

GitHub Template Repositories: Served Domains, Maintenance, and Practitioner Guidelines

Leuson Da Silva*, Altaf Allah Abbassi*, Imen Trabelsi*, Paulo Borba†, Foutse Khomh*

*Polytechnique Montréal, Canada

Emails: {leuson-mario-pedro.da-silva, altaf-allah.abbassi, imen.trabelsi, foutse.khomh}@polymtl.ca

†Centro de Informática, Universidade Federal de Pernambuco, Brazil

Email: phmb@cin.ufpe.br

Abstract—Over time, GitHub has introduced different strategies for sharing reusable code artifacts. In addition to fork-based reuse, *template repositories* provide a distinct feature for generating new projects from scaffolding. Although this feature has been available since 2019, little is known about the domains it supports, its maintenance characteristics, or the practices that guide practitioners for effective template design. To address this gap, we conduct a large-scale empirical study of GitHub template repositories across the five most used programming languages. First, we mine and categorize templates to analyze the domains they serve, exploring the LLM-as-a-judge strategy. Next, we explore the reliability of templates by evaluating the associations between repository characteristics and activity, and quality-related issues (e.g., code smells, vulnerabilities, and security hotspots) through statistical analysis. Finally, we qualitatively analyze a representative subset of *templates* to derive practical guidelines and recurring pitfalls for template design and management. Our results show that *Web Development* is the predominant domain across ecosystems, while maintenance and quality issues vary by programming language. We further find that high-quality templates tend to adopt established software engineering practices, while providing comprehensive documentation and clear guidance for use. Overall, our findings offer empirical insights and actionable guidance to support practitioners in designing and adopting high-quality template repositories.

Index Terms—Templates, GitHub, Guidelines, Pitfalls

I. INTRODUCTION

Code and related artifacts are largely shared among developers and practitioners during collaborative software development mainly through version control systems, like *Git* and *Mercurial*. Over time, these tools gained support from online services like GitHub and BitBucket, expanding the resources and features available to practitioners. With advancements in GitHub and related features, they have shaped the current ways practitioners share code. For example, *forks* allow developers to mainly extend repositories created and maintained by others, keeping parallel lines of development [1], while *pull requests* enable contributors to propose changes back to the original project, mediating collaboration and integration of external contributions [2]. On the other hand, *template repositories*¹ represent an alternative to share reusable code artifacts as starting points for new projects [3]. While the former focuses on building software through collaborative development, with external contributors fixing and implementing new features [4],

the latter aims to efficiently speed up the initial steps of a project that shares relevant characteristics with the intended system, such as similar technologies or configurations, with less manual upfront configuration [5], [6].

Knowing that these features have been used for a while, we observe a discrepancy in what is known about them. While forks and pull request practices have received significant attention in prior research [4], [7], templates have mostly been used to support research and educational activities [8], rather than to study how practitioners use them. Furthermore, considering the expressivity of templates, with some of them becoming more popular than regular repositories, like a Microsoft repository focusing on Web Development for beginners [9], it is important to further understand the extent to which developers adopt templates, the purposes they serve, and the practices they are based on. Understanding these dimensions is essential to reasoning about where templates are most prevalent, how they differ across ecosystems, and which risks and benefits they may propagate to downstream projects. Without such understanding, both practitioners and researchers lack guidance on how to design, evaluate, and evolve templates in ways that effectively support software development practices.

Aiming to address this gap, we investigate (i) the domains that *templates* serve for, (ii) how repository characteristics and activity relate to maintenance over time, and (iii) the practical guidelines and pitfalls that could support practitioners when creating and maintaining *templates*. To this end, we conduct an empirical analysis of GitHub *templates*, guided by three main research questions, across the five most widely used programming languages, according to the GitHub Octoverse ranking [10]: TYPESCRIPT, PYTHON, JAVASCRIPT, JAVA, and C#.

First, for our RQ1, we mine *templates* and, supported by LLMs, automatically categorize them to uncover their supported domains, drawing on taxonomies from prior work [11], [12]. Our goal is to investigate and compare the domains in which practitioners employ them and examine how template usage differs across programming languages. For RQ2, we focus on how repository characteristics and activity relate to maintenance and security issues, aiming to evaluate the reliability of templates as reusable artifacts and identify signals that may help practitioners anticipate potential downstream

¹Hereafter referred to as templates.

effort before adoption. To this end, we collect a set of metrics capturing quality-related aspects (e.g., code smells, vulnerabilities, and security hotspots) and analyze their associations with repository characteristics (e.g., stars, forks, and commits) using a regression model. Finally, for RQ3, we select and analyze a set of repositories aiming to identify common patterns and derive actionable guidelines and pitfalls for practitioners when creating and maintaining their *templates*.

Our results show that *Web Development* is the main domain supported by *templates* across all programming languages, whereas more specialized domains tend to be associated with specific ecosystems, such as *Machine Learning/AI*, primarily supported by PYTHON. For RQ2, we observe substantial variability in maintenance across *templates* and programming languages. While some *templates* consistently exhibit low occurrence of bugs, code smells, and vulnerabilities, others concentrate multiple issues. These differences vary across ecosystems, with no single pattern holding uniformly across programming languages. Overall, the results show that maintenance and quality issues are ecosystem-dependent and cannot be treated as homogeneous signals across different programming languages.

For RQ3, as guidelines, we emphasize that *templates* should adopt standard and good practices commonly found in regular GitHub repositories, such as automated pipelines and comprehensive documentation. Beyond these practices, we also reinforce the importance of educating practitioners on properly using *templates*, promoting reuse through families of templates, and aligning releases with supported technology versions to avoid duplication and improve sustainability. As pitfalls, we highlight that repurposing regular repositories as *templates* without proper adaptation, unclear or inactive maintenance status, and poor separation of concerns between templates and technologies can mislead practitioners and hinder effective template reuse.

Overall, our study makes the following contributions:

- A set of guidelines and pitfalls that practitioners should consider when providing and selecting *templates*;
- A holistic view about the adoption of *templates*, covering domains and maintainability;
- A dataset of *templates* covering the top 5 programming languages in GitHub.
- We make our data and results publicly available in our replication package [13].

II. DATA COLLECTION

In this section, we explain in detail our process to establish the sample of repositories to be analyzed, mainly covering the mining and filtering steps of GitHub *templates*. Similar to previous studies, we rely on the GitHub API [14] and follow common GitHub mining guidelines to avoid duplicated, inactive, or non-representative repositories [15]–[18]. Figure 1 summarizes the overall workflow.

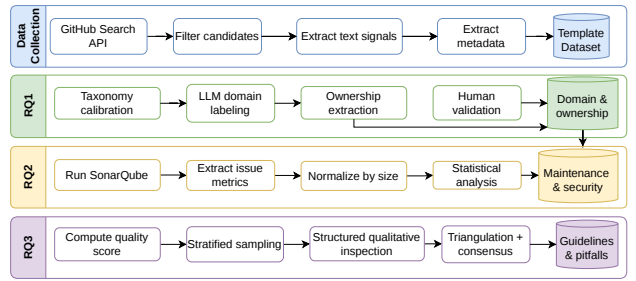


Fig. 1. Overview of the research approach.

A. Template Search Strategy

We rely on the GitHub API to collect our set of *templates*. To the best of our knowledge, the API does not provide official documentation or explicit support for searching exclusively for *templates* [19]. During an in-depth inspection of the API behavior, however, we discovered that including the parameter `template:true` in the search query restricts the results to repositories explicitly marked as *templates*. To assess the reliability of this undocumented mechanism, we contrasted its results with those obtained through a keyword-based heuristic commonly adopted in prior studies. Specifically, we performed an exploratory search using terms frequently associated with *templates*, including *template*, *boilerplate*, *starter*, *skeleton*, and *scaffold*, appearing in repository names, descriptions, and README files. Our comparison showed that all repositories retrieved by the keyword-based approach were also captured by the `template:true` query, which additionally returned a substantially larger set of repositories.² Based on this validation, we adopted the API-based approach relying on the `template:true` parameter for our data collection.

