Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

June 2025, Volume 3, Issue 1

Analyzing Answer-Type Preferences Among Expertise Shapes on Stack Overflow

Omid Mohammadi Kia✉, ORCID iD: 0000-0003-2225-4450
Shahid Beheshti Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran,
o_mohammadikia@sbu.ac.ir
Mahmood Neshati, ORCID iD: 0000-0002-3953-4857
Shahid Beheshti Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran,
m_neshati@sbu.ac.ir

Abstract
Developer expertise is a critical component of community-based question-answering platforms like Stack Overflow. However, expertise is not monolithic. Developers exhibit different "shapes" of expertise, such as deep specialists (I-shaped) or broad generalists with one specialty (T-shaped). While prior research has examined developer expertise and answer quality independently, how different expertise structures influence preferences for specific answer characteristics remains insufficiently understood. This paper investigates the relationship between a developer's expertise shape and their preference for specific answer characteristics.
We present a large-scale empirical study of over 48,000 Stack Overflow users, classifying them into I-shaped, T-shaped, Pi-shaped, and Comb-shaped profiles based on the distribution of tag-level reputation, following established expertise-shape modeling approaches. We then analyze the types of answers these user profiles tend to upvote and accept as solutions, focusing on characteristics such as answer length, inclusion of code snippets, use of images, and citation of external references.
Using separate analyses for community-level (upvotes) and task-resolution-level (accepted answers) preference signals, our findings reveal distinct and systematic differences across expertise shapes. I-shaped specialists favor technically deep, code-heavy answers, while T-shaped and Comb-shaped experts show a preference for more summarized, conceptual answers that include diagrams or references. These patterns are consistent across robustness checks and sensitivity analyses.
The results highlight that answer usefulness is user-dependent rather than universal, and they can help improve expertise-aware answer recommendation systems and foster more effective knowledge sharing on collaborative platforms.

Keywords: Stack Overflow, Expertise Shapes, T-Shaped, Answer Quality, Empirical Software Engineering, Recommender Systems.

1. INTRODUCTION

Community-based question-answering (CQA) platforms, with Stack Overflow (SO) as the premier example, have become indispensable tools in modern software development. Developers rely on these platforms for daily problem-solving, learning new technologies, and sharing knowledge within a global community. The continued success of such platforms depends not only on answer availability, but on the relevance of answers to the individual developer's needs.

The effectiveness of an answer, however, is not objective; it is highly dependent on the user seeking help. An answer considered perfect by one developer may be insufficient or overly simplistic for another. This subjectivity is closely tied to differences in how developers structure and apply their knowledge. To explore this, our work adopts the concept of *expertise shapes* [1] (e.g., I, T, Pi, Comb) from organizational theory and software engineering research, applying it to software developers. These shapes describe the balance between a developer's deep, specialized knowledge (the vertical bar of an 'I' or 'T') and their broad, generalist knowledge (the horizontal bar of a 'T' or 'Comb').

While prior work has extensively studied answer quality on SO [2], [3] and others have modeled developer expertise using signals such as reputation, activity, and contribution patterns [4], [5], [6], [7] research has also explored the classification of question types on the platform [8]. However, existing studies typically treat expertise as a scalar notion (e.g., novice vs. expert) or analyze answer quality independently of the consumer's knowledge structure. As a result, little is known about how different expertise profiles influence what developers actually consider a useful or preferable answer.

This paper addresses this gap by explicitly linking expertise structure to answer preference. We investigate how a developer's knowledge shape correlates with the characteristics of the answers they upvote and accept, viewing these actions as complementary signals of preference. This leads to our primary research questions:

- **RQ1:** What are the distinct answer characteristics (length, code, images, references) preferred by developers with different expertise shapes (I, T, Pi, Comb)?
- **RQ2:** How significant are these preferences, and what do they imply about the information-seeking goals of each expert type?

Submit Date: 2025-11-06
Revise Date: 2026-02-13
Accept Date: 2025-04-28
✉ Corresponding author

To answer these questions, we present an empirical study of over 48,000 active Stack Overflow users from the June 2024 data dump [9]. We use heuristic, reputation-based criteria—grounded in prior expertise-shape modeling work—to classify users into four expertise shapes based on how their knowledge is distributed across technology tags. We then statistically analyze the characteristics of over 1.2 million answers, considering upvotes and accepted answers in separate analyses to distinguish between community-level appreciation and task-resolution preferences.

Our key finding is that these preferences are systematic and significantly different across expertise shapes. For instance, I-shaped specialists strongly favor answers with substantial code blocks, while T-shaped and Comb-shaped experts show a statistically significant preference for summarized answers that include diagrams and external references. These differences persist across robustness checks and sensitivity analyses.

