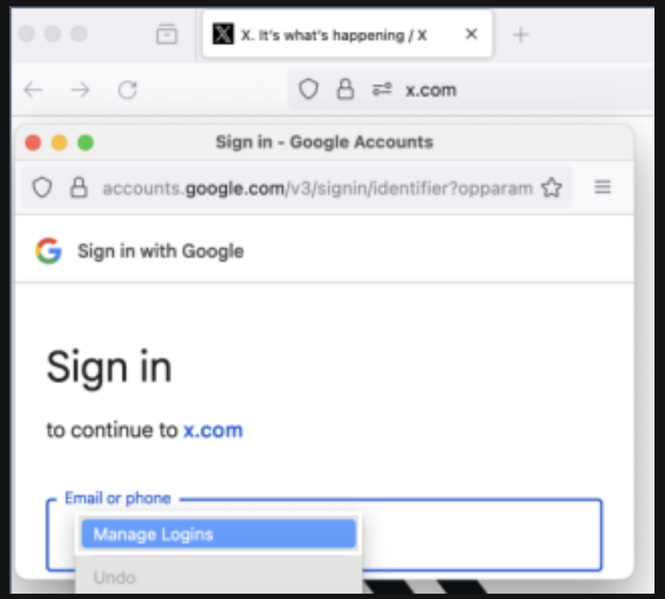
↳ PR to fix the issue

Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

↳ PR to fix the issue

↳ Issue 1937085: Password page opens in pop-up

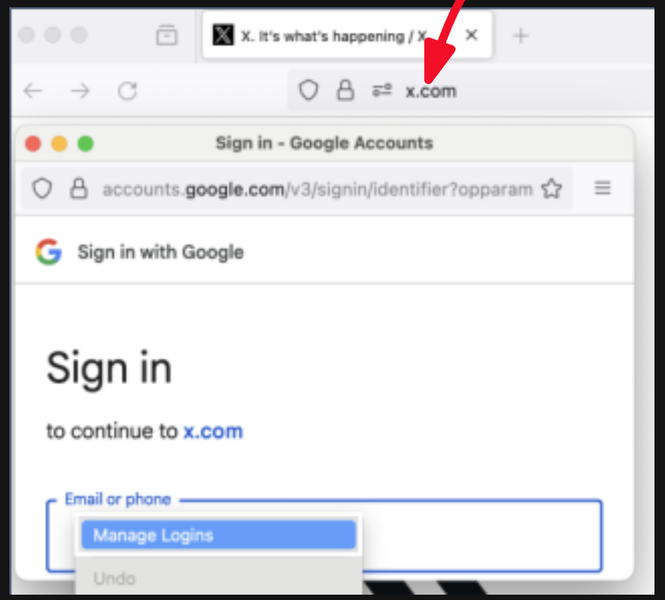


Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

↳ PR to fix the issue

↳ Issue 1937085: Password page opens in pop-up



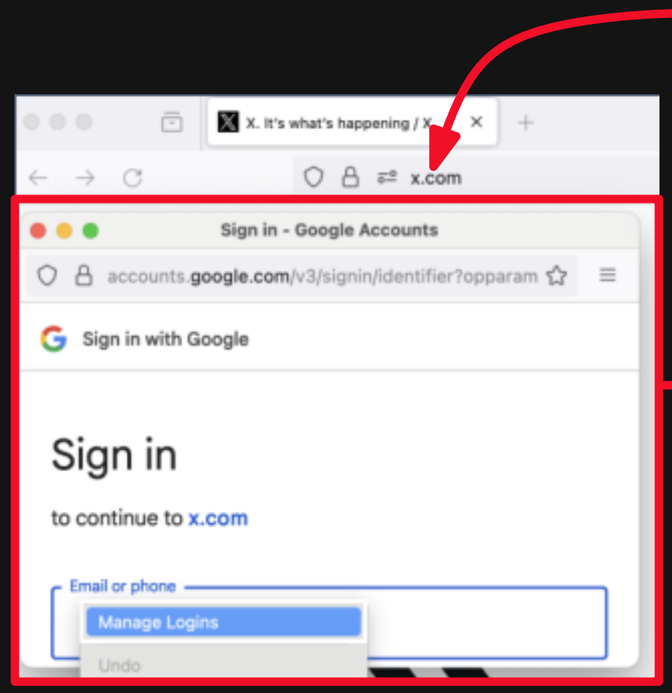
Open site with
pop-up login window

Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

↳ PR to fix the issue

↳ Issue 1937085: Password page opens in pop-up



Open site with pop-up login window

