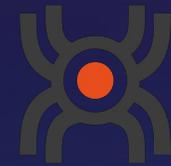




German
OWASP
Day 2025

YuraScanner



Leveraging LLMs for Task-driven Web App Scanning

Aleksei Stafeev, Tim Recktenwald,
Gianluca De Stefano, Soheil Khodayari,
Giancarlo Pellegrino

Coverage

More coverage

=

More opportunities to find vulnerabilities

Web Security Scanners

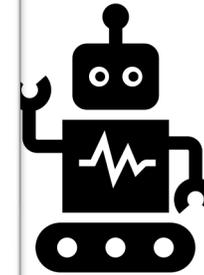
Main objective
Automatically find
target website

Tools:

- **Click** some buttons
- **Type** some letters
- **Beep** occasionally

The screenshot shows the eBay homepage with several key elements highlighted for a web security scanner:

- Navigation:** Links for home, my eBay, site map, and sign in are at the top right. A menu bar contains Browse, Sell, Services, Search, Help, and Community.
- Search:** A search bar with the text "what are you looking for?" and a "find it!" button. A "Smart Search" link is also present.
- Welcome Circle:** A blue oval containing "welcome new users" and buttons for "register", "new to eBay?", "how do I bid?", "how do I sell?", and "why eBay is safe". A red box highlights the "Bottom of welcome circle".
- Specialty Sites:** A list of links including Automotive, Business Exchange, Great Collections, and Half.com.
- Categories:** A vertical list of product categories such as Antiques & Art, Books, Movies, Music, Coins, Stamps, Collectibles, Computers, Dolls, Figures, Jewelry, Gemstones, Photo & Electronics, Pottery & Glass, Real Estate, Sports, Toys, Bean Bag Plush, and Everything Else.
- Hot Picks:** A red-bordered section featuring "CLASSIC" items like The Beatles, Breweriana, First Edition Books, Lincoln Logs, Half.com, and More Classic Items.
- Local Trading:** A blue-bordered section with links for Appliances, Furniture, Sporting Goods, and a "Pick a region" dropdown menu.
- Browse by Themes:** A green-bordered section with links for Gifts - Men, Gifts - Women, Gifts - Children, and an "All Themes" dropdown menu.
- Featured Items:** A green-bordered section listing items like "Miracle Diet Pills - Get Fit & Trim For Xmas!", "#1 BASS Guitar Study Program! Vol 1-4", and "#1 BLUES Guitar Study Program! Vol 1-7".
- Don't Miss...:** A yellow-bordered section with a "Don't Miss..." header and a small image of a woman.



Research on Crawling

• Page Similarity

- URL
- DOM Tree
- HTML (e.g., Hash-based)
- Screenshots
- Custom
- Randomized
- Reinforcement Learning
- Code coverage guided
- Rule-based

Randomized won!

Name	Algorithm	Tools
Page Similarity: Page URL		
URL Equality	True if the URL strings are the same	Black Widow, JAW, SecuBat, GNU Wget, w3af
Page Similarity: DOM Tree		
Tree Equality	True if the two trees are identical	RAPE
RTED	True if $RTED(t_i, t_j) > c$ for a threshold c , where RTED calculates the minimum of node edit operations that transform one tree into the other one	Crawljax, FeedEx
UI Controls	True if the ratio of common UI controls (e.g., input tags) is greater than a threshold c	AutoBlackTest
Root-Link Paths	True if the ratio of common root-to-link paths is greater than a threshold c	Crescenzi
Page Similarity: HTML Code		
SimiHash	True if the Hamming distance of two 64-bit fingerprint digests is greater than a threshold c	Crawljax, Manku
TLSH	True if the distance of two locality-sensitive hash digests is greater than a threshold c	Crawljax
Common Shingles	True if the fraction of common shingles is greater than a threshold c	Broder
TAF	True if $TAF(t_i, t_j) > c$, where TAF is the difference of the tag and attribute frequency function of two trees	Lucca
LevenSeq	True if $LevenSeq(s_i, s_j) > c$, where LevenSeq is the Levenstein distance between the sequences of the tags and attributes	Lucca
	True if the two color histograms is greater than a threshold c	Crawljax
	True if the two bit hash digests is greater than a threshold c	Crawljax
	True if the two hash digests is greater than a threshold c	Crawljax
	True if the two hash digests is greater than a threshold c	Crawljax
	True if the two hash digests is greater than a threshold c	Crawljax
	True if the two hash digests is greater than a threshold c	Crawljax
Dagger	True if the sequence of hashes per DOM level	
LoRE	True if the form prefix trees are the same	
	True if the rooted link prefix trees are the same (precondition)	
	True if the fingerprints match and SimiHash is greater than a threshold	
	True if the sets are the same	
	True if the request methods are the same	ZAP
	True if the request methods are the same	Wapiti
	True if the request methods are the same	Skipfish
		Crawl, LigRE

Look in paper for more!

SoK: State of the Krawlers – Evaluating the Effectiveness of Crawling Algorithms for Web Security Measurements

Aleksei Stafeev
CISPA Helmholtz Center
for Information Security

Giancarlo Pellegrino
CISPA Helmholtz Center
for Information Security

Abstract

Web crawlers are tools widely used in measurements whose performance has been studied so far.

Motivation

Search for a connection Show map 



 **Today, from 11:38** >  **Return journey** >  **Passenger, bicycles, BahnCards** >

Change outbound route Add 1 Person (aged 27-64), no discount

Stopovers > **Mode of transport** > **Transfer time** > **Book seat only** ?

None All Normal

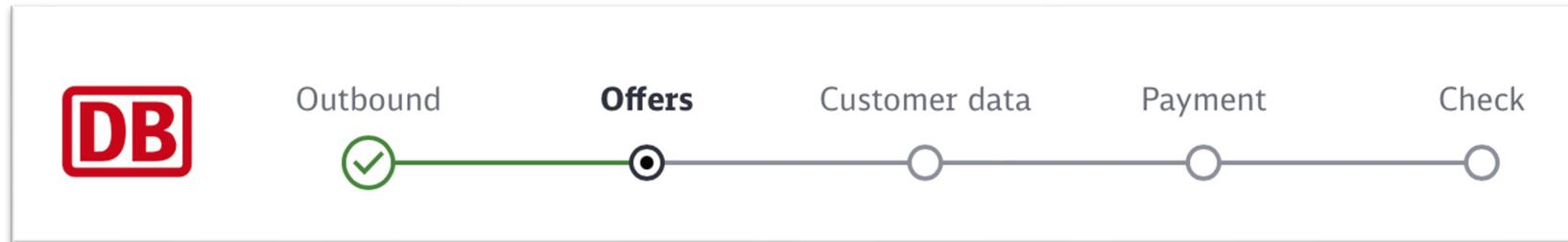
Show fastest connections Direct services only Bicycle transport possible

 What does that mean?

