

EASI-LLM: Ecosystem-Aware Software Implementation with LLMs

Muslim Chochlov (UL), Michael English (UL), Jim Buckley (UL)

1 Introduction, problem statement, motivation

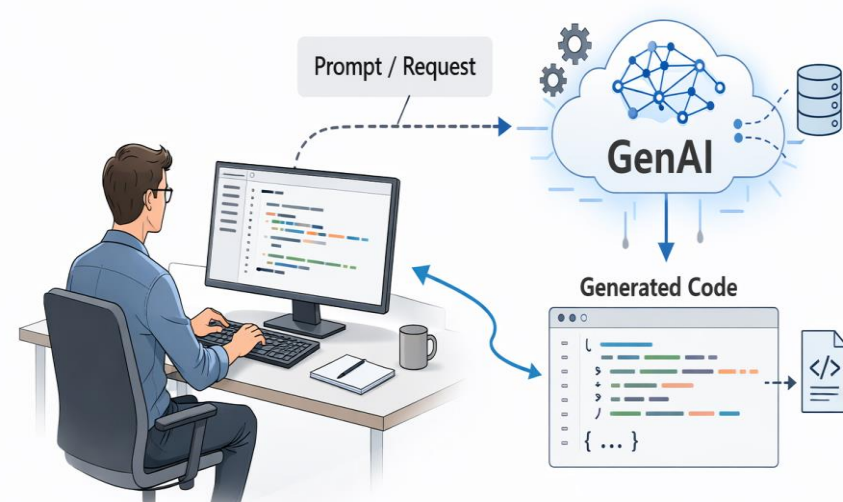
Context:

Programmers seem set to achieve significant coding efficiencies when adopting Generative AI approaches:

- Completing tasks 55.8% faster (Github Copilot)
- 26.08% increase in completed coding tasks (Microsoft, Accenture, and other Fortune 100 companies)
- Rankings in the top 54.3% of code competitions (AlphaCode)

The Problem:

Current GenAI code generation needs to go beyond code snippets and functional requirements towards non-functional requirements.