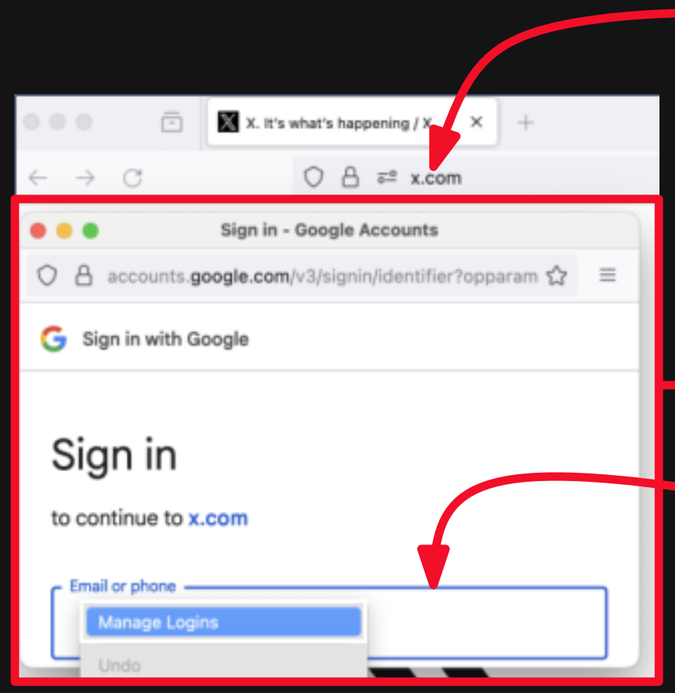
Click sign-in button

Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

↳ PR to fix the issue

↳ Issue 1937085: Password page opens in pop-up



Open site with pop-up login window

Click sign-in button

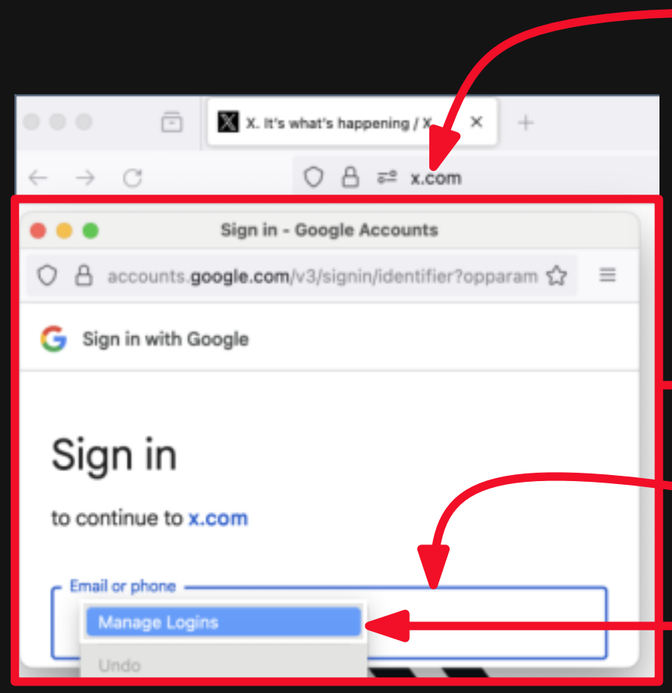
Right-click input field

Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

↳ PR to fix the issue

↳ Issue 1937085: Password page opens in pop-up



Open site with pop-up login window

Click sign-in button

Right-click input field

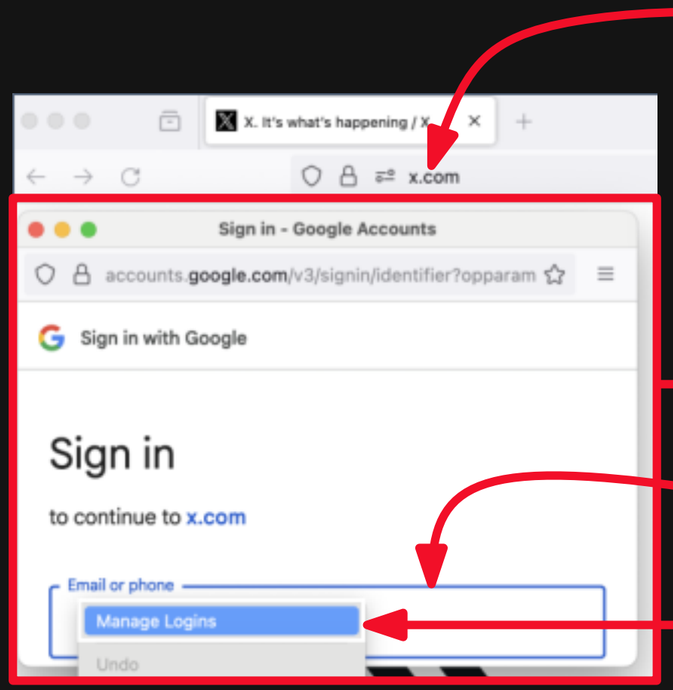
Select Manage Logins

Motivating Example (Firefox)