Deutschland-Ticket connections only

[→ Information for holders of the Deutschland-Ticket](#)

Motivation



Multi-step workflow

Motivation

- Traditional scanners struggle with exploring deeper states
- **Key limitation:** They lack awareness of multi-step workflows
- ML-based methods have been proposed to tackle this weakness
 - E.g., reinforcement learning on user-provided traces [1]
- Does not scale well!

Approach

- Instead of training a model, we opted to use [large language models \(LLMs\)](#)
- Non-academic approaches have proposed LLM-based [browsing agents](#) to [assist users](#) with tasks [2, 3]
 - E.g., “Book a hotel in Düsseldorf”
- Instead, we want to complete workflows and reach deeper states in web applications [without user interaction](#)

[2] N. Friedman. (2022) Natbot. <https://github.com/nat/natbot>.

[3] (2024) Skyvern. <https://github.com/Skyvern-AI/skyvern>.

You are a scanner...

Yu r a scanner...

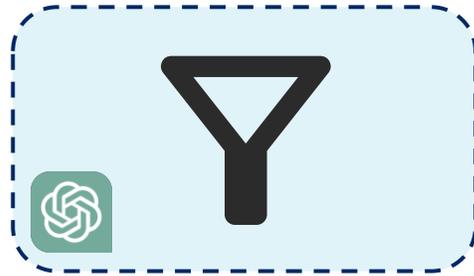
YuraScanner

YuraScanner

A **fully automated, task-driven** web application scanner

Architecture of YuraScanner

Task Extraction



Task Execution



Vulnerability Scanning



1. Add a new category for products.
2. Edit the information for an existing product.
3. Delete a previous order.

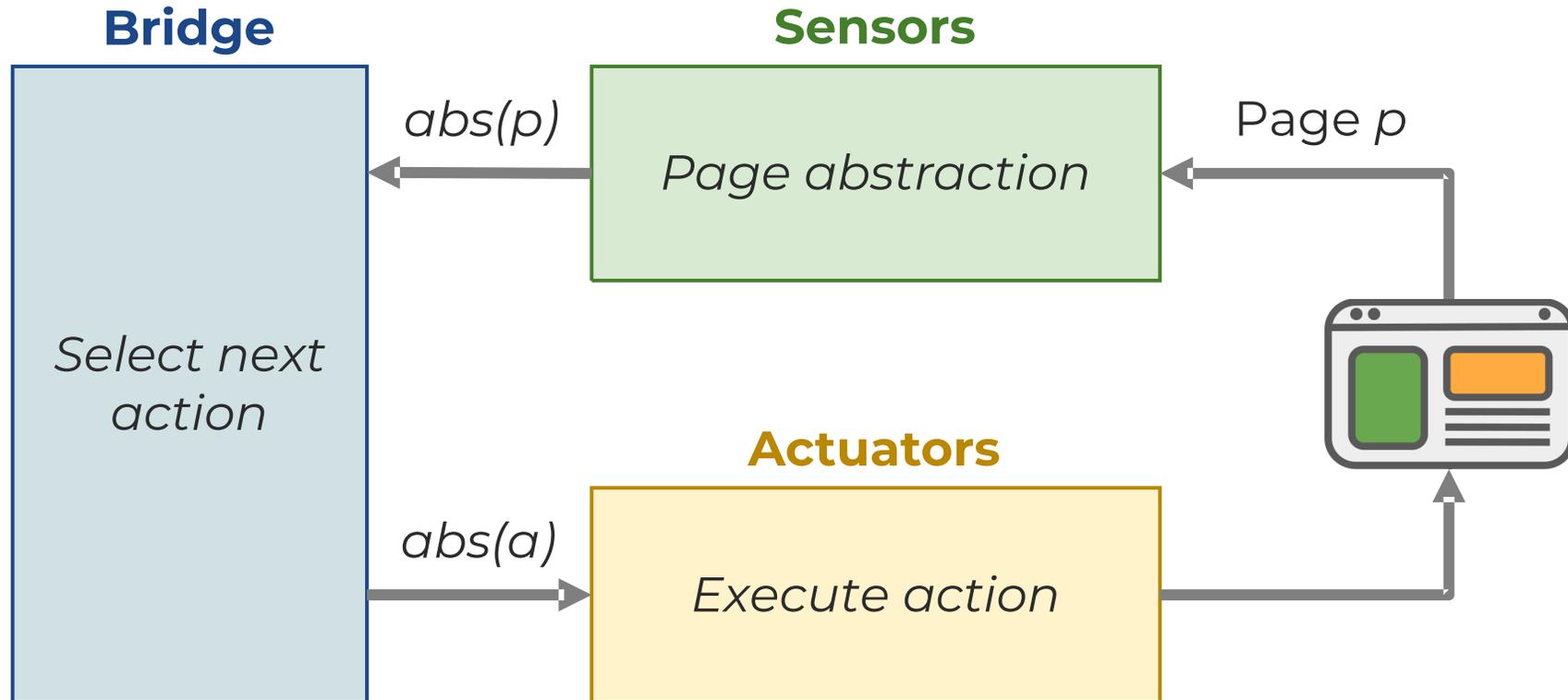
1. Do a shallow crawl
2. Extract all **interactable HTML elements**
3. Provide a list of elements to a **LLM** and request a list of tasks