B. Filtering Criteria

We create a PYTHON script with a set of configuration parameters to ensure stable and reproducible queries (e.g., number of repositories per page, maximum pages per query, timeouts, and delay intervals). Regarding the adopted search parameters, we followed well-established guidelines for empirical studies on GitHub, ensuring that our queries targeted active, non-forked, non-archived, and public repositories [16]–[18]. By excluding forked repositories, we aim to avoid duplication and promote a more diverse set of independent projects in our sample. To capture active projects, we further exclude archived repositories and those that have not been updated in the past two years. We only require repositories to have at least two stars as a minimal filter to remove trivial or toy projects, while still preserving a broad and diverse sample of *templates*. We did not impose any restriction on the number of forks, since *templates* are intended to be reused directly rather than forked. Finally, we did not apply any restrictions related to project structure (e.g., minimum number

²We executed the keyword-based heuristic for comparison, but all its results were included, and substantially surpassed, by those retrieved with the `template:true` query.

TABLE I
NUMBER OF MINED REPOSITORIES PER PROGRAMMING LANGUAGE
ACROSS RQ1–RQ3

Programming Language	Mined Templates	RQ1	RQ2/RQ3
TYPESCRIPT	6202	5804	6180
PYTHON	3363	2628	3356
JAVASCRIPT	3708	3176	3368
JAVA	938	740	262
C#	569	470	215
Total	14780	12818	13381

of files) or development history (e.g., minimum period of activity), as these aspects could limit our findings. Such criteria might be suitable for general-purpose repositories that evolve continuously, but they could bias the sample against valid *templates*, which could be seen as incomplete when compared with regular repositories.

Knowing that GitHub limits search results to the first 1,000 matches per query, we instruct our script to automatically divide the search space into adaptive star-count ranges. To estimate the number of repositories in a candidate range, we issue a lightweight search API request for the same query and star interval (with `per_page=1`) and use the `total_count` field returned by the API as an estimate of the number of matches. We then recursively split the star range (e.g., `stars:2..3`, `stars:4..7`, etc.) until each sub-range yields at most 1,000 results. For each star range, our script queries the GitHub Search API using the constructed search parameters. As the API returns at most 100 repositories per page, the script iterates through up to 10 pages per range, accumulating up to 1,000 repositories that meet the criteria. We repeated this process for all programming languages evaluated in this study. Finally, Table I presents the final distribution of mined templates (2nd column).

III. RQ1: WHAT DOMAINS AND OWNERSHIP PATTERNS CHARACTERIZE TEMPLATES?

A. Motivation

When *templates* are created, they are typically designed to address specific goals using particular technologies. These goals often reflect the needs of particular domains, considering the diversity of repositories and ecosystems on GitHub. Accordingly, identifying the domains served by *templates* in our final classification of the analyzed sample allows us to understand where *templates* are most prevalent and where gaps exist. Furthermore, this knowledge also reveals how domain-specific requirements (frameworks, tools, or patterns) might influence template design and maintenance. In addition to the domain, ownership, whether a template is created by an individual developer or an organization, also shapes its purpose, quality, and expected use. Understanding these ownership patterns provides complementary insights into who drives template creation and how *templates* propagate across the ecosystem.

B. Approach

With the final list of projects selected (Table I), we proceed to infer the domains these projects serve. Since we are not aware of prior studies using LLMs to classify the domains of GitHub repositories, we use prior LLM-based SE studies [20], [21] as motivation for employing LLMs as support tools in our classification and judgment tasks. To design and validate the domain taxonomy used in this process, we first conducted a manual analysis on a calibration subsample. Specifically, we randomly selected 100 projects from the PYTHON sample, and one researcher manually categorized them according to domain classifications proposed in prior work [11], [12]. This initial analysis showed that most repositories fit into eight existing categories. However, several projects could not be adequately captured by the original taxonomy. As a result, six additional domains emerged and were incorporated into the final classification schema: *Automation/DevOps*, *Infrastructure/Cloud*, *Research*, *Data Management*, *Robotics/IoT*, and *Data Apps/Visualization*.

Once the categories were defined, concise definitions for each category were created to guide the LLM-based classification. Guided by recent reporting guidelines for empirical SE studies involving LLMs [22], we report the LLM role, model configuration, prompt inputs, output format, and human validation procedure; additional details are available in our online appendix [13]. For each repository, the prompt included the project description, topics, and README content, when available. The model was instructed to (i) summarize the project, (ii) assign exactly one category according to the predefined taxonomy, (iii) justify its decision by briefly explaining the rationale behind the classification, and (iv) report its confidence in the classification using a score from 1 to 5. If a project did not fit any of the predefined categories, the model was allowed to propose a new one. The output was requested in JSON format, including the fields `category`, `reasoning`, and `confidence`. Two models were selected based on their capabilities (GPT-4o-mini and DeepSeek-V3.2), employed with a temperature setting of 0.2, using our manually classified templates as a reference. DeepSeek-V3.2 achieved higher agreement with the manually labeled sample (91/100) compared to GPT-4o-mini (74/100). Therefore, we selected DeepSeek-V3.2 for the remaining classification tasks. During this process, we observed that some project information was not available in English. To ensure consistency in the analysis, we implemented preprocessing routines using `langdetect` [23] for language identification, which filtered out non-English projects based on the language detected in their README files, retaining only English repositories for classification (see Table I, 3rd column). During this process, the LLM proposed only three new domains among more than 12,000 classifications. After manual inspection, we found that these labels were technology-specific refinements of existing categories rather than missing high-level domains. We therefore mapped them to the closest predefined category in our taxonomy.

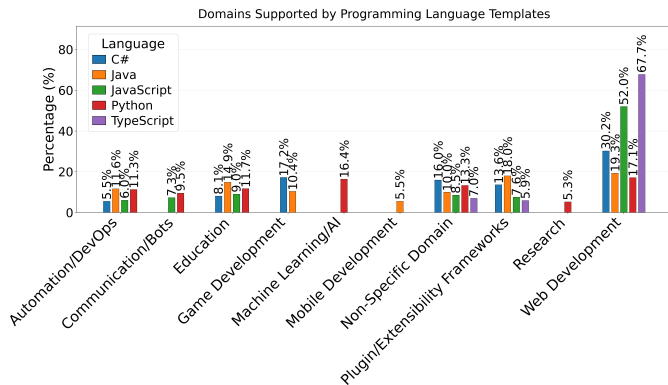


Fig. 2. Distribution of Domains supporting GitHub Templates.

As LLMs were used to classify template domains, this approach entails additional computational and financial costs. In total, the LLM-based classification incurred a cost of approximately \$6 USD. To evaluate the accuracy of the LLM’s classifications, two researchers independently categorized a different random subsample of projects, establishing a ground truth for comparison. Such a manual validation was performed on a statistically representative sample of 192 repositories (95% confidence level with a 7% margin of error). Two authors independently labeled the sampled repositories following the same protocol and using the resulting extended taxonomy. In cases of disagreement, the authors discussed the repositories and reached consensus without consulting the labels produced by the LLM. When comparing these consensus labels against the automated classifications, we observed that the LLM differed from the human consensus in 28 out of 192 cases, corresponding to 85.4% agreement, with Cohen’s $\kappa = 0.82$ specifically measuring the agreement between the automated classifications and the final human-consensus labels.