Issue 1858633: Password page opens with incorrect URL

↳ PR to fix the issue

↳ Issue 1937085: Password page opens in pop-up

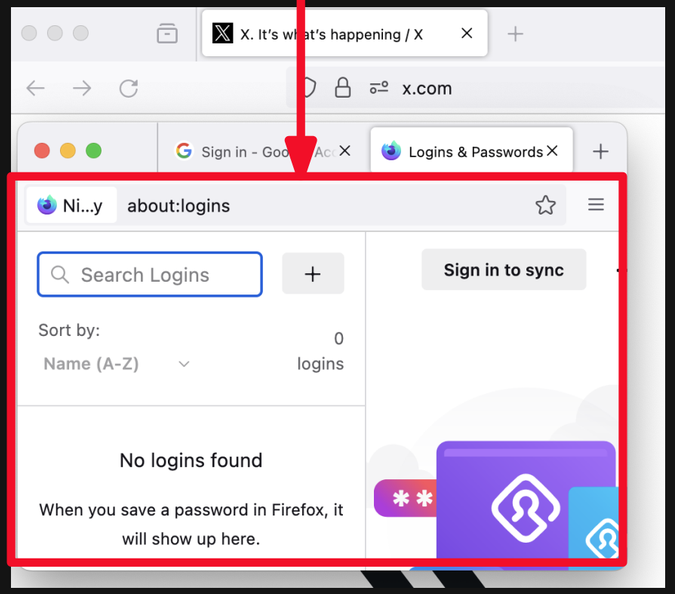


Open site with pop-up login window

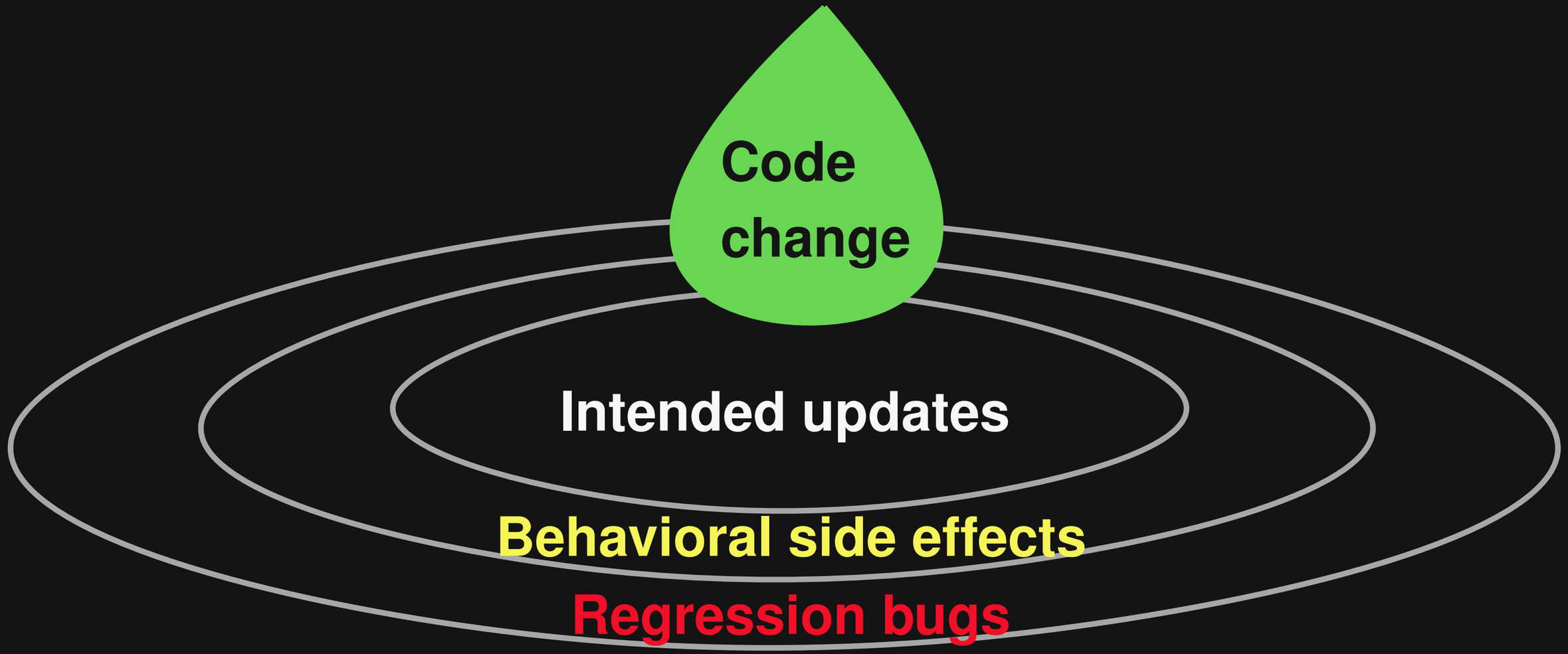
Click sign-in button

Right-click input field

Select Manage Logins



The Ripple Effect of Code Changes



How Often Does This Happen?