These findings have direct implications for the design of CQA platforms. By accounting for user expertise structure rather than assuming a universal notion of answer quality, a platform could personalize its answer-ranking mechanisms to surface the type of answer most likely to be useful to a specific user, thereby improving knowledge-sharing efficiency.

The remainder of this paper is organized as follows. Section 2 reviews related work on expertise modeling and answer quality. Section 3 describes our methodology for data collection, user classification, and analysis. Section 4 presents our findings. Section 5 discusses the implications and interpretations of these results. Finally, Section 6 concludes the paper and outlines avenues for future work.

2. BACKGROUND AND RELATED WORK

Modeling Expertise: From I-Shaped to Comb-Shaped

The concept of "T-shaped" individuals was first popularized in the 1990s to describe an ideal employee who possesses deep expertise in one area (the vertical bar) and a broad ability to collaborate across other disciplines (the horizontal bar) [1] .This model was later adapted to software engineering [4], [5] where a T-shaped developer might have deep expertise in back-end development while maintaining working knowledge of front-end and database systems.

Subsequent work has generalized this notion to capture a broader range of expertise distributions. An I-shaped individual represents a deep specialist with limited breadth, whereas a Pi-shaped individual possesses two distinct areas of deep expertise. Comb-shaped individuals, in contrast, exhibit multiple areas of specialization, forming a profile resembling a comb [6]. Rostami et al. [1] formalized these expertise shapes for team formation and talent identification in agile software development by modeling expertise as distributions over knowledge areas.

Existing work on expertise shapes has primarily focused on organizational outcomes such as team composition, collaboration efficiency, and talent discovery, rather than on information-seeking or content-evaluation behavior. In this study, we adopt these well-established expertise-shape definitions and apply them to Stack Overflow users, using them as an analytical lens to examine how developers with different knowledge structures interact with and evaluate technical content.

Answer Quality on Stack Overflow

A large body of research has sought to define and predict answer quality on SO. Studies have identified numerous features correlated with high-scoring or accepted answers. These include answer length, readability, sentiment, the inclusion of code snippets, the presence of external links, and even the author's reputation. For example, some studies found that longer, more comprehensive answers are more likely to be accepted[10], [11]. Others noted the critical importance of code snippets for solving technical problems.

A substantial body of research has investigated answer quality on Stack Overflow [2], [3], [12], [13], [14], [15] These studies typically operationalize quality using signals such as upvotes, accepted answers, or predictive models trained on these outcomes. Numerous answer-level features have been identified as correlates of high-quality or highly rated answers, including answer length, readability, sentiment, inclusion of code snippets, presence of external links, and author reputation [10], [16], [17].

Prior findings, however, are not always consistent. For example, some studies report that longer, more detailed answers are more likely to be accepted [15], [17], whereas others emphasize the importance of concise responses depending on context. Similarly, while many works highlight the critical role of code snippets in solving programming problems [18], [19], [20], [21], [22], [23], [24], [25] the relative importance of these features varies across studies and tasks.

Importantly, most answer-quality studies evaluate answers from an aggregate or platform-wide perspective, implicitly assuming that the same answer characteristics are universally desirable across users and situations.

Positioning and Research Gap

While the literature on developer expertise modeling and answer quality on Stack Overflow is extensive, these two research streams have largely evolved in parallel. Expertise studies focus on *who* the developers are and how their skills are distributed, whereas answer-quality studies focus on *what* constitutes a good answer based on observable features and community feedback. A limited number of studies have examined voting behavior in relation to user reputation or experience [10], [16], [17] but they do not account for the internal structure of a developer's expertise. As a result, existing work cannot explain why certain answer characteristics are valued differently by different groups of experienced users, nor why findings about "good" answers sometimes diverge across studies.

This paper contributes to the literature by explicitly connecting expertise structure to answer preference. Rather than redefining answer quality or proposing new expertise models, we investigate how established expertise shapes (I, T, Pi, Comb) are associated with systematic differences in the types of answers developers upvote or accept. By doing so, we provide a unifying perspective that helps contextualize prior answer-quality findings through the lens of user heterogeneity.

these two streams of research—developer expertise shapes and answer quality metrics—are well-established, no prior work has explicitly connected them. It is known what features constitute a good answer in general, and it is known that different developer expertise shapes exist. However, it remains unknown how an expert's knowledge structure (e.g., I-shaped vs. T-shaped) influences their preference for specific answer features (e.g., code vs. diagrams). This paper aims to bridge that gap by empirically linking user profiles to their demonstrated content preferences.

3. METHODOLOGY

This section details our step-by-step process for data collection, classification of users and answers, and statistical analysis. To support transparency and reproducibility, the data analysis scripts, classification heuristics, and anonymized datasets used in this study are publicly available on github[1].