Finally, to analyze ownership patterns, for each template, we extracted the owner type, distinguishing whether it was maintained by an individual user or an organization (all mined templates, see Table I, 2nd column). After collecting this information, we computed the frequency of each ownership type across all templates. Such an analysis allows us to quantify the distribution of templates between individual and organizational owners and to identify overarching trends in their prevalence.

C. Results

Domain Analysis: Figure 2 presents the distribution of the domains associated with the templates across the five programming languages evaluated here. We report only categories accounting for more than 5% of the repositories. Full details are provided in the online appendix [13]. Overall, we can observe a clear dominance of *Web Development* templates, which represent the most dominant domain across all programming languages. Those results, specifically for `TYPESCRIPT` (67.7%) and `JAVASCRIPT` (52%), are expected since these languages are heavily used in front-end and full-stack

projects, highly supported by web frameworks (React, Next.js, Angular) [10], [24]. For the remaining domains, we observe a more balanced distribution, highlighting *Non-specific Domain*, *Plugin/Extensibility*, and *Education* templates. While the *Non-specific Domain* category provides generic, domain-agnostic software templates for a wide range of software development purposes, the *Plugin/Extensibility* category supports scaffolding projects designed to extend existing resources or frameworks. Meanwhile, *Education* templates focus on offering a structured foundation for projects whose primary goal is teaching or learning, such as course materials, assignments, or programming competitions. These templates often include configurations and structures that facilitate integration with larger platforms (e.g., IDE extensions, web frameworks, CI/CD, and general tools), enabling practitioners to streamline the development of plugins, modules, or reusable components.

Different from *Regular Projects*, *Automation/DevOps* templates provide additional support for various types of automation, including testing environments, CI/CD configurations, and deployment workflows. The remaining domains exhibit either isolated highlights or uniformly small proportions. For example, we observe a high percentage of *Machine Learning/AI* templates in `PYTHON`, while a similar trend appears for *Game Development* in `C#` and `JAVA`. Such an observation aligns with the widespread use of `PYTHON` for ML projects on GitHub [25], as well as the well-established support for game development in `C#` and `JAVA`, considering key game engines, like Unity and JMonkey [10], [26].

Programming Languages Influence: While some programming languages concentrate their presence in a restricted number of domains (`TYPESCRIPT` and `JAVASCRIPT`), we observe that `PYTHON` is the top language covering all domains with substantial representation. Such an observation highlights the versatility and broad applicability of `PYTHON`, whose ecosystem supports diverse software development needs ranging from web and data management to automation and machine learning. In contrast, the restricted presence of other languages across domains underscores their more specialized or framework-oriented nature, suggesting that certain ecosystems foster narrower patterns of reuse and template creation. This finding calls attention to the need for further investigation into how language-specific communities shape the dissemination and generalizability of templates, and also the imminent need for templates in these domains.

Template Ownership: Figure 3 presents the distribution of ownership on the evaluated templates. On GitHub, a repository can be owned either by an organization (typically representing teams or companies) or by an individual user account [27]. Overall, we observe that individual users own the majority of templates, accounting for roughly two-thirds of the total. Although these ownership types primarily differ in purpose, we can infer common motivations for sharing templates. Specifically, organizations tend to share templates to promote standardization and streamline internal development processes, whereas individual users often publish templates to support the open-source community, demonstrate best practices, or

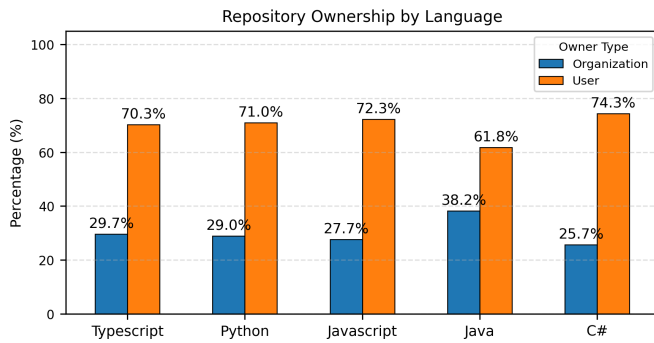


Fig. 3. Distribution of repository ownership types (Organization vs User) across programming languages.

showcase reusable project structures.

IV. RQ2: HOW DO REPOSITORY CHARACTERISTICS AND ACTIVITY OVER TIME RELATE TO MAINTENANCE?

A. Motivation

Templates are primarily designed to be reused as starting points for new projects, reducing initial setup effort. Unlike traditional repositories that evolve independently [28], adopting *templates* may directly propagate their structure, dependencies, and potential flaws to downstream projects [29], [30]. Consequently, the maintenance quality of templates is particularly important, as defects or outdated dependencies may be replicated across multiple projects and contribute to long-term technical debt [31], [32]. In practice, developers often rely on observable repository signals (e.g., activity, popularity, or ownership) when selecting templates, yet it remains unclear whether such signals relate to template reliability. Understanding the maintenance characteristics of templates is therefore essential to assess the risks they may pose across the ecosystem [33], [34]. In this RQ, we investigate how repository characteristics and activity relate to the presence of maintenance and security issues in templates, aiming to identify signals that may help practitioners assess template reliability before adoption.

B. Approach

When evaluating the occurrence of maintenance issues in GitHub repositories, previous studies have focused on several dimensions, including code smells, bugs, reliability issues, and security vulnerabilities. These studies typically rely on static analysis tools to detect, quantify, and report such issues at scale [35], [36]. Motivated by these studies, we aim to investigate the occurrence of quality issues specifically within the context of templates. Specifically, we treat the number of maintenance and security issues as dependent variables, and analyze their associations with repository characteristics and activity indicators.

Since our study covers five different programming languages and requires collecting comparable results across them, it was essential to adopt a tool with broad language support and consistent reporting capabilities. For this reason, we selected

SonarQube [37], a widely used and industry-standard static analysis platform. To operationalize the analysis, our scripts handle all the required steps to run SonarQube and collect metrics. For each template, the script clones the latest available repository’s version, invokes SonarQube on this snapshot, and extracts the resulting quality metrics: *bugs*, *code smells*, *security hotspots*, *general* and *informational vulnerabilities*, and their associated *severity levels*. For exploratory analyses, we normalized issue values by size (per KLOC) to enable fair comparisons across repositories. For compiled programming languages, successfully building the repository is a prerequisite for running SonarQube. However, a subset of JAVA and C# repositories could not be analyzed due to missing dependencies, configuration issues, or build failures. Specifically, 676 JAVA repositories and 354 C# repositories were excluded from the corresponding analyses. For the remaining programming languages, a small number of runs still failed due to SonarQube execution issues (e.g., timeouts or repository-specific configuration errors). These repositories were excluded, resulting in minor reductions of the final samples, as reported in Table I (4th column).

To examine the association between repository characteristics and quality issues, we considered a regression model, allowing us to control for multiple factors simultaneously and assess their independent effects. Such a combination of methods provides a more rigorous understanding of which repository attributes are meaningfully associated with the observed maintenance issues.

First, we merged the metadata extracted from GitHub (RQ1) with the quality metrics obtained from SonarQube: number of *recent commits* (last 12 months), *stars*, and *forks*. From these raw fields, we computed additional explanatory variables: *repo_age_days* (repository’s age in days), *staleness_days* (days since the last update), and logarithmic transformations of stars and forks to reduce skewness in popularity distributions. We also encoded the *repository owner* as a binary variable, indicating whether the owner was an organization or an individual. For regression modeling, we used raw issue counts and incorporated repository size as a logarithmic offset. Finally, we consider the Negative Binomial regression model, which allows us to estimate the independent contribution of each predictor while controlling for multicollinearity. We also remove extreme outcome values to reduce the influence of outliers and assess the multicollinearity among predictors using variance inflation factors (VIF) before model fitting. Finally, we also performed a Mann–Whitney U test to compare issue distributions between organization- and user-owned templates. Since our analyses involve multiple tests across programming languages, predictors, and outcomes, we applied the Benjamini–Hochberg false discovery rate correction within each programming language. Together, these steps provide a comprehensive statistical characterization of how different repository attributes relate to the quality issues detected in templates. We adopt a significance level of $\alpha = 0.05$ for all statistical tests.