97,347 PRs merged into Firefox

Search for issues introduced by these PRs

11,910 PRs (12.2%) of PRs introduced at least one new bug

(Conservative estimate due to missing links and unknown bugs)

Why Is It So Hard?

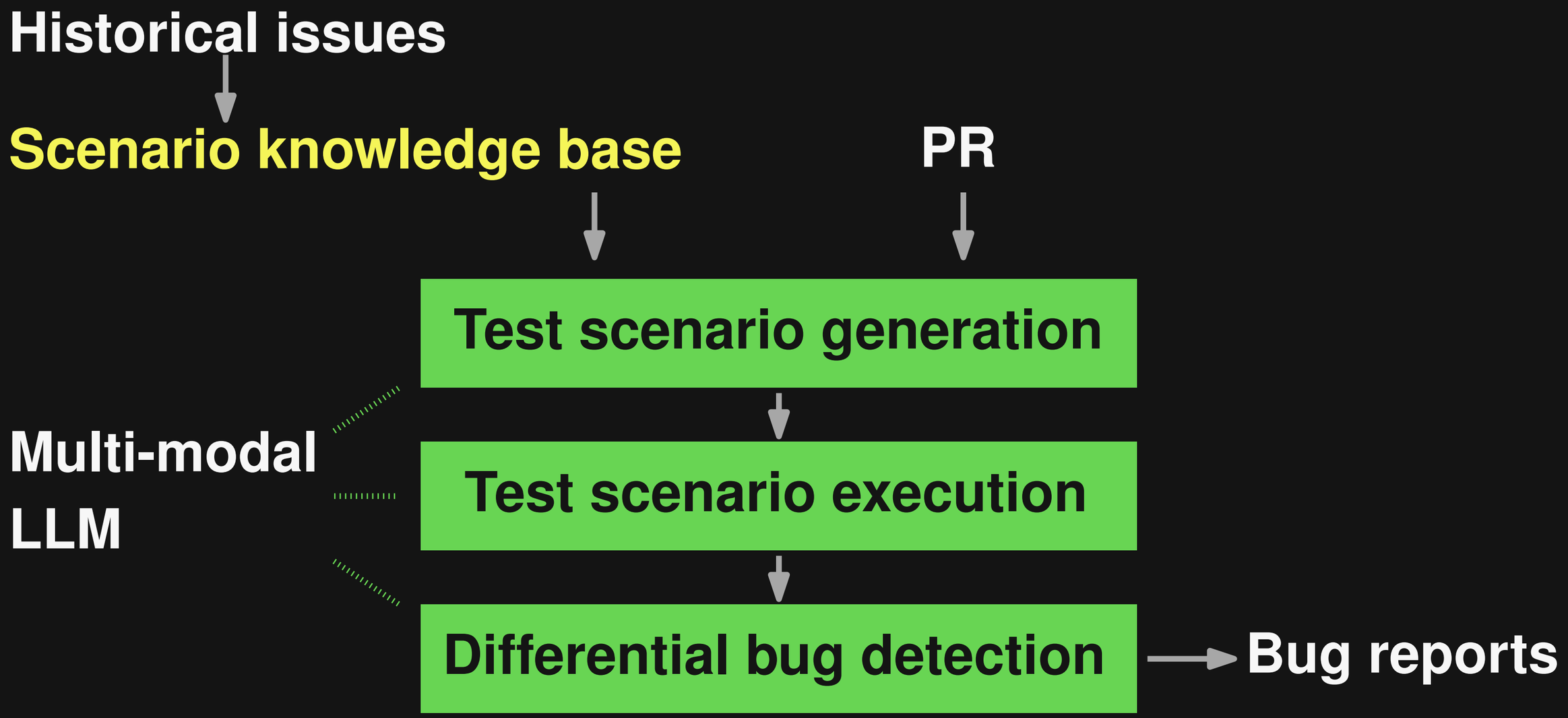
Two reasons:

1) **Challenging trigger steps**

- Specific event sequences
- Specific test data

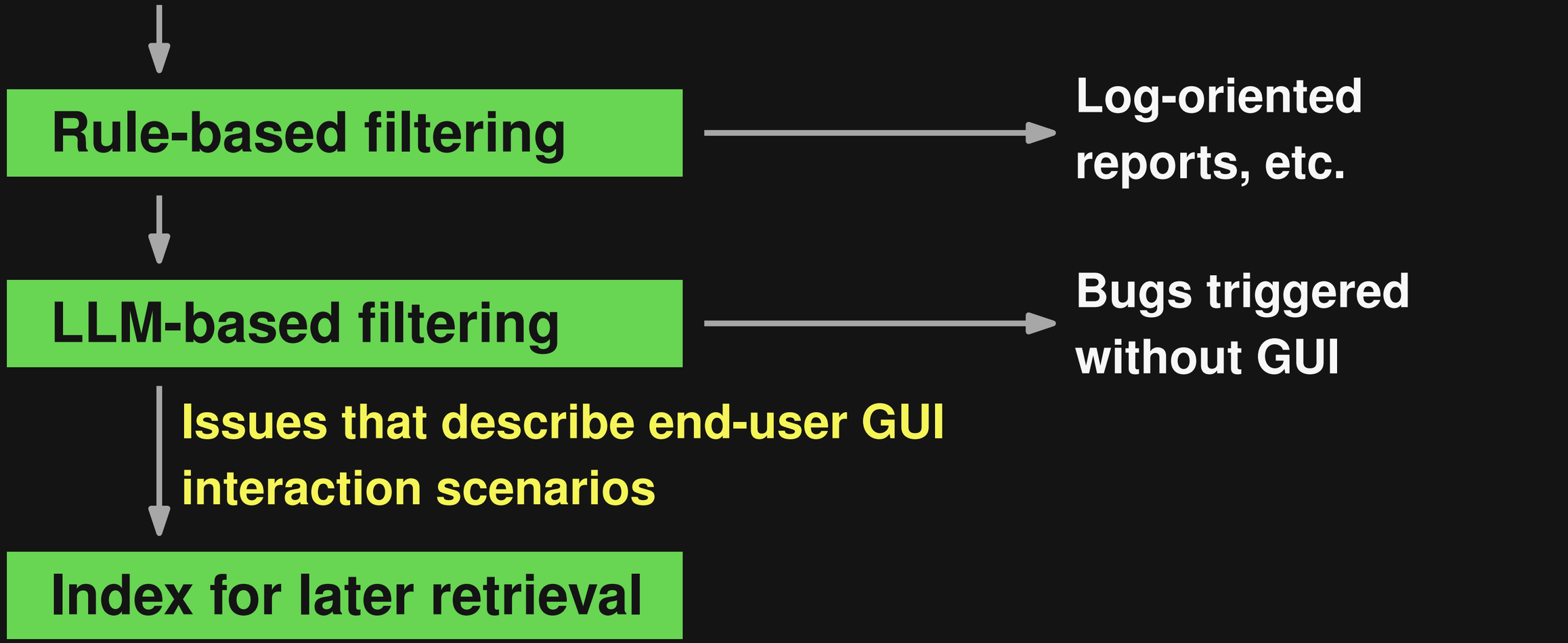
2) **Hard-to-predict oracles**

Overview of RippleGUItester



Scenario Knowledge Base

All issues in project history



Test Scenario Generation

Pipeline of structured LLM queries:

1) Analyze change impact and generate test scenarios
2) Enrich scenarios using retrieved past scenarios
3) Enhance test data

Test Scenario Generation

Pipeline of structured LLM queries:

1) **Analyze change impact and generate test scenarios**
2) **Enrich scenarios using retrieved past scenarios**
3) **Enhance test data**

Example:

- Understands that change affects how Passwords page is opened
- But: Unaware of all entry points to the Passwords page

Test Scenario Generation

Pipeline of structured LLM queries:

1) **Analyze change impact and generate test scenarios**
2) **Enrich scenarios using retrieved past scenarios**
3) **Enhance test data**

Example:

- Finds event sequences that reach Passwords page
- But: Input data remains vague (“from site A, open pop-up window”)