Data Collection

We used the official Stack Overflow data dump from June 2024 [9]. To ensure a focus on active and established technologies while avoiding sparsity issues, we restricted our analysis to the top 100 most frequently used Stack Overflow tags (e.g., *python*, *java*, *javascript*, *c#*, *react*). This tag-selection strategy is consistent with prior large-scale empirical studies on Stack Overflow[26], [27].

We filtered our dataset to include only users who had posted, answered, or voted on content within these 100 tags and had a reputation of at least 100, indicating a baseline level of sustained engagement. This threshold excludes sporadic or inactive accounts while retaining a broad spectrum of contributors. Our final dataset consisted of 48,215 unique users and 1,210,458 answers that these users had either upvoted or accepted.

Defining Expertise Shape Proxies (Operationalization)

We developed a set of heuristic rules to classify users into four expertise shapes based on the distribution of their reputation across different tags. Tag-level reputation is used as a proxy for topic-specific expertise, a common practice in prior Stack Overflow research. The operational definitions of T-shaped and Comb-shaped expertise follow established expertise-shape modeling approaches in software engineering research.

The classification heuristics are defined as follows:

- **I-Shaped (Specialist):** A user with 90% or more of their total tag-based reputation concentrated in a single tag.
- **T-Shaped (Generalist-Specialist):** A user with 50-70% of their reputation in one primary tag, and the remaining 30-50% distributed across 10 or more other tags.
- **Pi-Shaped (Dual-Specialist):** A user with two primary tags, each containing 30-45% of their reputation, with the two tags together accounting for at least 70% of their total.
- **Comb-Shaped (Multi-Specialist):** A user with 3 to 5 primary tags, each accounting for 15-25% of their reputation, with no single tag exceeding 30%.

Users who did not fit these heuristics (e.g., new users with sparse reputation) were excluded from the analysis. This classification resulted in 12,109 I-Shaped, 15,044 T-Shaped, 6,322 Pi-Shaped, and 8,011 Comb-Shaped users.

Let Ube the set of users and $T$the set of Stack Overflow tags. For a given user $u \in U$, let $R(u,t)$denote the reputation earned by user $u$under tag $t \in T$. The total tag-based reputation of user $u$is defined as:

$$R_{total}(u) = \sum_{t \in T} R(u,t)$$

We define the normalized reputation share of user $u$for tag $t$as:

$$p(u,t) = \frac{R(u,t)}{R_{total}(u)}$$

Let $T_u = \{t \in T \mid R(u,t) > 0\}$denote the set of tags in which user $u$has non-zero reputation. Tags are ordered such that:

$$p(u,t_1) \geq p(u,t_2) \geq \cdots \geq p(u,t_{|T_u|})$$

Using these definitions, expertise shapes are operationalized as follows:

- I-Shaped (Specialist):
$$\exists t_1 \in T_u \text{ such that } p(u,t_1) \geq 0.90$$
- T-Shaped (Generalist–Specialist):
$$0.50 \leq p(u,t_1) \leq 0.70 \ \wedge \ | T_u | \geq 10$$
- Pi-Shaped (Dual Specialist):
$$0.30 \leq p(u,t_1), p(u,t_2) \leq 0.45 \ \wedge \ p(u,t_1) + p(u,t_2) \geq 0.70$$
- Comb-Shaped (Multi-Specialist):
$$\exists \{t_1, \ldots, t_k\} \subseteq T_u,\ 3 \leq k \leq 5,\ 0.15 \leq p(u,t_i) \leq 0.25,\ \max_i \ p(u,t_i) \leq 0.30$$

Users who do not satisfy any of the above conditions are excluded from the analysis.

To assess robustness, we conducted a sensitivity analysis by varying all percentage thresholds by ±10%. The direction and relative magnitude of the results remained stable across all configurations, indicating that the findings are not artifacts of specific cutoff choices.

Defining Answer Characteristic Proxies

We analyzed the raw Markdown and HTML of each of the 1.2 million answers to classify them based on four key characteristics:

- **Summarized vs. Long:** Answers with fewer than 150 words were labeled *Summarized*, while answers with more than 400 words were labeled *Long*. These thresholds approximately correspond to the lower and upper quartiles of the answer-length distribution and capture extreme brevity and verbosity**.** Answers in the intermediate range (150–400

[1] https://github.com/omkia/StackoverflowUserTypes

words) were not assigned to either category. As a robustness check, we additionally modeled word count as a continuous variable.

- **With Code:** The answer contained one or more <code> blocks with at least 5 lines of code.
- **With Image:** The answer contained one or more <img> tags (typically used for screenshots or diagrams).
- **With Reference:** The answer contained one or more external <a> hyperlinks, excluding links to other Stack Overflow pages.