C. Results

Overall, we observe a strong dispersion, indicating that quality issues in templates are sparse, highly skewed, and dominated by a few repositories with disproportionately high issue counts. When considering individual metrics, an average of 39–54% of templates per language report zero issues, with the vulnerability metric exhibiting the highest sparsity (approximately 84–96% of repositories reporting zero vulnerabilities). Skewness coefficients range on average from approximately 6 to over 46, indicating long-tailed distributions dominated by a few repositories. These characteristics violate normality assumptions, motivating our reliance on Negative Binomial regression models.

Ownership differences. Organization-owned templates generally exhibit slightly lower issue densities, except for TYPESCRIPT, but the effects are mostly negligible or small. For JAVASCRIPT and PYTHON, significant differences mostly indicate lower issue densities for organization-owned templates, with negligible effects ($r_{rb} \approx -0.02$ to -0.10), except for PYTHON critical vulnerabilities ($r_{rb} = -0.12$). JAVA and C# show small negative effects for vulnerability density ($r_{rb} = -0.14$ and $r_{rb} = -0.11$), suggesting fewer vulnerabilities per KLOC in organization-owned repositories. For TYPESCRIPT, significant differences appear only for critical and informational vulnerability densities, both with negligible positive effects ($r_{rb} = 0.05$ to 0.07). Overall, ownership type explains only a minor portion of variability in vulnerability density and maintenance indicators.

Regression Analysis. We observe that repository-level metadata is associated with maintenance and security issues, but in language- and outcome-dependent ways. The strongest evidence is observed for JAVASCRIPT and TYPESCRIPT, where *log_forks* is positively associated with bugs, code smells, general and informational vulnerabilities, security hotspots, and vulnerability severity levels, with IRRs from 1.20 to 2.29. This suggests expected issue counts ranging from 20% to 129% higher for more diffuse templates in these ecosystems. In contrast, for PYTHON, the only association involving *log_forks* is negative for bugs (IRR = 0.43), indicating 57% lower expected bug counts. Thus, repository diffusion relates to issue counts in an ecosystem-specific rather than universal way.

Recent development activity is associated with issue counts only in specific ecosystems, with varying direction and magnitude. In TYPESCRIPT, *commits_last_12m* is positively associated with most evaluated issues, but with IRRs close to 1.00, indicating marginal increases of about 0.10% to 0.35% per additional commit. In C#, recent commits are positively associated only with security hotspots (IRR = 1.006), while JAVASCRIPT shows mixed and very small corrected associations. Thus, recent activity is a statistically detectable but modest signal, whose practical effect may only become meaningful cumulatively in highly active repositories.

Finally, organization-owned templates are associated with fewer issues in some cases (*owner_is_org*), including JAVA

vulnerabilities (IRR = 0.14), several JAVASCRIPT outcomes (IRR = 0.31–0.73), and C# bugs (IRR = 0.24). Conversely, they are associated with more bugs in JAVA and TYPESCRIPT (IRR = 2.98 and 2.79), as well as more TYPESCRIPT code smells and selected vulnerability severity levels (IRR = 1.25–1.62). Thus, ownership likely reflects ecosystem-specific practices, scale, and reporting processes, and should not be interpreted as a uniform quality signal across languages.

V. RQ3: WHAT ACTIONABLE INSIGHTS CAN SUPPORT PRACTITIONERS IN CREATING, MAINTAINING, AND ADOPTING TEMPLATES EFFECTIVELY?

A. Motivation

Templates are expected to accelerate project setup and promote reuse by providing structured scaffolding for new projects. However, as observed in RQ1 and RQ2, *templates* span diverse ecosystems and may exhibit maintenance and security issues. This highlights the importance of understanding how *templates* should be designed and maintained in practice. Therefore, beyond characterizing template repositories, it is important to derive actionable insights that can guide practitioners in the effective creation, management, and adoption of *templates*. In RQ3, we aim to extract practical guidelines and identify recurring pitfalls that can help practitioners better design, maintain, and use *templates*.

B. Approach

To identify actionable insights for practitioners, we performed a targeted qualitative analysis of a representative subset of templates. For that, first, we computed a composite quality score for each repository by combining the multiple maintenance and security metrics from RQ2, further normalizing them by project size (LOC, files, or functions). Then, each normalized metric was transformed into a percentile rank to ensure robustness to skewed distributions and outliers, as previously observed in our sample. The composite score corresponds to the average percentile rank across all available metrics, with lower values indicating better overall quality. Based on this score, we selected three complementary groups of repositories per programming language: best (20, lowest composite scores), worst (20, highest composite scores), and mixed repositories (10), characterized by high variability across quality dimensions. To mitigate sampling bias, selection was stratified by ownership type (organization vs. individual) and repository popularity (stars). This sampling strategy enables a focused yet representative qualitative examination of recurring quality patterns and practices that can support practitioners when selecting, reusing, and creating new templates.

To support a systematic qualitative analysis, we used predefined CSV templates to guide data extraction for each repository. The templates captured: (i) the repository goal and whether it was aligned with the target repository goal, (ii) quality and maintenance practices, and (iii) observed strengths, issues, and candidate actionable insights. Before the main analysis, two researchers conducted a calibration phase on a small random subset of templates outside the selected sample

to refine the extraction fields, align the unit of analysis, and define the criteria for identifying actionable observations. Next, both researchers independently analyzed the sampled repositories and recorded their observations in the CSV files. The resulting observations were compared and grouped into recurring patterns. Patterns associated with desirable practices were consolidated as *guidelines*, while recurring problematic practices were consolidated as *pitfalls*. Disagreements were resolved by re-examining the repositories and RQ2 metrics. Finally, the agreed set of guidelines and pitfalls was reviewed by the remaining authors.

C. Results

First, we discuss the set of guidelines observed as good practices, and next, the set of pitfalls.

1) Guidelines:

G1: Integrate maintenance and quality practices

Templates reported with the best scores in our analysis consistently adopted established software engineering practices. In particular, they provided ready-to-use project setups and integrated common quality mechanisms, such as linting, testing at appropriate levels, CI/CD pipelines, hooks, automation tools, and bots, for different purposes, like dependency updates and code review. These practices not only help promote maintainability and improve the quality of downstream projects instantiated from the templates, but may also serve an educational role by introducing good practices to new practitioners.

G2: Provide comprehensive/accessible documentation

Beyond explaining the project's purpose, templates must guide users in understanding, customizing, and extending the provided structure to fit their specific needs. To this end, effective documentation can take multiple forms, including step-by-step guides, code snippets with explanatory descriptions, links to external resources, and creation of wikis [38]. We also observe some templates providing documentation in different languages, supporting reuse while avoiding the need to create new templates to broadly support the same initial goal [39].

G3: Educate users on how to GitHub Template Feature

Although GitHub's templating feature is not new, we observe that several repositories provide inaccurate or incomplete instructions on how to properly use it. In particular, users are often instructed to clone the repository directly, rather than using the *Use this template* button [40]. Templates should explicitly guide users toward the correct instantiation mechanism and update their documentation accordingly, especially when existing repositories are later converted into templates.