Architecture of YuraScanner

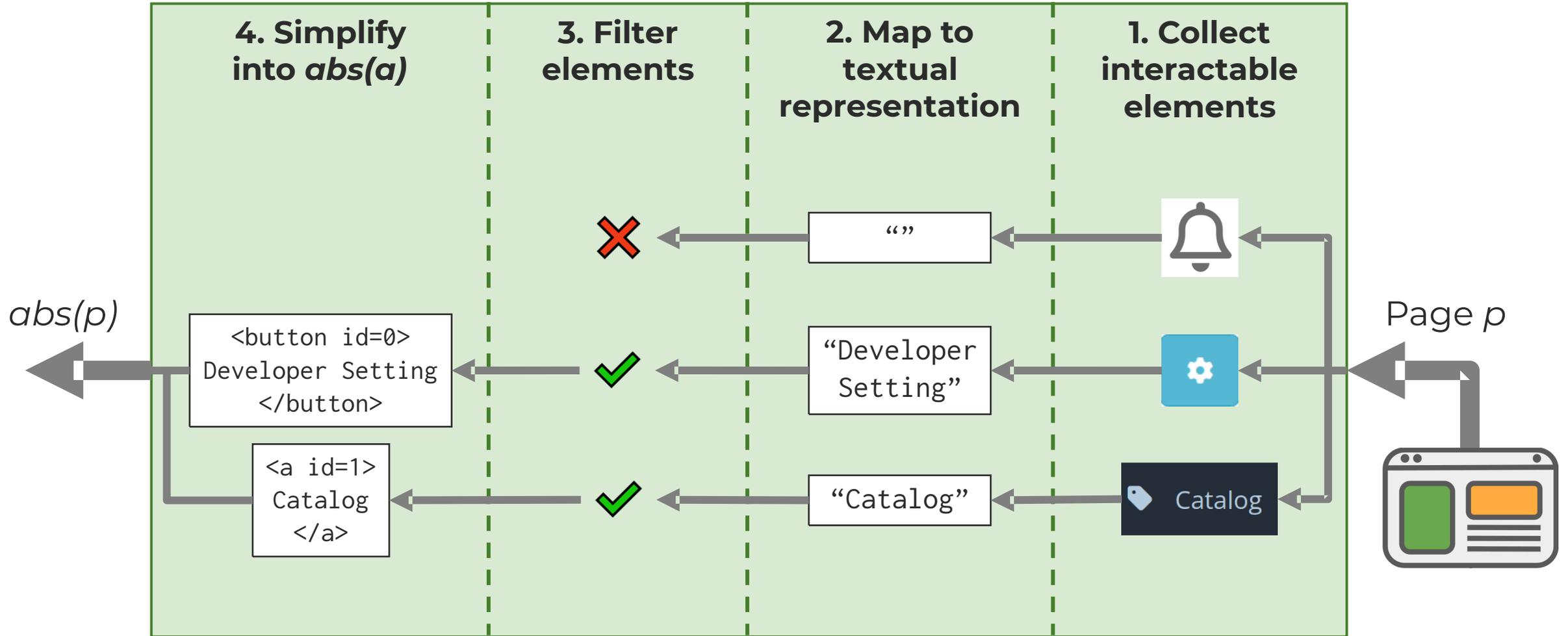


- Executes every task iteratively
- At each step of a task:
 - Generates a *simplified textual page representation*
 - Queries **LLM** for the next **command** (e.g., *click button #2*)
 - Executes the **command**