Analysis Method

To To answer our research questions, we analyzed the relationship between a user's expertise shape and the characteristics of the answers they preferred. We treat preference using two complementary signals: upvotes and accepted answers. Upvotes reflect community-level appreciation, while accepted answers represent explicit solution selection by the question asker.

To avoid conflating these signals, we conduct separate analyses for upvotes and accepted answers. For each expertise shape, we built logistic regression models where the dependent variable is a binary indicator of whether an answer was preferred by a user of that shape. The independent variables are the binary flags for the answer characteristics (Summarized, Long, With Code, With Image, With Reference).

Logistic regression was selected for its interpretability and suitability for binary outcomes. We report regression coefficients, odds ratios, and statistical significance, and we evaluate model quality using standard diagnostics, including AIC and ROC/AUC. Multicollinearity was assessed using Variance Inflation Factors (VIF), with all values below commonly accepted thresholds.

RESULTS

Our analysis reveals systematic and statistically significant differences in answer preferences across expertise shapes. The results of the logistic regression models are summarized in Table 1, which reports the regression coefficients (B) for each answer characteristic. A positive coefficient indicates that the presence of a feature increases the likelihood of an answer being preferred by users of a given expertise shape, whereas a negative coefficient indicates aversion.

Table 1 summarizes the results of the upvote-based models. Corresponding models using accepted answers as the dependent variable are reported separately in the Appendix and show consistent directional effects.

Table 1: Logistic Regression Coefficients for Answer Preference by Expertise Shape

Answer Feature	I-Shaped (Specialist)	T-Shaped (Generalist)	Pi-Shaped (Dual)	Comb-Shaped (Multi)
Answer Length (Long)	+0.21*	-0.18*	+0.10	-0.22**
Answer Length (Summ.)	-0.15	+0.24**	+0.05	+0.31***
Includes Code	+0.72***	+0.29**	+0.45***	+0.18*
Includes Image	-0.04	+0.38***	+0.19*	+0.44***
Includes Reference	+0.10	+0.41***	+0.28**	+0.39***

*Significant at $p < .05$, **Significant at $p < .01$, ***Significant at p < .001*

This table presents our core findings, which we elaborate on below in the context of our research questions.

RQ1: What are the distinct answer characteristics preferred by each shape?

I-Shaped experts exhibit the strongest preference for answers that include code snippets (B = 0.72, $p < .001$), with an effect size substantially larger than that observed for any other group**.** This indicates a clear prioritization of concrete, implementation-level content. I-shaped users also show a mild preference for long answers, while images and external references do not significantly influence their preferences.

T-Shaped experts display a markedly different preference profile. They show strong positive associations with answers that include references (B = 0.41, $p < .001$) and images (B = 0.38, $p < .001$), along with a significant preference for summarized answers (B = 0.24, $p < .01$). Conversely, long answers are negatively associated with preference (B = -0.18, $p < .05$), suggesting an aversion to overly detailed responses.

Pi-Shaped experts demonstrate a hybrid pattern. Similar to I-shaped specialists, they show a strong preference for code-containing answers (B = 0.45, $p < .001$). At the same time, they also exhibit significant positive associations with references (B = 0.28, $p < .01$) and images (B = 0.19, $p < .05$), aligning them partially with T-shaped users.

Comb-Shaped experts show the strongest overall preference for summarized and visually supported answers. Images (B = 0.44, $p < .001$), summarized answers (B = 0.31, $p < .001$), and references (B = 0.39, $p < .001$) are all strongly associated with preference. This group also exhibits the strongest aversion to long answers (B = -0.22, $p < .01$). To reduce redundancy and facilitate comparison, Table 2 summarizes the preferred and disfavored answer characteristics across all four expertise shapes.

Table 2. Comparative Summary of Answer Preferences by Expertise Shape

Expertise Shape	Strongly Preferred Answer Characteristics	Neutral / Mixed	Disfavored Characteristics
I-Shaped (Specialist)	Code-heavy answers; technically detailed solutions	Long answers	Images; external references
T-Shaped (Generalist-Specialist)	Summarized answers; images; external references	Code snippets	Long answers
Pi-Shaped (Dual Specialist)	Code snippets; external references	Images; answer length	None strongly disfavored
Comb-Shaped (Multi-Specialist)	Summarized answers; images; external references	Code snippets	Long answers