G4: Communicate evolution through a roadmap

GitHub repositories usually provide roadmaps to communicate upcoming changes and planned features. Templates can equally benefit from this practice, as a roadmap helps users

understand how the template is expected to evolve and how it can be extended over time. This is particularly valuable for templates targeting the *Education* domain, where roadmaps can guide learners by informing tasks to be performed and expected outcomes [41], [42].

G5: Organize templates as a reusable family

Templates can be designed as a reusable family [43], where new templates are derived from existing ones to preserve and propagate previously established good practices while introducing technology-specific variations. We observe this approach when practitioners define a default template and subsequently use it as the basis for creating additional templates targeting different technologies, ecosystems, and programming languages [44].

G6: Align technology versions with releases

Our analysis shows that many templates are created to support the use of specific technologies, often alongside libraries or frameworks provided by organizations or individual developers. As these technologies evolve, templates should explicitly track the supported technology versions and associate them with corresponding template releases [45]. Providing clear release information, for example, through versioned releases, allows practitioners to identify compatible template versions while supporting reproducibility over time [46].

G7: Encourage community and social support

Collaborative software development inherently relies on community engagement through discussion, contribution, and shared problem-solving. For templates, this engagement can be impacted by how practitioners consume, adapt, and extend the provided templates. In our analysis, we observed some initiatives aimed at fostering active communities around templates, such as dedicated communication channels (e.g., Discord servers) to support users. Similarly, some template providers include calls for sponsorship, helping sustain long-term maintenance in ways comparable to traditional GitHub repositories [47].

2) Pitfalls:

P1: Turning non-templates into templates

Many repositories labeled as templates were originally designed as full implementations, such as end-to-end applications, libraries, or frameworks [48]. While some templates may provide useful examples of how to use specific technologies, non-templates often lack the scaffolding, guidance, and extensibility expected from a template. As a result, practitioners should avoid turning regular repositories into *templates* unless they are explicitly adapted to support reuse, learning, and extension. This issue can be mitigated by following Guideline G3, which helps ensure that repositories are intentionally designed and used as reusable starting points rather than being repurposed from full implementations.

P2: Unclear or inactive repository status

Several inactive templates fail to clearly communicate their current status to practitioners [49]. To avoid misleading users, template providers should explicitly signal inactivity, for example, by archiving the repository or clearly documenting its status. While we observed some cases where inactivity was mentioned in the README, such information is less visible and may be overlooked compared to GitHub’s native archival mechanisms [50]. This issue can be mitigated by following Guidelines G4 and G6, which help communicate the evolution and maintenance status of templates over time.

P3: Multiple templates in a single repository

GitHub repositories are typically designed as single entities with a clear and focused goal, and *templates* are no exception. However, we observed *templates* targeting different programming languages or ecosystems within the same repository, distributed across folders, or even separate *branches* [51]. Such an organization increases complexity and can negatively impact discoverability, maintenance, and reuse. To mitigate this issue, practitioners should identify the common core across such variants and structure them as a family of templates, following the approach discussed in Guideline G5.

P4: Mixing technology and template concerns

Repositories hosting technologies on GitHub typically serve as primary points for implementation, documentation, issue tracking, and releases. *Templates*, however, address a different concern: providing reusable scaffolding rather than serving as the primary implementation artifact. During our analysis, we observed cases where technology implementations and their corresponding templates were colocated within the same repository [52]. Such an organization can be confusing for practitioners and may hinder long-term evolution by mixing product-level changes with template-specific concerns. Guideline G5 helps address this issue by promoting the organization of templates as a family across technologies.

P5: Empty templates and missing documentation

While *templates* are expected to provide reusable scaffolding, we observe some empty repositories, where no files were available, only the structure of folders. Similarly, some *templates* also lack basic documentation, including a repository description or a README file. These *templates* provide insufficient guidance and offer limited support for practitioners aiming to understand, adapt, or extend them. *Template* providers should include at least minimal functional content and documentation, as well as an overview of the repository structure, tooling (e.g., pipelines), and, when appropriate, high-level graphical representations [53]. Guideline G2 helps ensure that *templates* provide sufficient documentation to users.

3) *Evaluating Guidelines/Pitfalls Adherence through LLMs*: After defining the guidelines and pitfalls, we explored whether LLMs could support practitioners in preliminarily assessing

repository adherence to them. For that, we used GPT-5.2 with web search capabilities to evaluate a subset of projects previously analyzed in RQ3 (Section V), providing each guideline/pitfall and its description as input. A co-author not involved in defining or validating the guidelines manually evaluated the same projects, enabling us to compare human and LLM assessments. The total cost of the LLM evaluation was approximately USD 1.40. The comparison shows moderate agreement between human and LLM evaluations, with Cohen’s $\kappa = 0.46$, accuracy of 0.80, and macro F1-score of 0.73. These results suggest that LLMs can reasonably approximate human judgments when assessing adherence to the proposed guidelines and pitfalls; however, human validation remains necessary for more interpretative criteria. Agreement was higher for structural or easily observable criteria, such as G1 (software engineering practices) and P4/P5 (mixing technology and template concerns, and empty templates) ($\kappa = 0.60$ – 0.77). In contrast, lower agreement was observed for more interpretative criteria, G2 (documentation quality) and G7 (community engagement), which require deeper contextual understanding and subjective interpretation. Overall, these results indicate that LLMs can serve as useful assistants for preliminary template evaluation, while human validation remains important for criteria requiring deeper contextual interpretation.

VI. DISCUSSION

Our findings suggest that templates should be understood as upstream reusable artifacts rather than simple project starters. As reported in RQ2 (Section IV), templates may exhibit multiple quality issues, including code smells, bugs, and vulnerabilities. Since repositories instantiated from templates are not automatically synchronized with their original *template* repositories by default, issues present at instantiation time may persist downstream, while later fixes are not automatically propagated. If left unresolved, these inherited issues can contribute to template-specific technical debt over time [54], [55].

Template Consumers and Authors: For template consumers, our results indicate that adoption decisions should go beyond signals such as stars, forks, or ownership. Although these signals may suggest popularity or organizational support, RQ2 shows that their relationship with quality issues is ecosystem-dependent. Consumers should therefore inspect dependency freshness, documentation quality, automation support, supported versions, and known maintainability/security issues before adoption. For template authors, our findings suggest that templates should be maintained as reusable products rather than static examples. Authors should provide clear versioning, deprecation/status signals, migration guidance, and separation between reusable scaffolding and full implementations. Finally, since templates can be copied across projects and used as learning resources, poor template design may normalize inadequate practices, increase adaptation effort, and amplify local maintenance problems across broader ecosystems.

Although our primary classification focuses on served domains rather than an explicit taxonomy of template purposes,

our qualitative analysis suggests that templates play multiple roles in practice. Some templates act as educational or tutorial scaffolds, helping newcomers learn technologies and adopt good development practices. Others serve as framework starters or organizational boilerplates, aiming to standardize project setup, dependencies, automation, and release practices. Such a variation in purpose helps to explain why documentation, versioning, and separation between template and implementation concerns are central to our guidelines and to the pitfalls.

Researchers: When selecting and evaluating repositories from GitHub, researchers should be aware of the potential inclusion of templates in their samples. Suppose the goal is to analyze the characteristics or practices of actual software development, then templates should be filtered out, as they do not represent complete or functional projects but rather scaffolds intended for future development. Failing to identify and exclude them may introduce bias in empirical analyses, especially those examining development activity, team collaboration, or software evolution.

GitHub Maintainers: To facilitate future studies on templates, it is essential that GitHub provides better support for identifying and collecting templates. Currently, the search API allows users to query exclusively for templates, although we could not find official documentation. Without such support, researchers could adopt keyword heuristics, as we initially did here. However, such a decision can negatively affect dataset quality and introduce bias in empirical studies. Providing dedicated filters or metadata for templates would improve the reproducibility and accuracy of research that depends on large-scale repository mining.

