

Catching UX Flaws in Code

Leveraging LLMs to Identify Usability Flaws at the Development Stage



Nolan Platt

*Department of Computer Science
Virginia Tech*



Ethan Luchs

*Department of Computer Science
Virginia Tech*



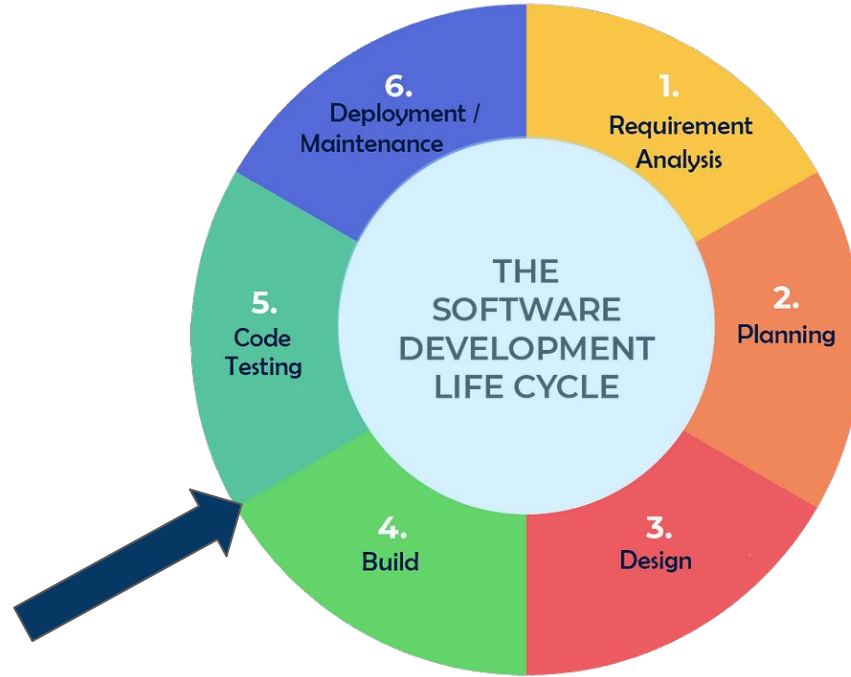
Dr. Sehrish Basir Nizamani

*Department of Computer Science
Virginia Tech*

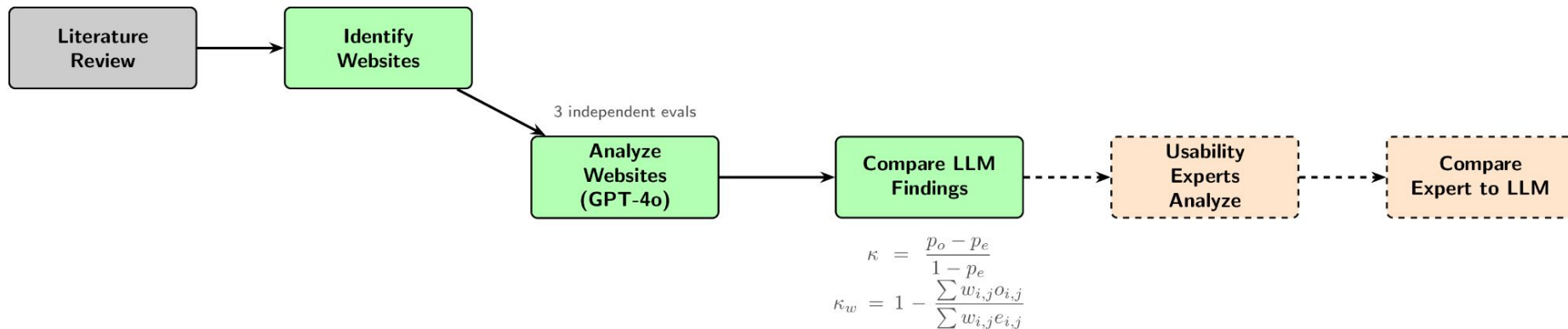
Gaps in Usability Testing

- Usability testing → expose UX and design flaws
 - AB Testing
 - **Heuristic Evaluations**
- Difficult to implement in smaller teams or fast environments (startups, research)