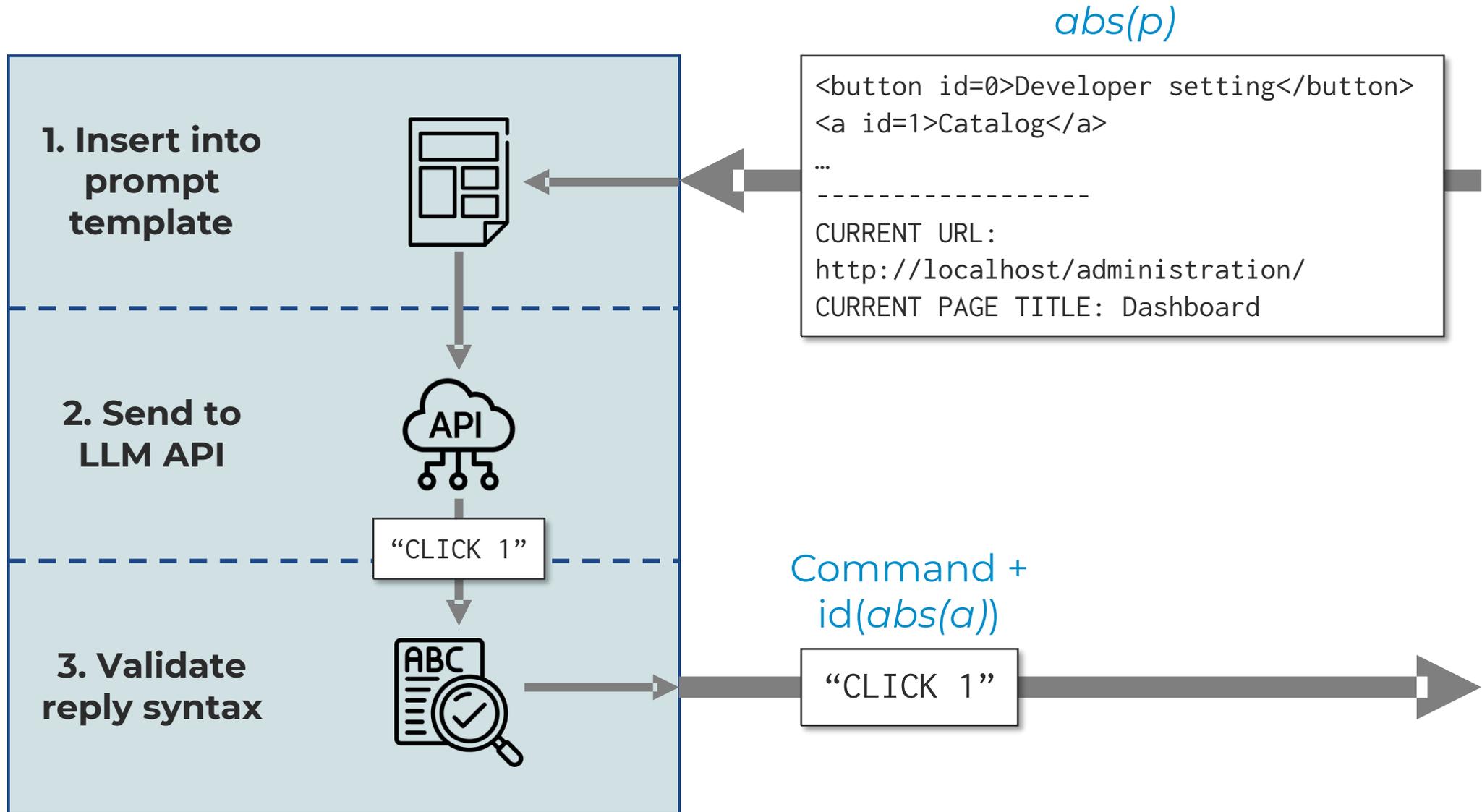
Task-driven Crawling



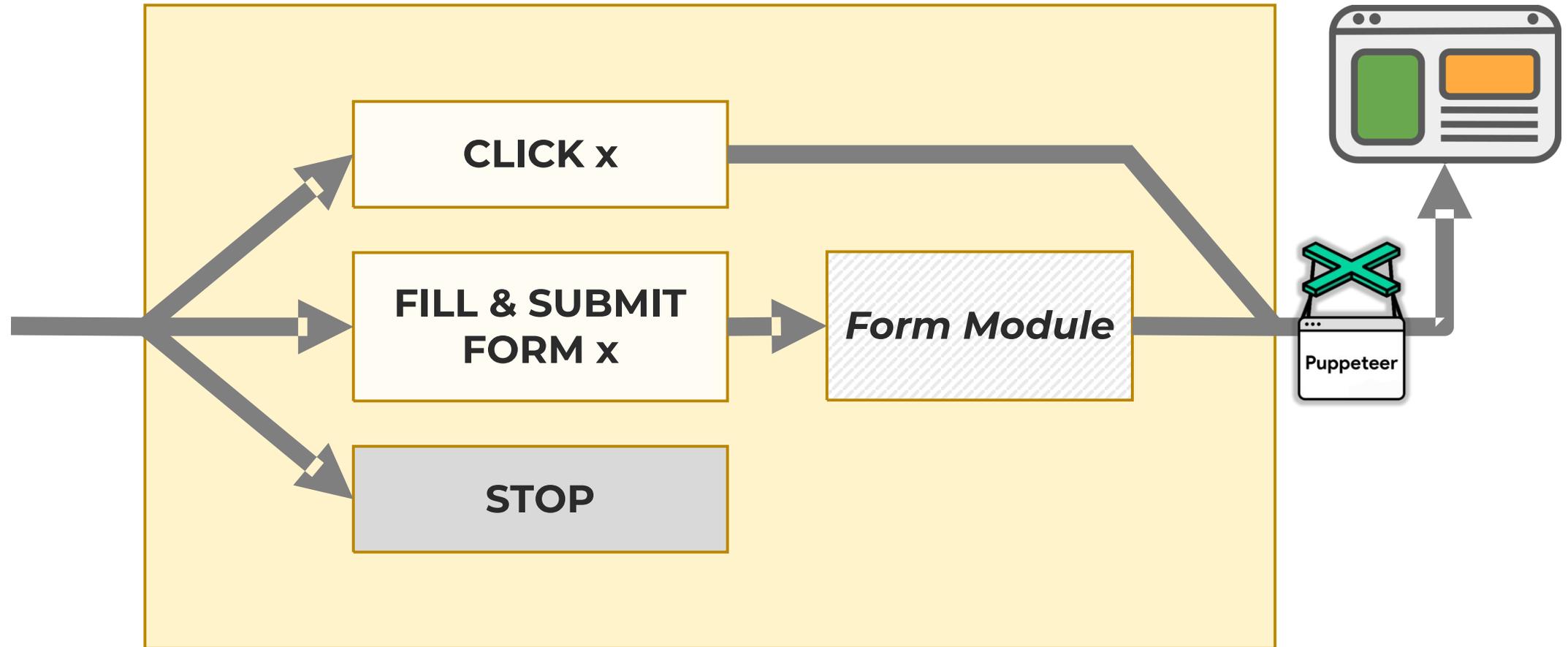
Sensors



Bridge



Actuators



Architecture of YuraScanner

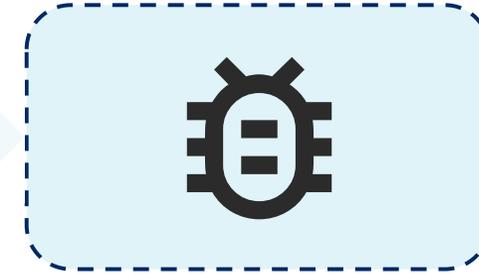
Task Extraction



Task Execution



Vulnerability Scanning



1. Visits **every form** collected during **task execution**
2. Uses the **XSS detection** engine of **Black Widow** [4]

[4] Benjamin Eriksson, Giancarlo Pellegrino, and Andrei Sabelfeld: "BlackWidow: Blackbox data-driven web scanning," in *2021 IEEE Symposium on Security and Privacy (SP)*

Evaluation

- We evaluated **YuraScanner** on 20 popular, modern web applications
- We divided our testbed into two sets:

1. Task Extraction and Execution

- Random subset of 10 web apps
- **Manually label** valid tasks and their success rate during task execution

2. Vulnerability Detection

- All 20 web apps
- Inspect vulnerabilities found by the attack module



Evaluation Results: Task Extraction



2,361 tasks

- 2,361 tasks were generated in total across 10 web applications

Evaluation Results: Task Extraction



- 2,361 tasks were generated in total across 10 web applications
- 77% of the tasks were valid (1,818 tasks)

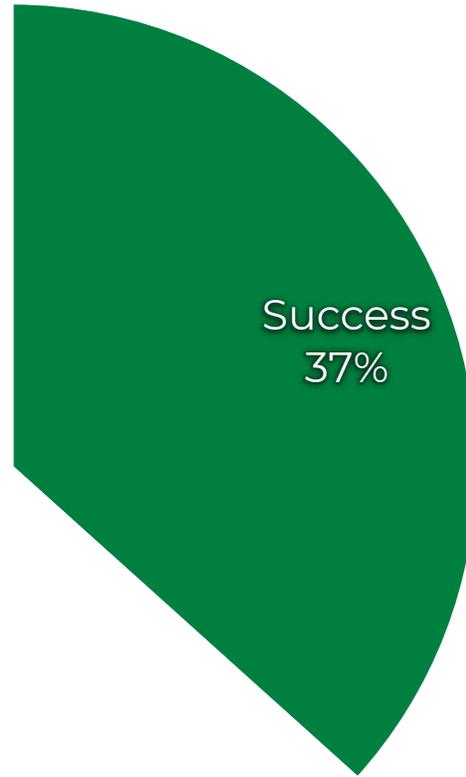
Evaluation Results: Task Extraction



- 2,361 tasks were generated in total across 10 web applications
- 77% of the tasks were valid (1,818 tasks)
- “Invalid” = Functionality does not exist in the web application
- Invalid task generation mainly occurred on pages with insufficient context
 - E.g., login page with only one button

MediaWiki *Get detailed information about membership or subscription plans*

Evaluation Results: Task Execution

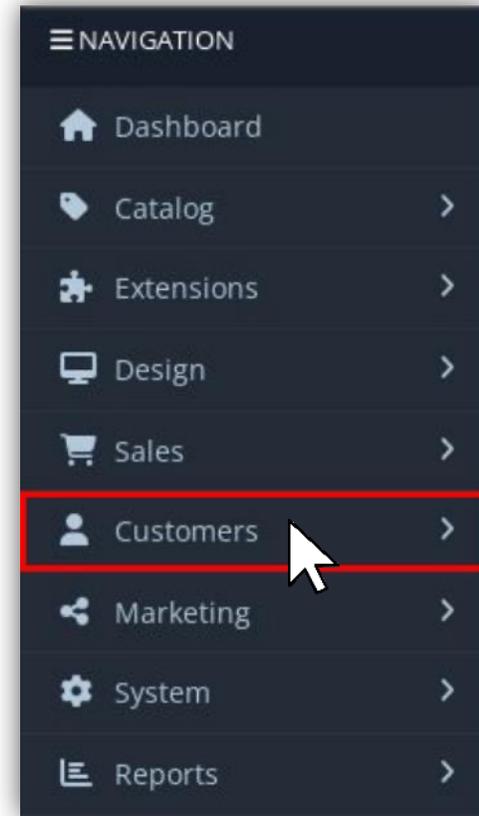


Task Execution Classification
(1,818 valid tasks)