VII. RELATED WORK

A. Template Repositories

Some studies primarily focused on using and evaluating the adoption of templates as a tool to promote and support students in courses [56]–[58]. For example, Smith et al. [5] explored how GitHub templates can facilitate software engineering education by providing a repository template as a structured starting point for student projects. For that, the authors propose a course model based on GitHub templates to scaffold learning activities, streamline project setup, and standardize evaluation. They report that the adoption of templates effectively reduced configuration overhead, encouraged good development practices, and supported reproducible coursework outcomes. Such an educational perspective complements our results presented here, illustrating one of the supported domains behind template creation (*Education*).

Similarly, Tu et al. [6] investigate the role of GitHub in supporting structured project development in the educational context. Specifically, the authors create repositories using them as templates to support students in initializing their projects with standardized structures, predefined configurations, and consistent development practices. Although these repositories were not formally labeled as *templates*, their purpose aligns with the fundamental idea of GitHub templates, providing

reusable foundations that ensure consistency and reduce configuration effort. Such a pedagogical perspective reinforces one of the main motivations behind template creation observed in our study: lowering entry barriers and promoting uniform project organization.

While the previous studies focus on the educational domain, our findings demonstrate that similar principles of reusability and standardization also drive template adoption across industrial and open-source domains. Together, these studies suggest that GitHub templates are evolving into important vehicles for both learning and professional software engineering practice, bridging the gap between scaffolding for education and reuse in real-world development.

Beyond template-based studies, previous work covering GitHub README files provides an important and relevant perspective on understanding reusable scaffolding artifacts. In this context, README files have been studied as a central artifact for communicating project goals, installation steps, usage examples, contribution expectations, and quality-related information [59]–[62]. These studies show that repository documentation is closely related to how projects are understood, adopted, and maintained. Such a perspective is particularly relevant for template repositories, since templates must not only provide code but also explain how the repository should be instantiated, customized, and evolved. While these previous studies analyze documentation practices in general GitHub repositories, our work complements them by studying repository-level templates as reusable artifacts and deriving guidelines that include documentation, maintenance, and reuse-specific concerns.

B. Collaborative Development and Structured Templates

Over time, GitHub has become the largest ecosystem for collaborative software development, being extensively studied across multiple dimensions. For example, previous studies have investigated pull-based development workflows [2], fork-based reuse [4], and project popularity [17]. Previous studies have also investigated software evolution in depth, covering how collaborations occur and diverge with each other [63], [64], quality and maintenance characteristics evolve [65], and the impact of adopting automation tools [66]. Researchers have also proposed guidelines for conducting empirical studies using GitHub data [15], [67], while other studies have examined best practices and pitfalls related to communication and collaboration within GitHub projects [68]. However, most of this literature focuses on traditional repositories, exploring different mechanisms, such as forks and pull-requests (PRs), leaving templates largely unexplored.

With respect to repository management, GitHub provides multiple features exploring the general concept of *templates*, including support for issue reports, pull requests, and GitHub Action workflows [69]. Prior research has examined the adoption and impact of such templates in contribution-level contexts. For example, Sülün et al. [70] analyzed 350 templates from 100 projects while collecting insights from practitioners, showing that structured issue templates reduce reporting time

and improve information quality. Similarly, Zhang et al. [71] conducted an empirical study on pull request templates, observing that structured templates enhance review consistency and reduce the time and effort required during code review. However, these studies focus on contribution-level templates rather than repository-level templates, as we investigate here.

VIII. THREATS TO VALIDITY

Construct to Validity. As previously informed, the GitHub API does not officially support searching for templates. To mitigate this limitation, we initially employed a keyword-based heuristic as a proxy for identifying potential templates. However, during further experimentation, we discovered that including the parameter `template:true` in the query reliably returned repositories explicitly marked as templates, which we adopted to strengthen the accuracy and completeness of our dataset. Nevertheless, to ensure consistency and reduce potential bias, we restricted our analysis to repositories explicitly labeled as templates in our final dataset.

Our taxonomy of domains was defined through manual analysis conducted by a single researcher, possibly introducing classification bias. Aiming to mitigate this risk, two independent researchers performed additional analyses, including validation against LLM-generated categorization, to improve consistency and robustness. When comparing human- and LLM-based domain classifications for RQ1, we observed some disagreements. By manually inspecting them, most conflicts occurred in repositories crossing multiple domains, where classification boundaries are inherently ambiguous rather than completely incorrect. Although minor differences in categorization were observed, the overall domain distribution and reported trends remained consistent. Therefore, we believe that these discrepancies may not directly affect the validity of the reported results.

Internal Validity. When investigating the factors associated with quality issues in RQ2, we highlight that the associations should not be interpreted causally. Repository-level characteristics such as recent activity, size, and popularity may be inter-related, and confounding factors, like the domain complexity or framework usage, may influence the observed relationships. Although we controlled for size and assessed multicollinearity, unobserved factors may still affect the results.

Threats to Reliability. Although qualitative analysis inherently involves interpretation, we mitigated individual researcher bias through independent analysis and consensus-based consolidation, reducing the influence of personal perspectives on the reported insights. Our LLM-based domain classification also used a single run per repository. While we used a low temperature setting and manually validated a representative sample, we acknowledge that we do not directly measure run-to-run variability in the generated labels.

Conclusion Validity. To assess potential multicollinearity among predictors, we computed variance inflation factors (VIF) for all models. All variables exhibit VIF values below commonly accepted thresholds (around 1), indicating that multicollinearity does not materially affect our estimates and that

observed changes in coefficient significance reflect conditional effects rather than unstable regression artifacts.

External Validity. By explicitly relying on repositories marked as *templates*, we may have excluded repositories that serve similar purposes but are not properly marked as such. This limitation is relevant in our context, considering that the template feature was only introduced by GitHub in 2019 [72]. Although our study covers the five most used programming languages on GitHub, our results may not generalize to templates written in other programming languages.

IX. CONCLUSION

Templates represent a relatively recent mechanism for sharing code on GitHub, focusing on providing an initial structure that practitioners can learn from and extend when building new solutions. In this study, we aimed to further understand the domains that templates support, the factors associated with maintenance and quality issues, and the guidelines and pitfalls practitioners should consider when proposing or selecting templates. To achieve this goal, we conducted a large-scale empirical study of GitHub templates across the five most popular programming languages on the platform (TYPESCRIPT, PYTHON, JAVASCRIPT, JAVA, and C#).

Our results show that templates span diverse domains, with *Web Development* being the most prevalent, a pattern consistently observed across programming languages. Regarding maintenance and quality issues, we observe substantial variability, with repositories exhibiting different levels and densities of issues. These findings indicate that maintenance and quality issues in templates are strongly ecosystem-dependent rather than uniformly distributed. Finally, we provide a set of actionable guidelines and recurring pitfalls, highlighting how design, documentation, automation, versioning, and governance decisions in templates can directly influence their reusability and the quality of projects instantiated from them.

By providing the first large-scale empirical analysis of GitHub *templates*, this study contributes to a deeper understanding of how reusable scaffolding repositories shape modern software development practices. For researchers, our findings open avenues for future work on the long-term impact of templates on software quality, the evolution of template families, and the role of templates in disseminating best practices across ecosystems. Furthermore, we also report implications for the adoption of templates in empirical research and discuss the need for improved support from GitHub maintainers to facilitate mining template-related information.

AI-ASSISTED CONTENT DISCLOSURE

LLMs were used to assist the authors with grammar revision and support for some scripts used in this study. All outputs were reviewed and validated by the authors.

ACKNOWLEDGMENTS

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institute for Advanced Research, and the Canada Research Chairs Program.

