Towards Neural Synthesis for SMT-Assisted Proof-Oriented Programming in F*

Saikat Chakraborty, Gabriel Ebner, Siddharth Bhat, Sarah Fakhoury,
Sakina Fatima, Shuvendu Lahiri, Nikhil Swamy

arXiv: cs.PL/2405.01787

Taste of F*

Goal

```
val quicksort: #a:eqtype -> f:total_order a -> l:list a ->
  Tot (m:list a{sorted f m /\ is_permutation a l m})
  (decreases (length l))
let rec quicksort #a f l =
```

Solution

```
  match l with
  | [] -> []
  | pivot::tl ->
    let hi, lo = partition (f pivot) tl in
    let m = quicksort f lo @ pivot :: quicksort f hi in
    permutation_app_lemma pivot tl (quicksort f lo) (quicksort f hi);
    m
```

Dataset

Projects

- F*
- Karamel
- EverParse
- HACL*
- miTLS-F*
- EverQuic-Crypto
- Merkle-Tree
- Steel

Soon:

- Pulse
- Zeta
- Starmada
- Noise*
- DICE*

Total: **~940kLOC**

(this is a living, growing dataset)

Checker

- Python API
  - Takes care of all include paths, etc.

- Filter out:
  - Definitions that cannot be checked
  - That can be solved with `let … = ()`

Classification of definitions

- “Simply” typed
  - `int -> int`
  - `(a -> b) -> list a -> list b`

- Proofs
  - `forall xs. xs @ [] == xs`

- Dependently typed

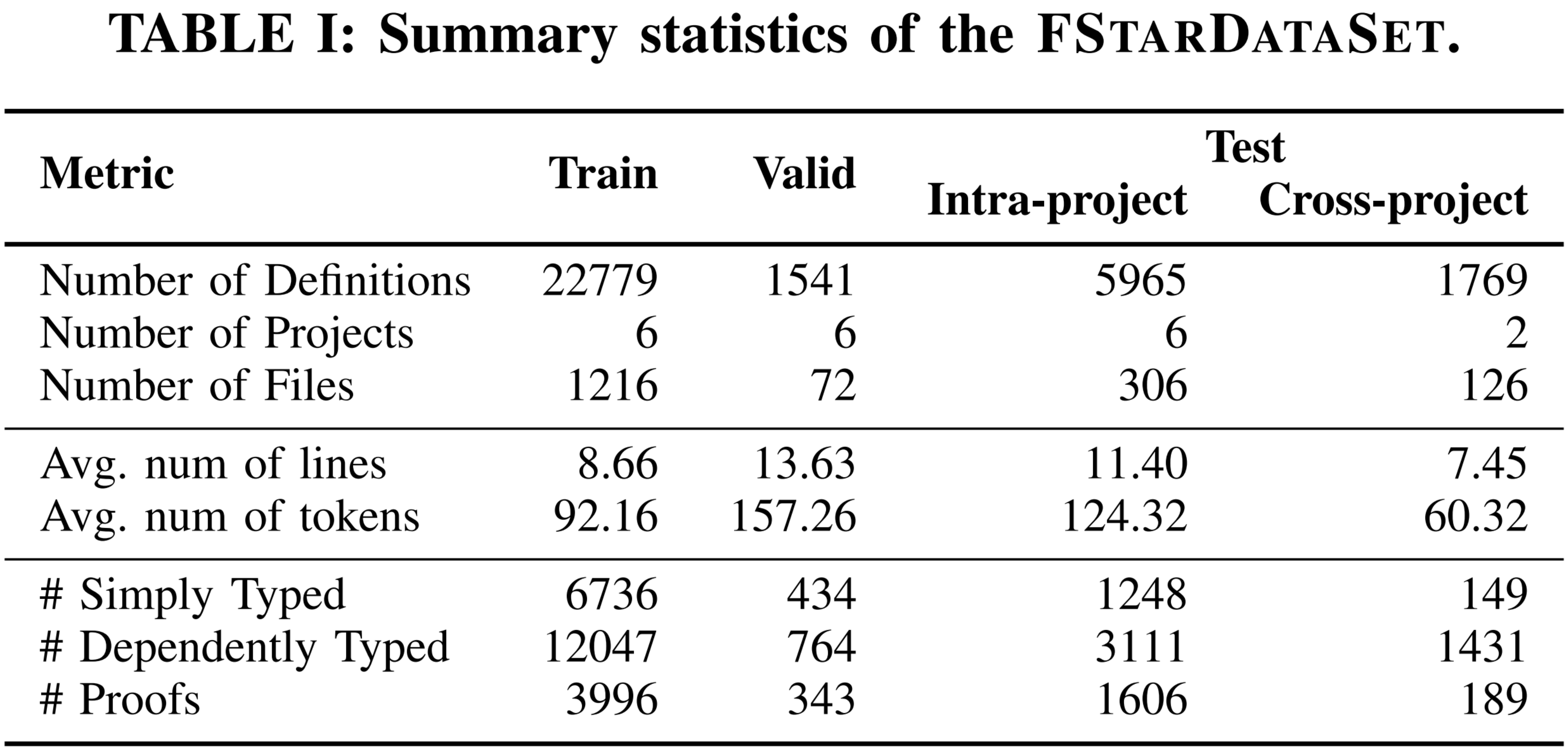


TABLE I: Summary statistics of the FStarDataSet.