Evaluation Results: Task Execution



Task Execution Classification
(1,818 valid tasks)

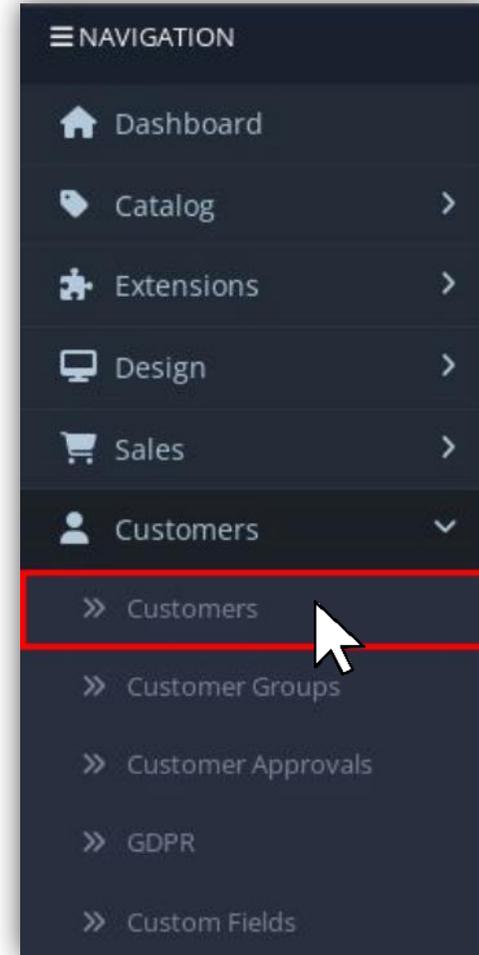


Task: Add new “Customers” to the database

Evaluation Results: Task Execution



Task Execution Classification
(1,818 valid tasks)

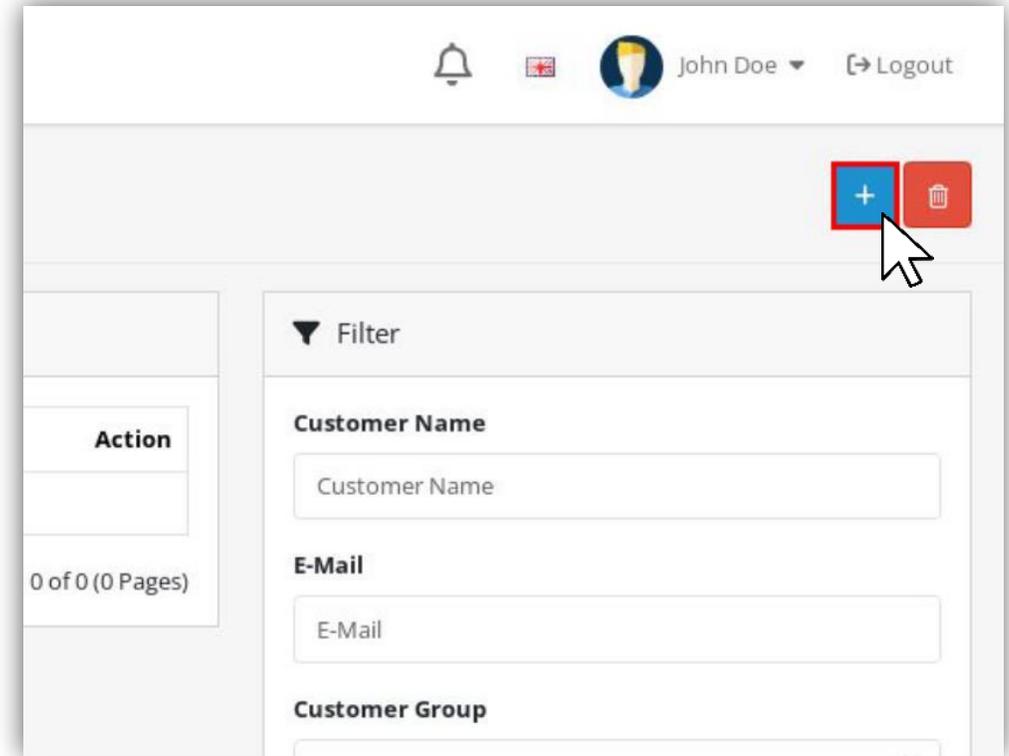


Task: Add new “Customers” to the database

Evaluation Results: Task Execution



Task Execution Classification
(1,818 valid tasks)



Task: Add new "Customers" to the database

Evaluation Results: Task Execution



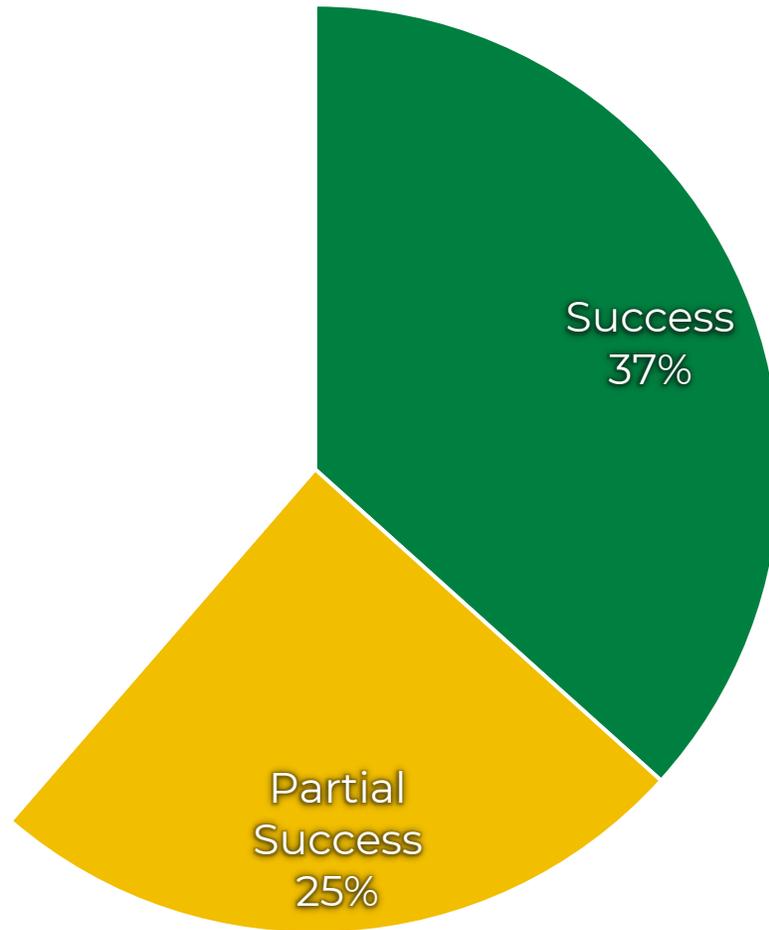
Task Execution Classification
(1,818 valid tasks)

A screenshot of a web form for adding a new customer. The form is divided into two sections: 'Customer Details' and 'Password'. The 'Customer Details' section includes fields for Store (Default), Customer Group (Default), First Name (Alice), Last Name (Smith), E-Mail (alice.smith@example.com), and Telephone (1234567890). The 'Password' section includes fields for Password and Confirm, both masked with dots.