Test Scenario Generation

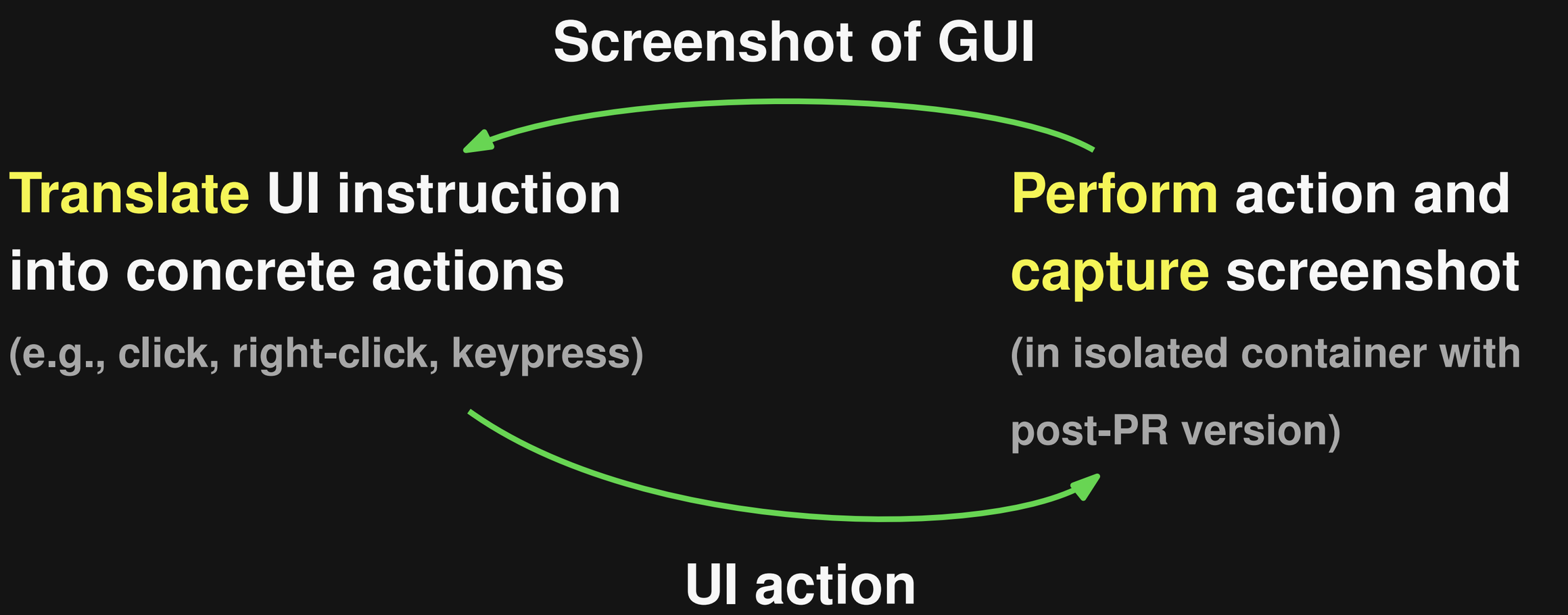
Pipeline of structured LLM queries:

1) **Analyze change impact and generate test scenarios**
2) **Enrich scenarios using retrieved past scenarios**
3) **Enhance test data**

Example:

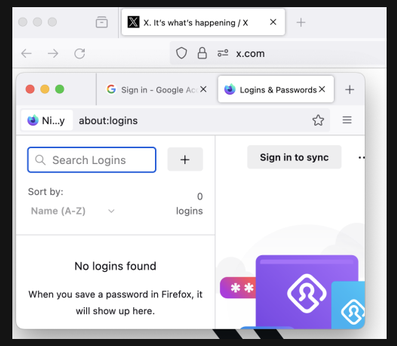
- Instantiates abstract input data requirements
- Generates a script that reliably creates a pop-up

Tests Scenario Execution

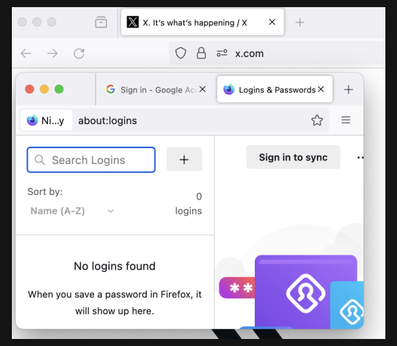


Differential Bug Detection

Pre-PR execution



Post-PR execution



Pixel-based comparison

LLM-based bug detection

LLM-based bug deduplication and filtering

Bug report

Experimental Setup

- **Target projects**
  - Firefox (30 PRs with known regressions)
  - Zettlr (38 recent PRs), Godot (25 recent PRs), JabRef (18 recent PRs)
- **LLMs**
  - GPT-5-nano for filtering historical issues
  - GPT-5.2 for scenario generation and bug detection
  - Claude Opus 4.5 for scenario execution