RQ2: How significant are these preferences and their implications?
The preferences are highly significant ($p < .001$ in many cases), indicating that these are not random chance but clear behavioral patterns. The implications are clear: the "best" answer is perceived differently by each group. I-Shaped users seek deep, technical solutions, while T-Shaped and Comb-Shaped users seek high-level concepts, diagrams, and links for further learning.
The observed effects are statistically significant across most answer characteristics and expertise shapes, with many coefficients reaching $p < .001$**.** This indicates that the differences are unlikely to be due to random variation and instead reflect systematic behavioral patterns.
Taken together, the results suggest that answer usefulness is not perceived uniformly across developers. I-shaped experts predominantly seek technically detailed, code-centric solutions, whereas T-shaped and Comb-shaped experts favor concise, high-level explanations enriched with diagrams and external references. Pi-shaped experts occupy an intermediate position, combining preferences for concrete implementations with contextual and integrative information.
Importantly, these patterns persist in models restricted to accepted answers only, reinforcing the interpretation that they reflect genuine problem-solving preferences rather than general popularity effects.

4- Accepted-Answer-Only Analysis
To disentangle general community preference from definitive problem resolution, we conducted an additional logistic regression analysis using **accepted answers only** as the dependent variable.
Overall, the accepted-answer-only models **reinforce and strengthen** the patterns observed in the upvote-based analysis. As summarized in **Table 3**, I-shaped and Pi-shaped experts continue to show a strong preference for answers containing code, while T-shaped and Comb-shaped experts remain significantly associated with summarized answers, images, and external references.
Specifically, **I-shaped experts exhibit an even stronger preference for code-heavy answers** in the accepted-only model ($B = 0.85$, $p < .001$), indicating that executable solutions are particularly critical when selecting a final answer. A similar pattern is observed for **Pi-shaped experts** ($B = 0.58$, $p < .001$), reinforcing their reliance on concrete implementations when resolving interdisciplinary problems.
In contrast, **T-shaped and Comb-shaped experts show heightened associations with summarized answers, images, and references**. For Comb-shaped users, summarized answers ($B = 0.37$, $p < .001$) and images ($B = 0.49$, $p < .001$) exhibit the strongest effects, while long answers remain negatively associated with acceptance. These findings suggest that high-level clarity and visual explanation play a central role in solution selection for broadly skilled developers.
Effect sizes are generally larger than those observed in the upvote-based models, indicating that these answer characteristics are especially influential when users select a **single accepted solution**, rather than expressing general approval through voting.

Table 3. Accepted-Answer-Only Logistic Regression Coefficients by Expertise Shape

Answer Feature	**I-Shaped**	**T-Shaped**	**Pi-Shaped**	**Comb-Shaped**
Answer Length (Long)	+0.28*	−0.24*	+0.12	−0.31**
Answer Length (Summ.)	−0.19	+0.29**	+0.07	+0.37***
Includes Code	+0.85***	+0.34**	+0.58***	+0.22*
Includes Image	−0.06	+0.42***	+0.24*	+0.49***
Includes Reference	+0.14	+0.46***	+0.33**	+0.44***

* $p < .05$, ** $p < .01$, *** $p < .001$

5 -DISCUSSION
The results indicate that developer expertise shape is systematically associated with differences in information-seeking behavior on Stack Overflow**.** Rather than suggesting causal mechanisms, we interpret these patterns as *consistent behavioral tendencies* that reflect how developers with different knowledge structures evaluate and consume technical content.
I-Shaped (The Specialist): I-shaped experts show a strong and consistent preference for code-heavy answers, along with a mild tolerance for longer responses. This pattern aligns with their specialization-oriented expertise structure, where developers operate primarily within a narrow technical domain. For such users, concrete implementations appear to provide the most direct and efficient path to problem resolution, whereas high-level explanations, images, or external references offer limited additional value. Importantly, this interpretation reflects observed preference behavior rather than assumptions about intent or task complexity.
T-Shaped (The Generalist-Specialist): T-shaped experts exhibit preferences for summarized answers enriched with images and external references. This combination suggests that these users benefit from concise conceptual explanations supported by visual structure and authoritative sources. Such preferences are consistent with an expertise profile that balances depth in a primary domain with breadth across adjacent areas, where rapid contextual understanding is often required. The aversion to long answers further indicates that excessive detail may hinder, rather than help, efficient information uptake for this group**.**
Pi-Shaped (The Integrator): Pi-shaped experts demonstrate a hybrid preference pattern, showing strong associations with both code inclusion and contextual features such as references and images**.** This suggests that users with dual specializations may seek answers that both demonstrate practical implementation and explain how solutions operate across multiple technical domains. These findings support the view that Pi-shaped expertise involves integrative problem-solving rather than isolated specialization.

Comb-Shaped (The Architect): Comb-shaped experts show the strongest preference for summarized answers and visual content, along with a clear aversion to long, detail-heavy responses**.** Images and references appear particularly valuable to this group, likely because they facilitate high-level reasoning across multiple domains. This pattern is consistent with expertise structures that emphasize coordination, abstraction, and system-level understanding rather than low-level implementation details**.**

.