Field	Value
Store	Default
Customer Group	Default
* First Name	Alice
* Last Name	Smith
* E-Mail	alice.smith@example.com
Telephone	1234567890
* Password
* Confirm

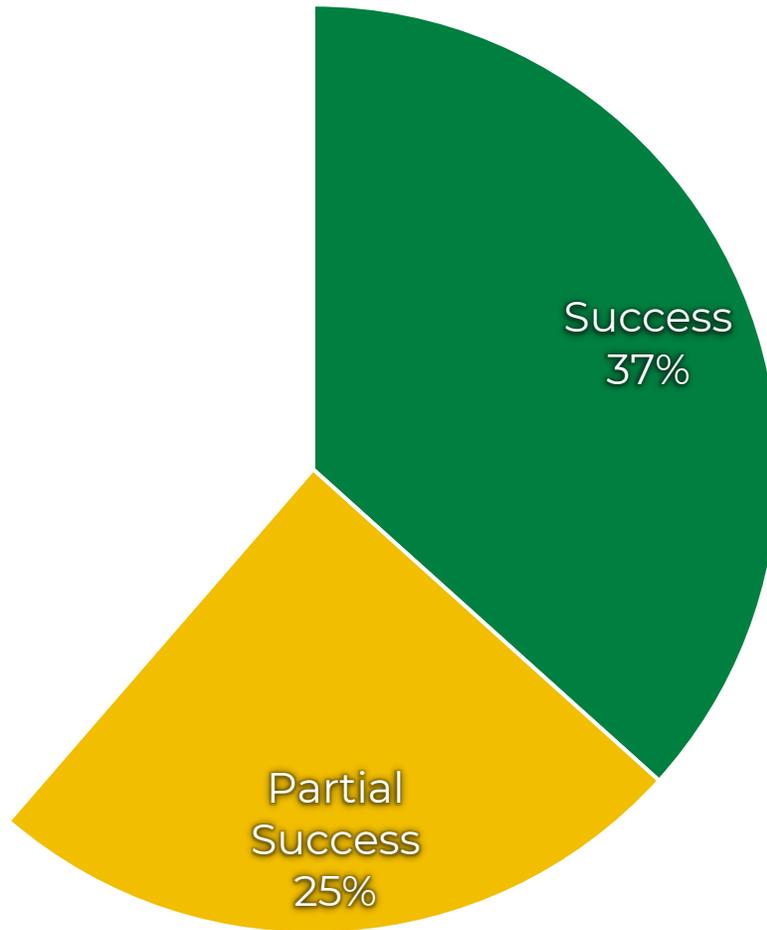
Task: Add new “Customers” to the database

Evaluation Results: Task Execution



Task Execution Classification
(1,818 valid tasks)

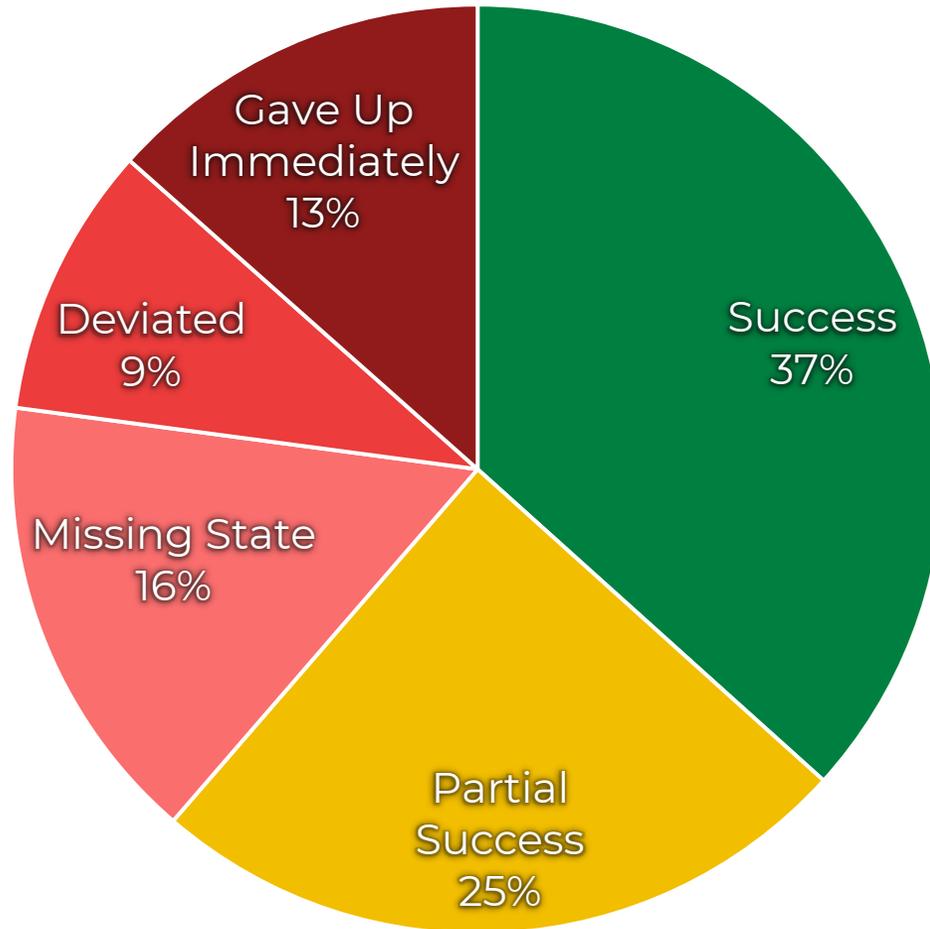
Evaluation Results: Task Execution



Task Execution Classification
(1,818 valid tasks)

The image shows a warning message box at the top with a red background and a white border. The message reads: "Warning: Please check the form carefully for errors!" followed by a close button (X). A large blue arrow points downwards from this warning box to a screenshot of a web form titled "Add Category". The form has tabs for "General", "Data", "SEO", and "Design". The "General" tab is active, showing fields for "Category Name" (filled with "Electronics"), "Description" (with a rich text editor), "Meta Tag Title" (filled with "Electronic Devices"), "Meta Tag Description" (filled with "Meta description for electronic devices including top brand names."), and "Meta Tag Keywords" (filled with "electronics, electronic devices, gadgets"). A small warning icon is visible in the top right corner of the form, which is highlighted by a dashed blue box.