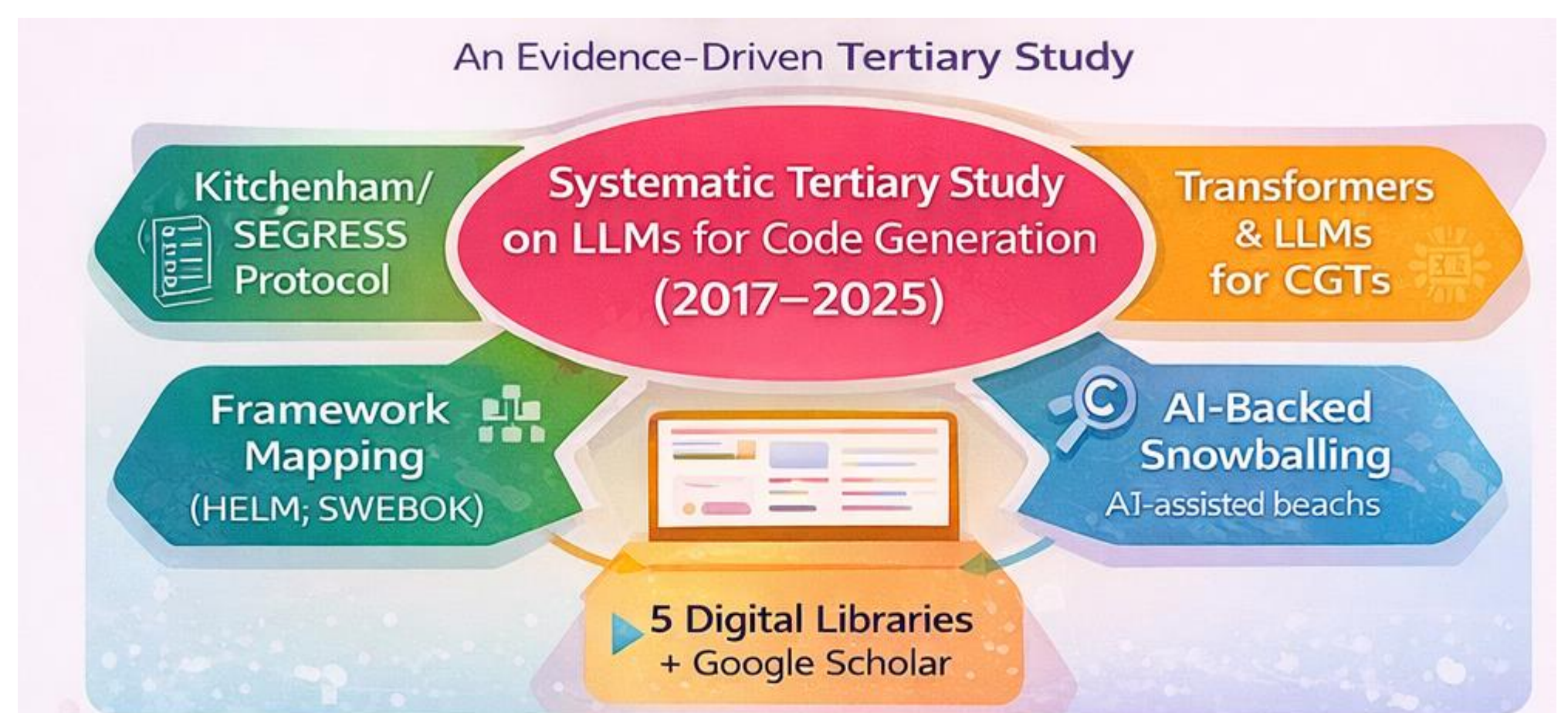


Research Objective:

A (more) holistic, efficiency-ranking protocol for evaluating AI-generated code

2 Literature review, methodology

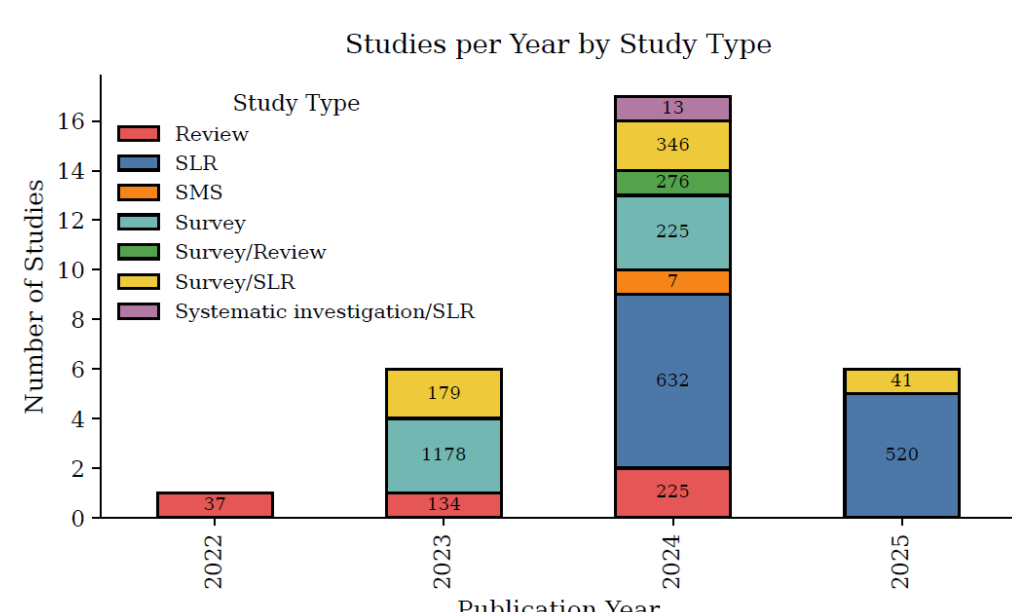
Scarcity of existing secondary/tertiary studies addressing code generation with a focus of real-world applicability. We conducted our own:



Identified 30 relevant secondary studies, that discuss **CODE GENERATING TASKS (CGTs)**

3 Results, Findings

Evolving but seemingly maturing field, with distinguishable concerns (efficiency, accuracy, robustness, toxicity & bias), challenges, and directions