REFERENCES

- [1] L. Ren, S. Zhou, and C. Kästner, "Forks insight: Providing an overview of github forks," in *Proceedings of the 40th International Conference on Software Engineering: Companion Proceedings*, 2018, pp. 179–180.
- [2] G. Gousios, M. Pinzger, and A. v. Deursen, "An exploratory study of the pull-based software development model," in *Proceedings of the 36th international conference on software engineering*, 2014, pp. 345–355.
- [3] B. Jackson, "Generate new repositories with repository templates," <https://github.blog/developer-skills/github/generate-new-repositories-with-repository-templates/>, 2019, blog post, published June 6, 2019; updated May 14, 2021.
- [4] J. Jiang, D. Lo, J. He, X. Xia, P. S. Kochhar, and L. Zhang, "Why and how developers fork what from whom in github," *Empirical Software Engineering*, vol. 22, no. 1, pp. 547–578, 2017.
- [5] S. Smith, C. W. Schankula, L. Dutton, and C. K. Anand, "A software engineering capstone course facilitated by github templates," *arXiv preprint arXiv:2410.12114*, 2024.
- [6] Y.-C. Tu, V. Terragni, E. Tempero, A. Shakil, A. Meads, N. Giacaman, A. Fowler, and K. Blincoe, "Github in the classroom: Lessons learnt," in *Proceedings of the 24th Australasian Computing Education Conference*, 2022, pp. 163–172.
- [7] S. Zhou, B. Vasilescu, and C. Kästner, "How has forking changed in the last 20 years? a study of hard forks on github," in *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*, 2020, pp. 445–456.
- [8] Google Scholar, "Search results for "GitHub template repository"," https://scholar.google.com/scholar?start=10&q=%22github+template+repository%22&hl=en&as_sdt=0,5, 2026, accessed: 2026-05-30.
- [9] Microsoft, "24 lessons, 12 weeks, get started as a web developer," <https://github.com/microsoft/Web-Dev-For-Beginners>, accessed: 2026-02-06.
- [10] GitHub Staff. (2025, October) Octoverse: A new developer joins github every second as ai leads typescript to #1. GitHub Blog. Accessed: 2026-01-26. [Online]. Available: <https://github.blog/news-insights/octoverse/octoverse-a-new-developer-joins-github-every-second-as-ai-leads-typescript-to-1/>
- [11] C. Sas, A. Capiluppi, C. Di Sipio, J. Di Rocco, and D. Di Ruscio, "Gitranking: A ranking of github topics for software classification using active sampling," *Software: Practice and Experience*, vol. 53, no. 10, pp. 1982–2006, 2023.
- [12] F. Zanartu, C. Treude, B. Cartaxo, H. S. Borges, P. Moura, M. Wagner, and G. Pinto, "Automatically categorising github repositories by application domain," *arXiv preprint arXiv:2208.00269*, 2022.
- [13] L. D. Silva, A. A. Abbassi, I. Trabelsi, P. Borba, and F. Khomh, "Replication package for "github template repositories: Served domains, reliability, and practitioner guidelines"," 2026. [Online]. Available: <https://github.com/leusonmario/template-repository-analysis>
- [14] V. Cosentino, J. Luis, and J. Cabot, "Findings from github: methods, datasets and limitations," in *Proceedings of the 13th international conference on mining software repositories*, 2016, pp. 137–141.
- [15] E. Kalliamvakou, G. Gousios, K. Blincoe, L. Singer, D. M. German, and D. Damian, "An in-depth study of the promises and perils of mining github," *Empirical Software Engineering*, vol. 21, no. 5, pp. 2035–2071, 2016.
- [16] N. Munaiah, S. Kroh, C. Cabrey, and M. Nagappan, "Curating github for engineered software projects," *Empirical Software Engineering*, vol. 22, no. 6, pp. 3219–3253, 2017.
- [17] H. Borges, A. Hora, and M. T. Valente, "Understanding the factors that impact the popularity of github repositories," in *2016 IEEE international conference on software maintenance and evolution (ICSME)*. IEEE, 2016, pp. 334–344.
- [18] F. Zampetti, F. Belias, C. Zid, G. Antoniol, and M. Di Penta, "An empirical study on the fault-inducing effect of functional constructs in python," in *2022 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2022, pp. 47–58.
- [19] GitHub Community, "Unsupported Search for Template Repositories on GitHub," <https://github.com/orgs/community/discussions/68572>, 2023, accessed: 2026-01-26.
- [20] A. A. Abbassi, L. Da Silva, A. Nikanjam, and F. Khomh, "Unveiling inefficiencies in llm-generated code: Toward a comprehensive taxonomy," *arXiv preprint arXiv:2503.06327*, 2025.
- [21] Z. Sollenberger, J. Patel, C. Munley, A. Jarmusch, and S. Chandrasekaran, "Llm4vv: Exploring llm-as-a-judge for validation and verification test suites," in *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, 2024, pp. 1885–1893.
- [22] S. Baltes, F. Angermeir, C. Arora, M. M. Barón, C. Chen, L. Böhme, F. Calefato, N. Ernst, D. Falessi, B. Fitzgerald *et al.*, "Guidelines for empirical studies in software engineering involving large language models," *arXiv preprint arXiv:2508.15503*, 2025.
- [23] Mimino666, "langdetect," <https://pypi.org/project/langdetect/>, 2025, version 1.0.9; accessed: 2026-01-26.
- [24] S. DI MEGLIO, L. L. L. Starace, V. Pontillo, R. Opdebeeck, C. De Roover, and S. Di Martino, "Investigating the adoption and maintenance of web gui testing: Insights from github repositories," *Available at SSRN 5182165*, 2025.
- [25] S. Raschka, J. Patterson, and C. Nolet, "Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information*, vol. 11, no. 4, p. 193, 2020.
- [26] K. H. Sharif and S. Y. Ameen, "Game engines evaluation for serious game development in education," in *2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 2021, pp. 1–6.
- [27] N. Zöllner, J. H. Morgan, and T. Schröder, "A topology of groups: What github can tell us about online collaboration," *Technological Forecasting and Social Change*, vol. 161, p. 120291, 2020.
- [28] I. Rehman, D. Wang, R. G. Kula, T. Ishio, and K. Matsumoto, "New-comer oss-candidates: Characterizing contributions of novice developers to github," *Empirical Software Engineering*, vol. 27, no. 5, p. 109, 2022.
- [29] Y. Wu, Z. Yu, M. Wen, Q. Li, D. Zou, and H. Jin, "Understanding the threats of upstream vulnerabilities to downstream projects in the maven ecosystem," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2023, pp. 1046–1058.
- [30] B. Lazarine, Z. Zhang, A. Sachdeva, S. Samtani, and H. Zhu, "Exploring the propagation of vulnerabilities from github repositories hosted by major technology organizations," in *Proceedings of the 15th Workshop on Cyber Security Experimentation and Test*, 2022, pp. 145–150.
- [31] R. He, H. He, Y. Zhang, and M. Zhou, "Automating dependency updates in practice: An exploratory study on github dependabot," *IEEE Transactions on Software Engineering*, vol. 49, no. 8, pp. 4004–4022, 2023.
- [32] F. R. Cogo, G. A. Oliva, and A. E. Hassan, "Deprecation of packages and releases in software ecosystems: A case study on npm," *IEEE Transactions on Software Engineering*, vol. 48, no. 7, pp. 2208–2223, 2021.
- [33] F. Palomba, D. A. Tamburri, F. A. Fontana, R. Oliveto, A. Zaidman, and A. Serebrenik, "Beyond technical aspects: How do community smells influence the intensity of code smells?" *IEEE transactions on software engineering*, vol. 47, no. 1, pp. 108–129, 2018.
- [34] S. H. B. I. Kumar, L. R. Sampaio, A. Martin, A. Brito, and C. Fetzer, "A comprehensive study on the impact of vulnerable dependencies on open-source software," in *2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2024, pp. 96–107.
- [35] J. Lefever, Y. Cai, H. Cervantes, R. Kazman, and H. Fang, "On the lack of consensus among technical debt detection tools," in *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 2021, pp. 121–130.
- [36] V. Lenarduzzi, N. Saarimäki, and D. Taibi, "Some sonarqube issues have a significant but small effect on faults and changes. a large-scale empirical study," *Journal of Systems and Software*, vol. 170, p. 110750, 2020.
- [37] SonarSource, "SonarQube Community Build," <https://www.sonarsource.com/products/sonarqube/>, 2025, version Community Build v25.11.0.114957; accessed: 2026-01-26.
- [38] rife2, "Bootstrap project structure to create a rife2 web application with gradle," <https://github.com/rife2/rife2-gradle-hello>, accessed: 2026-02-06.
- [39] lgwk42, "Spring boot + java — auth template," <https://github.com/lgwk42/auth-template>, accessed: 2026-02-06.
- [40] gpauloski, "Python package template," <https://github.com/gpauloski/python-template>, accessed: 2026-02-06.
- [41] The-Marcy-Lab-School, "Technical-prework-1," <https://github.com/The-Marcy-Lab-School/technical-prework-1>, accessed: 2026-02-06.
- [42] hezean, "Sustc: Template for project 2 of sustech principles of database systems," <https://github.com/hezean/sustc>, accessed: 2026-02-06.