Evaluation Results: Task Execution



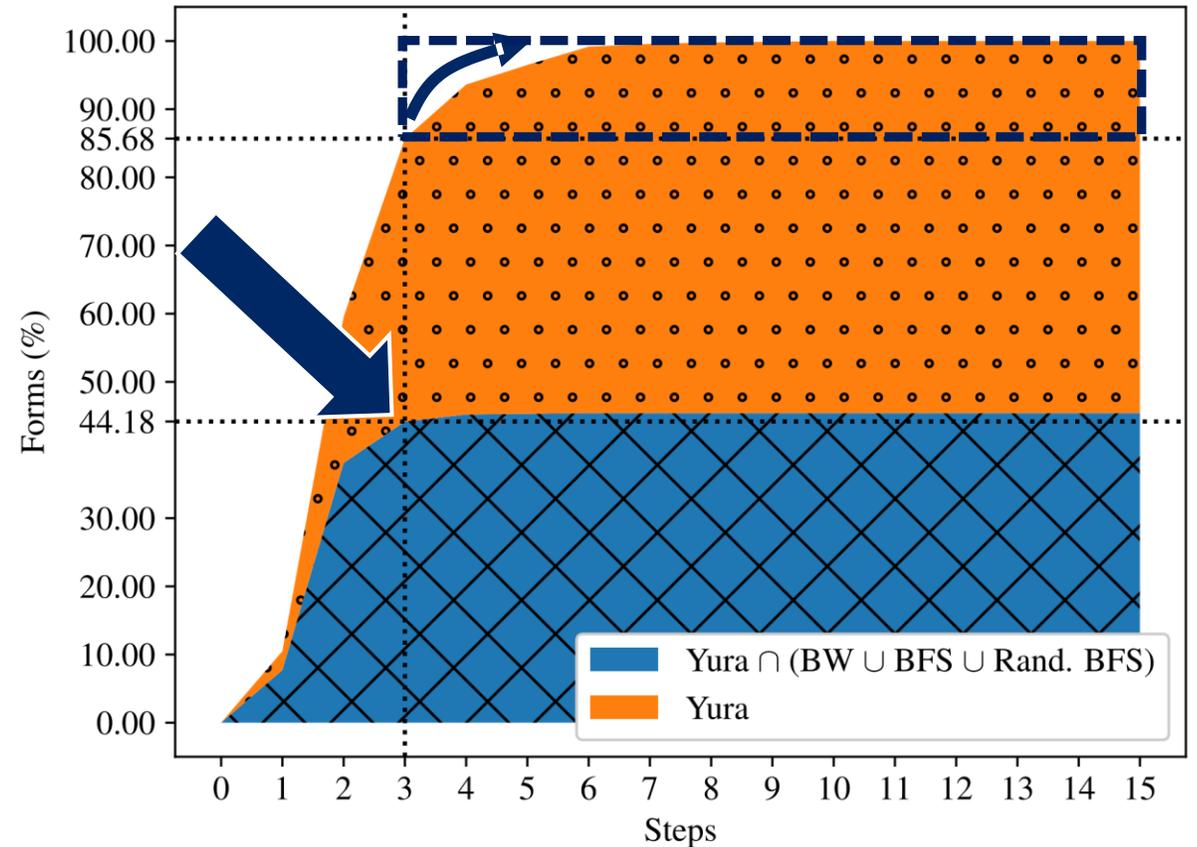
Task Execution Classification
(1,818 valid tasks)

- Missing State
 - E.g., tried to edit an object **before** creating one
- Deviated
 - Clicked on a **wrong** button
- Gave Up Immediately
 - Issued a **STOP** command

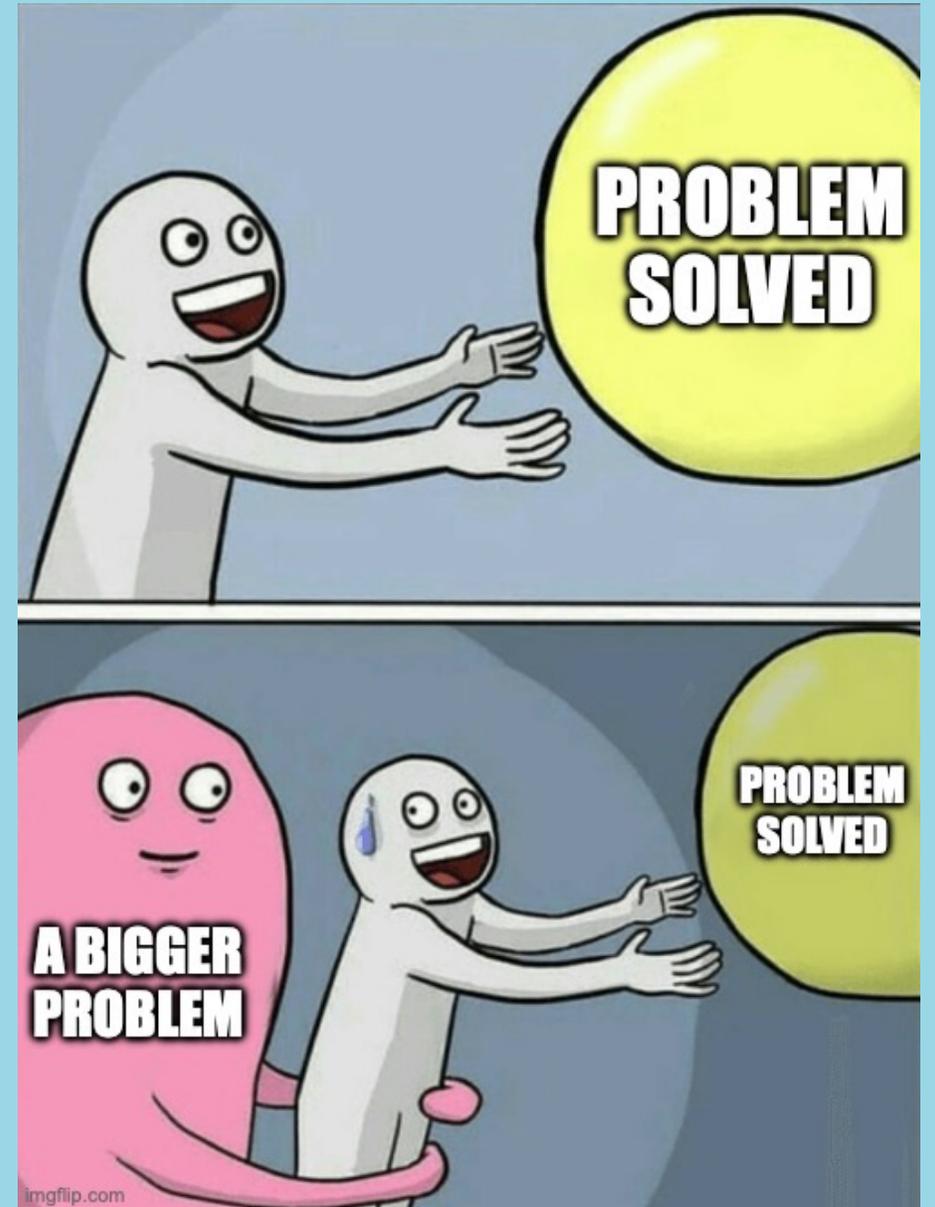
Characterization of the New Attack Surface

Is the *new attack surface* found by YuraScanner “*deeper*”?