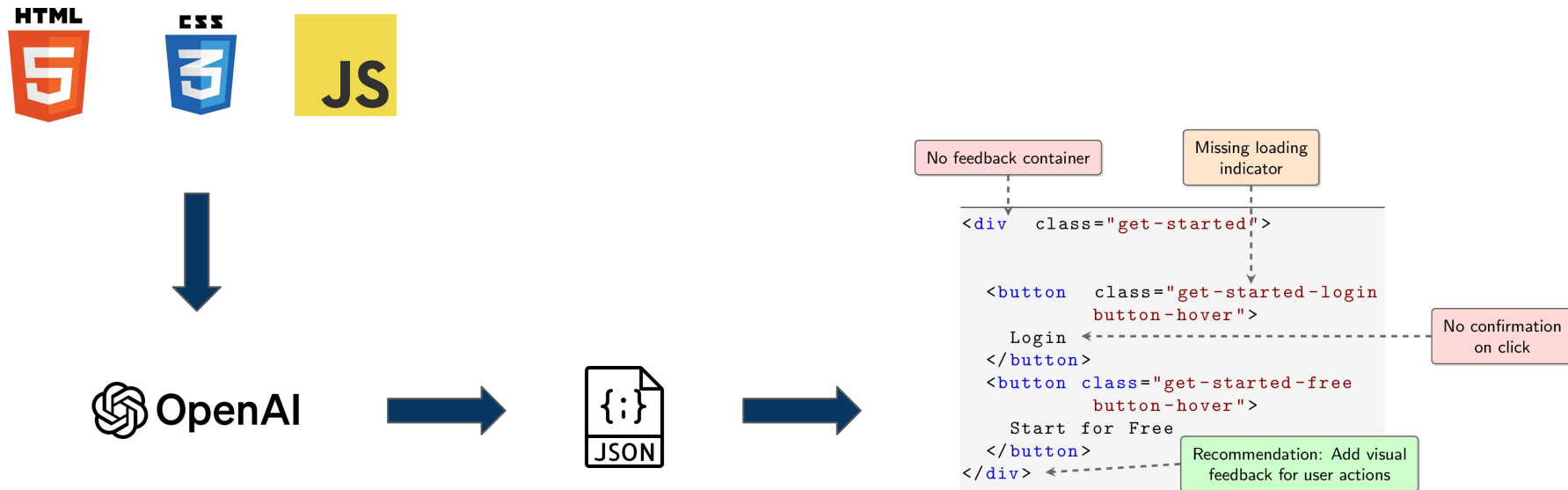
RQ: Can large language models (LLMs) provide reliable and consistent heuristic usability assessments in web applications at the development stage by analyzing source code?



Methodology



LLM Pipeline



Issue Detection Consistency

- **86%** of evaluations identified usability issues
- **.47-.53 Cohen's Kappa** (moderate agreement)
- **82-85% exact agreement** across evaluation pairs

UNWEIGHTED COHEN'S KAPPA ACROSS ALL ISSUE PAIRS

Comparison	Cohen's κ	Agreement	Exact Percent
eval1-eval2	0.471	247/300	82.3
eval1-eval3	0.530	255/300	85.0
eval2-eval3	0.502	254/300	84.7

∴ GPT-4o shows moderate consistency in detecting issues

Severity Rating Variability

DISTRIBUTION OF SEVERITY RATINGS

- **.60-.67 Weighted Cohen's Kappa (moderate agreement)**
- **56-57% exact agreement in severity ratings**

Severity Level	Count
0 (No Issue)	164
1 (Cosmetic)	251
2 (Minor)	306
3 (Major)	155
4 (Catastrophic)	9

∴ Severity judgments show reasonable agreement, more variability

What Does This Mean for Automated Usability Testing?

- **LLMs can perform rudimentary early-stage heuristic evaluations**
 - **Internal consistency $\neq$ correctness**
 - **Requires expert validation**
- **Automated screening tool to flag for human review**

Future Work

- **Compare against expert evaluations**
- **Train local LLM w/ usability criteria and reasoning examples**
 - **Synthetic Q+A training data**
 - **Diverse evaluation and code samples**
- **Agentic UX analysis, implemented directly into software development lifecycle**

Questions?

Nolan Platt*

*Department of Computer Science
Virginia Tech*

nolanplatt@vt.edu

Ethan Luchs**

*Department of Computer Science
Virginia Tech*

ethanluchs@vt.edu

Dr. Sehrish Basir Nizamani

*Department of Computer Science
Virginia Tech*

sehrishbasir@vt.edu