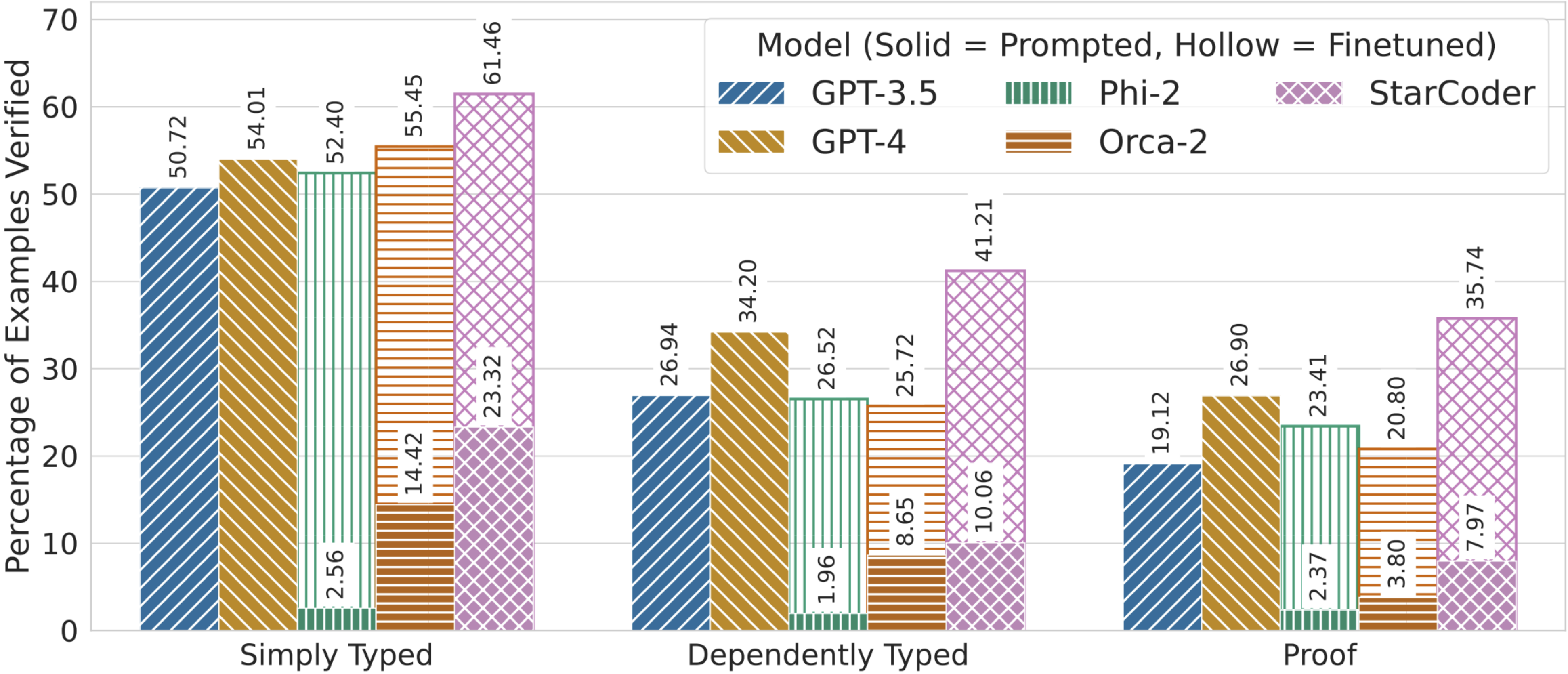
Metric	Train	Valid	Test	
			Intra-project	Cross-project
Number of Definitions	22779	1541	5965	1769
Number of Projects	6	6	6	2
Number of Files	1216	72	306	126
Avg. num of lines	8.66	13.63	11.40	7.45
Avg. num of tokens	92.16	157.26	124.32	60.32
# Simply Typed	6736	434	1248	149
# Dependently Typed	12047	764	3111	1431
# Proofs	3996	343	1606	189

Model results

Prompt setup

- Related examples (RAG)
  - Type similarity using OpenAI embeddings
- Premises
  - What global identifiers are expected?
  - Finetuned embedding model
- Type of the definition to generate (“goal”)

Success rate (verified @ 10)



Small finetuned models outperform GPT-4 on definition synthesis!

Future Directions

- Control: Force LLM to only complete valid identifiers using AICI
- Repair: Iterate generation based on error messages
- Augment: Insert proofs into ML-like code

arXiv: cs.PL/2405.01787