Bug Detection: Real-World Findings

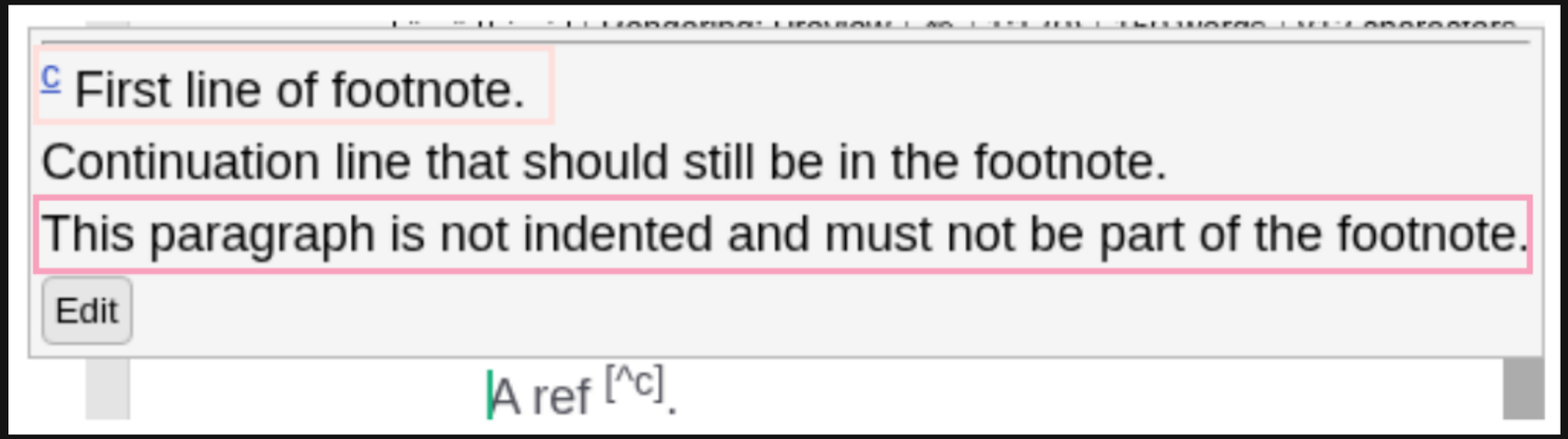
- 26 previously unknown findings reported to developers
- 18 fixed, 3 confirmed, 5 intended behavior

Bug Detection: Real-World Findings

- 26 **previously unknown findings** reported to developers
- 18 fixed, 3 confirmed, 5 intended behavior

Example from Zettlr publication workbench:

Pre-PR:

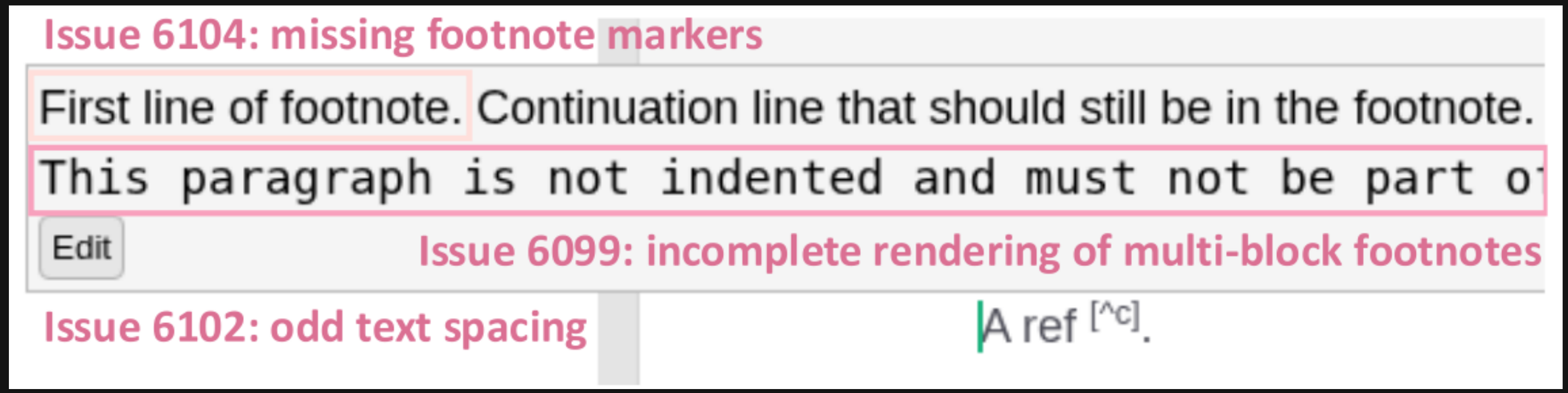


Bug Detection: Real-World Findings

- 26 **previously unknown findings** reported to developers
- 18 fixed, 3 confirmed, 5 intended behavior

Example from Zettlr publication workbench:

Post-PR:



Bug Detection: Precision & Recall

SUT	PR #	Bug #	TP #	FP #	Prec.
Firefox	30	29	52	41	0.559
Zettlr	38	63	67	95	0.414
JabRef	18	15	17	23	0.425
Godot	25	10	10	14	0.417
Total	111	117	146	173	0.458

Bug Detection: Precision & Recall

SUT	PR #	Bug #	TP #	FP #	Prec.
Firefox	30	29	52	41	0.559
Zettlr	38	63	67	95	0.414
JabRef	18	15	17	23	0.425
Godot	25	10	10	14	0.417
Total	111	117	146	173	0.458

On average, a PR introduces more than one new bug!