Implications for Stack Overflow

These findings challenge the implicit assumption that answer quality is universal and user-independent**.** Instead, they suggest that usefulness is shaped by the expertise structure of the consumer.

From a practical perspective, Stack Overflow and other CQA platforms could benefit from incorporating expertise-aware personalization mechanisms**.** For example, inferred expertise shapes could be used to re-rank existing answers—prioritizing code-centric solutions for specialists and concise, visually supported explanations for breadth-oriented users. Such personalization could be implemented without altering content creation practices, potentially improving efficiency and user satisfaction.

Threats to Validity

This study is subject to several limitations that should be considered when interpreting the results.

First, the classification of expertise shapes relies on heuristic, reputation-based proxies, which may not fully capture a developer's true knowledge structure. Although sensitivity analysis indicates robustness to threshold variation, this remains a construct validity threat. Future work could incorporate qualitative validation through surveys or interviews, or enrich expertise modeling with additional behavioral signals.

Second, the analysis is restricted to the top 100 Stack Overflow tags, which may bias results toward mainstream technologies and limit generalizability to niche or emerging domains. This constitutes a threat to external validity.

Third, we interpret upvotes and accepted answers as signals of preference, yet voting behavior on Stack Overflow is influenced by multiple factors, including visibility, author reputation, and social dynamics. While separating upvote-based and accepted-answer-based analyses mitigates this concern, preference inference remains an approximation rather than a direct measure of intent.

Finally, the observational nature of the study precludes causal conclusions**.** The reported associations should therefore be interpreted as descriptive patterns rather than explanatory mechanisms.

6 - CONCLUSION AND FUTURE WORK

In this paper, we investigated the relationship between developer expertise shapes (I-shaped, T-shaped, Pi-shaped, and Comb-shaped) and their answer preferences on Stack Overflow. Using a large-scale empirical dataset, we operationalized expertise shapes through reputation-based heuristics and analyzed the preferences of over 48,000 users across 1.2 million answers.

The primary contribution of this study is empirical evidence that developer expertise is better understood as a structured concept rather than a single, uniform category**.** Our results show that I-shaped specialists tend to prefer technically detailed, code-centric answers, whereas T-shaped and Comb-shaped experts exhibit preferences for more summarized, conceptual, and visually supported responses that include external references. Pi-shaped experts demonstrate a hybrid preference profile, combining elements of both specialization-oriented and breadth-oriented information needs. These findings indicate that a developer's knowledge structure is systematically associated with how they evaluate and select answers, rather than assuming a universal notion of answer usefulness.

While the proposed expertise-shape classification relies on heuristic proxies, sensitivity analysis suggests that the observed patterns are robust to reasonable threshold variations. Nonetheless, future work should explore richer and more flexible modeling approaches, such as interpretable machine learning techniques, to infer expertise structures using additional behavioral and temporal signals beyond reputation alone.

An important direction for future research is qualitative validation**.** Surveys or interviews with developers could help assess how well the inferred expertise shapes align with self-perceived knowledge structures and professional roles. Additionally, the framework introduced in this study could be extended beyond answer evaluation to examine whether developers with different expertise shapes ask different types of questions or exhibit distinct problem-framing behaviors.

Overall, this work highlights the importance of accounting for user heterogeneity in community-based question–answering platforms. By recognizing differences in expertise structure, future systems can move toward more effective, personalized, and context-aware knowledge support for software developers.

Appendix A. Logistic Regression Model Details

To improve the transparency and interpretability of the statistical analysis, this appendix reports detailed results for all logistic regression models used in the study. For each expertise shape, we report regression coefficients (B), odds ratios (OR), 95% confidence intervals (CI), and model diagnostics.

Odds ratios are provided to facilitate interpretation of effect sizes, indicating how the presence of a given answer characteristic affects the likelihood that an answer is preferred by users with a specific expertise shape. Confidence intervals quantify the uncertainty associated with these estimates.

Model quality was evaluated using the Akaike Information Criterion (AIC) and the Area Under the Receiver Operating Characteristic Curve (AUC). Across all models, AUC values ranged from **0.64 to 0.71**, indicating acceptable discriminative performance for behavioral data. AIC values suggest adequate model fit without evidence of overfitting.

To assess multicollinearity among predictors, Variance Inflation Factors (VIF) were computed for all independent variables. All VIF values were below **2.0**, well under commonly accepted thresholds, indicating no multicollinearity concerns.

Tables **A1–A4** present the full regression results for **I-shaped, T-shaped, Pi-shaped, and Comb-shaped** users, respectively.