Efficiency Concerns

- Commonly raised



Accuracy

- Strong on benchmarks but weakly supported for real-world generalization



Robustness

- Fragile across tasks and configurations



Toxicity & Bias

- Under-reported



Dominant Challenges

- Economic feasibility
- Evaluation validity
- Socio-technical integration



Future Directions

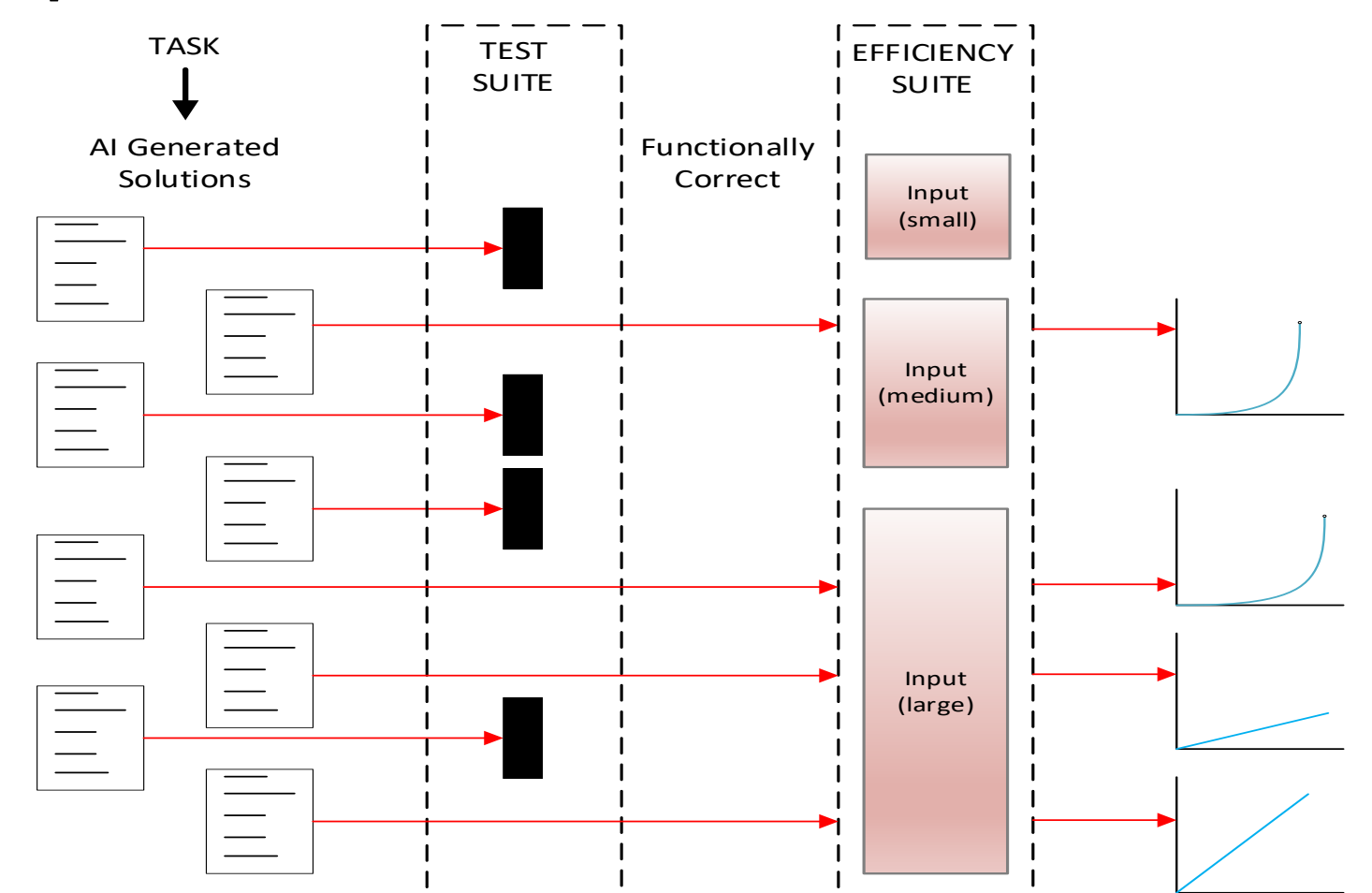
- Domain-aware model improvement
- Holistic, standardized evaluation needed

4 Ongoing Work

Following the identified need for holistic, standardized evaluation, we are working on a novel LLM evaluation benchmark and protocol, focusing on efficiency to support code ecosystems.

Envisaged characteristics include:

- Consisting of 100+ coding tasks
- Cloud-evaluated on 10+ LLMs
- Hardware Independent
- Asymptotic



(Under review at TOSEM) Chochlov Muslim et al.. "A Tertiary Review of Large Language Model-Based Code Generating Tasks: Trends, Challenges, and Future Directions"

HOST INSTITUTION



PARTNER INSTITUTIONS



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



DCU
Ollscoil Chathair Bhaile Átha Cliath
Dublin City University



DUNDALK
DUN DIALGÁN
Dundalk Institute of Technology
Dundalk City University



MTU
Munster Technological University
Munster Technological University



Maynooth University
National University of Ireland Maynooth
Maynooth University



SE TU
South East Technological University
South East Technological University

FUNDED BY