Bug Detection: Precision & Recall

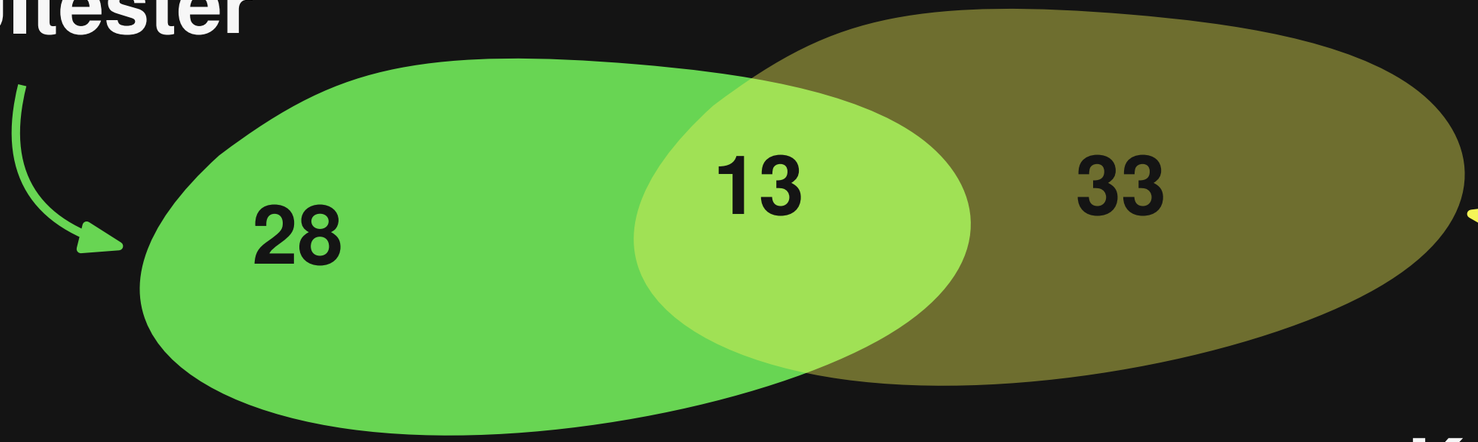
SUT	PR #	Bug #	TP #	FP #	Prec.
Firefox	30	29	52	41	0.559
Zettlr	38	63	67	95	0.414
JabRef	18	15	17	23	0.425
Godot	25	10	10	14	0.417
Total	111	117	146	173	0.458

Common reasons for false positives:

- GUI rendering and temporal instability (45%)
- LLM reasoning errors (28%)

Bug Detection: Precision & Recall

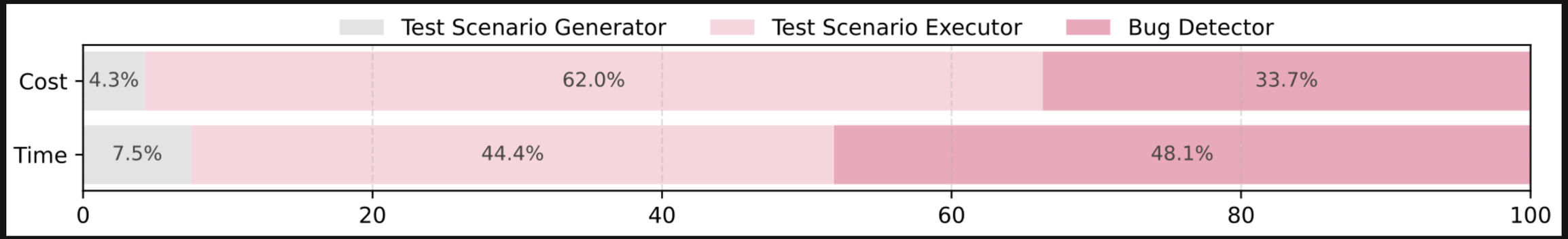
Bugs detected by RippleGUItester



Known bugs in Firefox

Costs and Efficiency

- Scenario knowledge base (**one-time effort**):
  3.22–50.3 hours, USD 0.50–20.82
- **Per-PR costs**:
  54.8 minutes, USD 6.00



Conclusions

- **Code changes have ripple effects**
- **RippleGUItester: Change-aware exploratory testing**
  - GUI-level test scenarios
  - Visual oracle enabled by multi-modal LLMs
- **Envisioned usage:**
  Check changes before or shortly after merging