Table A1. Logistic Regression Results for I-Shaped Experts

Predictor	B	OR	95% CI (OR)	p-value	VIF
Long Answer	0.21	1.23	[1.04, 1.45]	0.021	1.32
Summarized Answer	-0.15	0.86	[0.72, 1.03]	0.108	1.28
Includes Code	0.72	2.05	[1.78, 2.36]	<.001	1.41
Includes Image	-0.04	0.96	[0.81, 1.14]	0.612	1.19
Includes Reference	0.10	1.11	[0.94, 1.32]	0.207	1.22

Model diagnostics:
AIC = 412,305 | AUC = 0.71

Table A2. Logistic Regression Results for T-Shaped Experts

Predictor	B	OR	95% CI (OR)	p-value	VIF
Long Answer	-0.18	0.84	[0.72, 0.98]	0.032	1.36
Summarized Answer	0.24	1.27	[1.09, 1.48]	0.004	1.29
Includes Code	0.29	1.34	[1.15, 1.57]	0.001	1.44
Includes Image	0.38	1.46	[1.28, 1.66]	<.001	1.33
Includes Reference	0.41	1.51	[1.32, 1.73]	<.001	1.38

Model diagnostics:
AIC = 398,742 | AUC = 0.67

Table A3. Logistic Regression Results for Pi-Shaped Experts

Predictor	B	OR	95% CI (OR)	p-value	VIF
Long Answer	0.10	1.11	[0.93, 1.32]	0.261	1.27
Summarized Answer	0.05	1.05	[0.89, 1.24]	0.544	1.22
Includes Code	0.45	1.57	[1.32, 1.87]	<.001	1.46
Includes Image	0.19	1.21	[1.01, 1.45]	0.038	1.31
Includes Reference	0.28	1.32	[1.12, 1.56]	0.002	1.35

Model diagnostics:
AIC = 287,910 | AUC = 0.65

Table A4. Logistic Regression Results for Comb-Shaped Experts

Predictor	B	OR	95% CI (OR)	p-value	VIF
Long Answer	-0.22	0.80	[0.69, 0.93]	0.006	1.39
Summarized Answer	0.31	1.36	[1.18, 1.58]	<.001	1.26
Includes Code	0.18	1.20	[1.02, 1.42]	0.029	1.41
Includes Image	0.44	1.55	[1.34, 1.78]	<.001	1.34
Includes Reference	0.39	1.48	[1.28, 1.72]	<.001	1.37

Model diagnostics:
AIC = 361,088 | AUC = 0.64

7 - REFERENCES

[1] P. Rostami and M. Neshati, “T-shaped grouping: Expert finding models to agile software teams retrieval,” *Expert Syst Appl*, vol. 118, pp. 231–245, Mar. 2019, doi: 10.1016/j.eswa.2018.10.015.

[2] M. Neshati, “On early detection of high voted Q&A on Stack Overflow,” *Inf Process Manag*, vol. 53, no. 4, pp. 780–798, Jul. 2017, doi: 10.1016/j.ipm.2017.02.005.

[3] C. Shah and J. Pomerantz, “Evaluating and predicting answer quality in community QA,” *SIGIR 2010 Proceedings - 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 411–418, 2010, doi: 10.1145/1835449.1835518.

[4] S. S. Gharebagh, P. Rostami, and M. Neshati, “T-shaped mining: A novel approach to talent finding for agile software teams,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10772 LNCS, pp. 411–423, 2018, doi: 10.1007/978-3-319-76941-7_31.

[5] P. Rostami and A. Shakery, “A deep learning-based expert finding method to retrieve agile software teams from CQAs,” *Inf Process Manag*, vol. 60, no. 2, Mar. 2023, doi: 10.1016/j.ipm.2022.103144.

[6] P. Rostami and M. Neshati, “Intern retrieval from community question answering websites: A new variation of expert finding problem,” *Expert Syst Appl*, vol. 181, Nov. 2021, doi: 10.1016/j.eswa.2021.115044.

[7] M. Neshati, H. Beigy, and D. Hiemstra, “Expert group formation using facility location analysis,” *Inf Process Manag*, vol. 50, no. 2, pp. 361–383, Mar. 2014, doi: 10.1016/j.ipm.2013.10.001.

[8] O. M. Kia, M. Neshati, and M. S. Alamdari, “Open-domain question classification and completion in conversational information search,” *2020 11th International Conference on Information and Knowledge Technology, IKT 2020*, pp. 98–101, Dec. 2020, doi: 10.1109/IKT51791.2020.9345613.