- [43] L. Montalvilho and O. Díaz, "Tuning github for spl development: branching models & repository operations for product engineers," in *Proceedings of the 19th International Conference on Software Product Line*, 2015, pp. 111–120.
- [44] thiago roock, "Microservice-application: Modelo para criação de microsserviços no estado da arte," <https://github.com/thiago-roock/Microservice-Application>, accessed: 2026-02-06.
- [45] Jaxelr, "Vuesimpletemplate: A minimal dotnet template for vue.js 3.x," <https://github.com/Jaxelr/VueSimpleTemplate>, accessed: 2026-02-06.
- [46] BunnyWay, "Simple middleware template," <https://github.com/BunnyWay/es-simple-middleware>, accessed: 2026-02-06.
- [47] SpongePowered, "A template for making a sponge plugin for spongeapi 8.0.0," <https://github.com/SpongePowered/sponge-plugin-template>, accessed: 2026-02-06.
- [48] lschoe, "Mpyc: Multiparty computation in python," <https://github.com/lischoe/mpyc>, accessed: 2026-02-06.
- [49] learnthisacademy, "learnthisacademy/nodejs-template," <https://github.com/learnthisacademy/nodejs-template>, accessed: 2026-02-06.
- [50] EnderJs-CreatorGL, "Enderjs-creatorgl/lobo-bot-lite," <https://github.com/EnderJs-CreatorGL/Lobo-Bot-Lite>, accessed: 2026-02-06.
- [51] CleanroomMC, "Mod development template for cleanroom," <https://github.com/CleanroomMC/CleanroomModTemplate>, accessed: 2026-02-06.
- [52] abersheeran, "abersheeran/mingshe," <https://github.com/abersheeran/mingshe>, accessed: 2026-02-06.
- [53] NikolayIT, "Asp.net-core-template," <https://github.com/NikolayIT/ASP.NET-Core-Template>, accessed: 2026-02-06.
- [54] A.-J. Molnar and S. Motogna, "Long-term evaluation of technical debt in open-source software," in *Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2020, pp. 1–9.
- [55] G. Digkas, A. Chatzigeorgiou, A. Ampatzoglou, and P. Avgeriou, "Can clean new code reduce technical debt density?" *IEEE Transactions on Software Engineering*, vol. 48, no. 5, pp. 1705–1721, 2020.
- [56] A. Van Craen, M. Breyer, and D. Pflüger, "Introduction to parallel and distributed programming using n-body simulations," in *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, 2024, pp. 347–354.
- [57] M. D. Beckman, M. Çetinkaya-Rundel, N. J. Horton, C. W. Rundel, A. J. Sullivan, and M. Tackett, "Implementing version control with git and github as a learning objective in statistics and data science courses," *Journal of Statistics and Data Science Education*, vol. 29, no. suppl, pp. S132–S144, 2021.
- [58] A. D. Snowberger and C. H. Lee, "A workflow for practical programming class management using github pages and github classroom," *Journal of Practical Engineering Education*, vol. 15, no. 2, pp. 331–339, 2023.
- [59] G. A. A. Prana, C. Treude, F. Thung, T. Atapattu, and D. Lo, "Categorizing the content of github readme files," *Empirical Software Engineering*, vol. 24, no. 3, pp. 1296–1327, 2019.
- [60] Y. Liu, E. Noei, and K. Lyons, "How readme files are structured in open source java projects," *Information and Software Technology*, vol. 148, p. 106924, 2022.
- [61] T. Wang, S. Wang, and T.-H. P. Chen, "Study the correlation between the readme file of github projects and their popularity," *Journal of Systems and Software*, vol. 205, p. 111806, 2023.
- [62] M. Gaughan, K. Champion, S. Hwang, and A. Shaw, "The introduction of readme and contributing files in open source software development," in *2025 IEEE/ACM 18th International Conference on Cooperative and Human Aspects of Software Engineering (CHASE)*. IEEE, 2025, pp. 191–202.
- [63] L. Da Silva, P. Borba, T. Maciel, W. Mahmood, T. Berger, J. Moisakis, A. Gomes, and V. Leite, "Detecting semantic conflicts with unit tests," *Journal of Systems and Software*, vol. 214, p. 112070, 2024.
- [64] L. Da Silva, P. Borba, and A. Pires, "Build conflicts in the wild," *Journal of Software: Evolution and Process*, vol. 34, no. 4, p. e2441, 2022.
- [65] V. Lenarduzzi, N. Saarimäki, and D. Taibi, "The technical debt dataset," in *Proceedings of the fifteenth international conference on predictive models and data analytics in software engineering*, 2019, pp. 2–11.
- [66] T. Kinsman, M. Wessel, M. A. Gerosa, and C. Treude, "How do software developers use github actions to automate their workflows?" in *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. IEEE, 2021, pp. 420–431.
- [67] E. Kalliamvakou, G. Gousios, K. Blincoe, L. Singer, D. M. German, and D. Damian, "The promises and perils of mining github," in *Proceedings of the 11th working conference on mining software repositories*, 2014, pp. 92–101.
- [68] H. Hata, N. Novielli, S. Baltes, R. G. Kula, and C. Treude, "Github discussions: An exploratory study of early adoption," *Empirical Software Engineering*, vol. 27, no. 1, p. 3, 2022.
- [69] Z. Li, Y. Yu, T. Wang, Y. Lei, Y. Wang, and H. Wang, "To follow or not to follow: Understanding issue/pull-request templates on github," *IEEE Transactions on Software Engineering*, vol. 49, no. 4, pp. 2530–2544, 2022.
- [70] E. Sülün, M. Saçakçı, and E. Tüzün, "An empirical analysis of issue templates usage in large-scale projects on github," *ACM Transactions on Software Engineering and Methodology*, vol. 33, no. 5, pp. 1–28, 2024.
- [71] M. Zhang, H. Liu, C. Chen, Y. Liu, and S. Bai, "Consistent or not? an investigation of using pull request template in github," *Information and Software Technology*, vol. 144, p. 106797, 2022.
- [72] GitHub, "Generate new repositories with repository templates," <https://github.blog/developer-skills/github/generate-new-repositories-with-repository-templates/>.