

Ⓒ BalancedDatasetBuilder
clean_path injected_path label_mapping : dict output_path smelly_path
balance_classes(smelly, injected, target_per_smell) build_full_dataset(target_clean, target_per_smell, max_injected) build_injected_only_dataset(max_clean, max_injected) extract_python_code(code) load_json(file_path) normalize_labels(labels) process_injected_functions(injected_functions, max_injected) process_smelly_functions(smelly_functions) remove_comments(code) save_json(data, file_path)

Ⓒ CodeSmellAnalyzer
clean_results : list dataset dataset_path file_cache : dict inspector : Inspector label_mapping : dict log_interval : int logger : NoneType, RootLogger max_workers : int output_dir smelly_results : list
analyze_dataset_parallel() analyze_function_in_file(function_data) load_dataset() run() save_results()

Ⓒ DatasetEvaluator
dataset inspector : Inspector output_path temp_dir : str
analyze_code_similarity(similarity_matrix, threshold) compute_code_similarity(dataset) compute_similarity_distribution(similarity_matrix) exclude_invalid_syntax() extract_python_code(input_string) is_valid_syntax_using_inspector(code) load_dataset(path) load_smell_definitions(path) plot_similarity_distribution(similarity_matrix) process_and_save_results() save_invalid_entries(invalid_entries)

Ⓒ QwenLLM
model_name : str
generate_response(prompt)

Ⓒ FunctionDatasetBuilder
libraries : list repo_path
build_dataset() extract_functions(file_path) get_python_files() save_dataset(dataset, output_path)

Ⓒ RepositoryDownloader
github libraries : list output_folder : str processed_repos : set processed_repos_path : str
load_processed_repos() save_processed_repos() search_and_process_topics(topics, max_repos_per_topic, stars, pushed, language, download)

Ⓒ BaseLLM
<i>generate_response(prompt)</i>

Ⓒ CodeSmellInjector
llm_model max_smells : int smell_descriptions : dict
inject_smells(clean_function, num_smells)

Ⓒ InjectedSmellsDatasetBuilder
checkpoint_path injector input_path output_path timeout_seconds : int
load_checkpoint() process_dataset() process_function_with_timeout(func, index, total_functions) save_checkpoint(processed_indices) save_incremental_output(smelly_functions)