- Comparison with *BlackWidow*, *BFS*, *Random BFS*
- 44% percent of the forms were discovered *by all scanners*
- The *remaining forms* were found **exclusively** by YuraScanner
 - Notably, 14.3% were found at **depth > 3**
 - **Out of reach** for traditional tools!



Suffering from Success

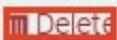


Suffering from Success

Task: “Delete a user from the ‘User Management’ section.”

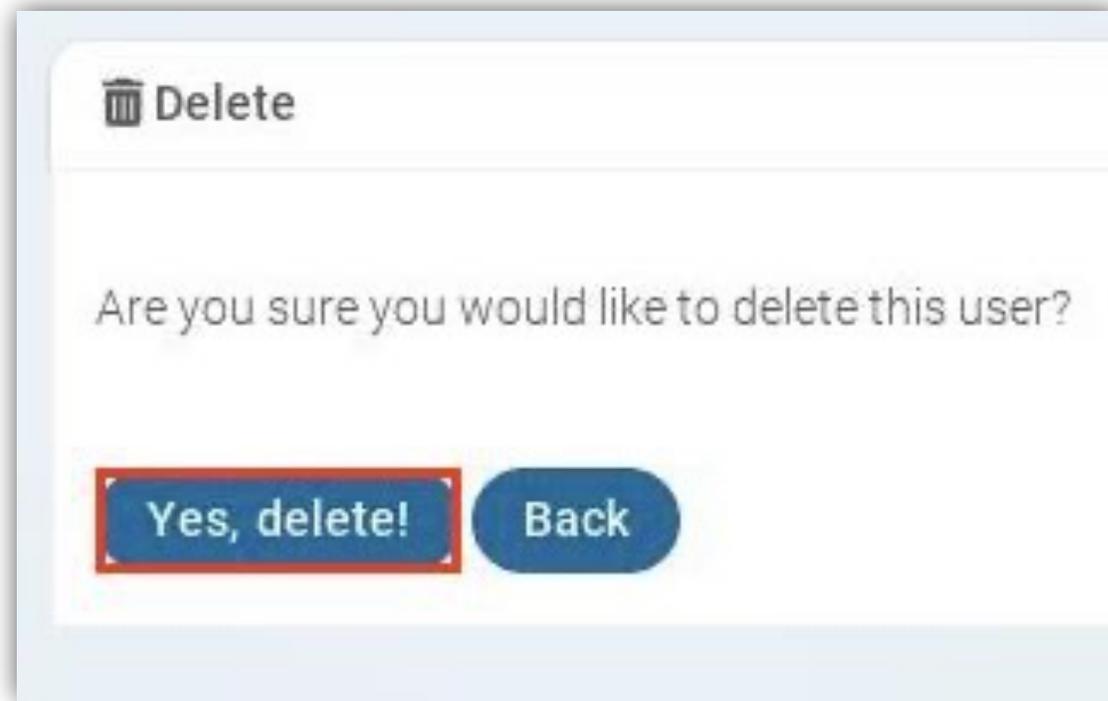


The screenshot displays a user management table with the following structure:

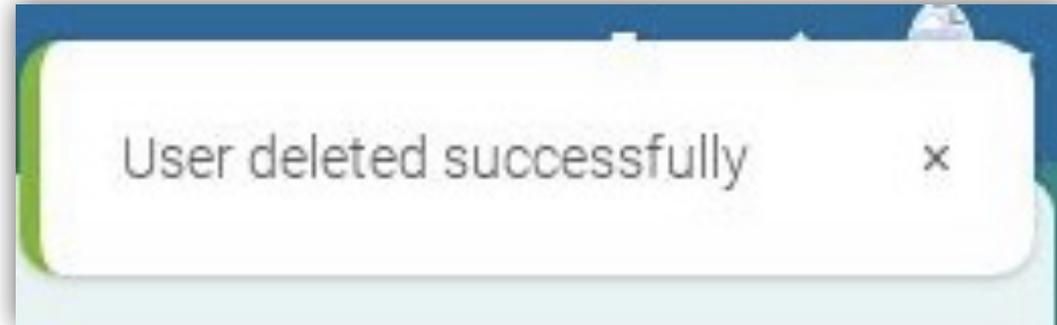
Name ▲	Email ▼	Two-factor Authentication ▼
John Doe	jaekpot@local	

Below the table, there is a pagination control showing "Showing 1 to 1 of 1 entries" and a "Show 100 ▼ entries" dropdown menu. A "Previous" link is also visible on the right side of the pagination area.

Suffering from Success



Suffering from Success



Suffering from Success

Login

Username or password incorrect!

Email

Enter email address

Password

Enter password

[Forgot password?](#)

Login

A login form with a white background and rounded corners. At the top left, the word "Login" is written in a blue, sans-serif font. Below it is a large, rounded rectangular input field. Inside this field, a smaller, rounded rectangular box with a red border contains the text "Username or password incorrect!". Below the error message are two more input fields. The first is labeled "Email" and contains the placeholder text "Enter email address". The second is labeled "Password" and contains the placeholder text "Enter password". To the right of the password field, there is a link that says "Forgot password?". At the bottom of the form is a solid blue button with the word "Login" written in white, centered on it.

Directed by
YURASCANNER

Vulnerability Detection



App	Total	Unique	YuraScanner		Black Widow	
			Stored XSS	Reflected XSS	Stored XSS	Reflected XSS
<i>Redacted</i>	12	11	4	7	-	1
Moodle	2	1	1	-	1	-
Leantime	1	1	-	-	1	-

- 13 unique [zero-day](#) vulnerabilities discovered
- 12 of them found by **YuraScanner**
- Located between two to four clicks away from the front page

Takeaways

- Coverage is vital for the vulnerability detection
- Traditional scanners struggle with **multi-step workflows**
 - 👍 LLMs is a promising solution with **62%** workflows executed fully or partially
- Task-driven crawling reaches **deeper functionalities** compared to traditional scanners
- LLMs may put the web application in an unrecoverable state
 - ⚠️ Need more **safeguarding**

