	Data Collection	Pre-processing	Data Validation	SE Testbeds Generation
	1. Repository Filter. 2. Commit filtering. 3. Method Code & Comments extraction.	1. AST Parsing 2. Feature. computation 3. Database push. 4. Distinct deduplication.	1. Meaningfulness validation. 2. Code sequences usefulness evaluation. 3. Feature validation.	1. Raw testbed filtering. 2. *RandomCut* code generation. 3. *Docstring* filters. 4. *Commit* messages filter. 5. Jaccard similatiry de-duplication
Artifacts	{ JSON }	My		{ JSON }