[9] “stackexchange directory listing.” Accessed: Nov. 03, 2025. [Online]. Available: https://archive.org/download/stackexchange

[10] B. Li, T. Jin, M. R. Lyu, I. King, and B. Mak, “Analyzing and predicting question quality in community question answering services,” *WWW’12 - Proceedings of the 21st Annual Conference on World Wide Web Companion*, pp. 775–782, 2012, doi: 10.1145/2187980.2188200.

[11] C. Meng, N. Arabzadeh, A. Askari, M. Aliannejadi, and M. De Rijke, “Query Performance Prediction Using Relevance Judgments Generated by Large Language Models,” *ACM Trans Inf Syst*, vol. 43, no. 4, Jul. 2025, doi: 10.1145/3736402/ASSET/22299729-9445-484B-BC37-E4CDAE747F2F/ASSETS/IMAGES/LARGE/TOIS-2024-0348-F08.JPG.

[12] A. Y. K. Chua and S. Banerjee, “So fast so good: An analysis of answer quality and answer speed in community Question-answering sites,” *Journal of the American Society for Information Science and Technology*, vol. 64, no. 10, pp. 2058–2068, Oct. 2013, doi: 10.1002/ASI.22902.

[13] Y. Lin and H. Shen, “SmartQ: A Question and Answer System for Supplying High-Quality and Trustworthy Answers,” *IEEE Trans Big Data*, vol. 4, no. 4, pp. 600–613, Aug. 2017, doi: 10.1109/TBDATA.2017.2735442.

[14] A. Merchant, D. Shah, G. S. Bhatia, A. Ghosh, and P. Kumaraguru, “Signals matter: Understanding popularity and impact of users on stack overflow,” *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, pp. 3086–3092, May 2019, doi: 10.1145/3308558.3313583.

[15] R. Selleras, *Predictive model: using text mining for determining factors leading to high-scoring answers in stack overflow*. 2020. Accessed: Nov. 03, 2025. [Online]. Available: https://search.proquest.com/openview/7813a3cc073b1ad7cafe1373d4605ad8/1?pq-origsite=gscholar&cbl=18750&diss=y

[16] M. R. Tavakoli, A. Heydarnoori, and M. Ghafari, “Improving the quality of code snippets in stack Overflow,” *Proceedings of the ACM Symposium on Applied Computing*, vol. 04-08-April-2016, pp. 1492–1497, Apr. 2016, doi: 10.1145/2851613.2851789.

[17] L. X. Zhao, L. Zhang, and J. Jiang, “Hot question prediction in Stack Overflow,” *IET Software*, vol. 15, no. 1, pp. 90–106, Feb. 2021, doi: 10.1049/SFW2.12013.

[18] S. Baltes and S. Diehl, “Usage and attribution of Stack Overflow code snippets in GitHub projects,” *Empir Softw Eng*, vol. 24, no. 3, pp. 1259–1295, Jun. 2019, doi: 10.1007/S10664-018-9650-5.

[19] Y. Wu, S. Wang, C. P. Bezemer, and K. Inoue, “How do developers utilize source code from stack overflow?,” *Empir Softw Eng*, vol. 24, no. 2, pp. 637–673, Apr. 2019, doi: 10.1007/S10664-018-9634-5.

[20] M. Tavakoli, A. Heydarnoori, M. G.-P. of the 31st Annual, and undefined 2016, “Improving the quality of code snippets in stack overflow,” *dl.acm.org*, Accessed: Nov. 03, 2025. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2851613.2851789

[21] C. Ragkhitwetsagul, J. Krinke, M. Paixao, G. Bianco, and R. Oliveto, “Toxic code snippets on stack overflow,” *ieeexplore.ieee.org*, no. Y, p. 1, 2018, doi: 10.1109/TSE.2019.2900307.

[22] Z. Gao, X. Xia, D. Lo, J. Grundy, X. Zhang, and Z. Xing, “I Know What You Are Searching for: Code Snippet Recommendation from Stack Overflow Posts,” *ACM Transactions on Software Engineering and Methodology*, vol. 32, no. 3, Apr. 2023, doi: 10.1145/3550150.

[23] M. Ahmad, M. C.-2019 I. 16th International, and undefined 2019, “Impact of stack overflow code snippets on software cohesion: A preliminary study,” *ieeexplore.ieee.org*, Accessed: Nov. 03, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8816758/

[24] D. Yang, A. Hussain, and C. V. Lopes, “From query to usable code: An analysis of Stack Overflow code snippets,” *Proceedings - 13th Working Conference on Mining Software Repositories, MSR 2016*, pp. 391–401, May 2016, doi: 10.1145/2901739.2901767.

[25] S. Baltes, S. D.-E. S. Engineering, and undefined 2019, “Usage and attribution of Stack Overflow code snippets in GitHub projects,” *Springer*, 2018, Accessed: Nov. 03, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s10664-018-9650-5

Omid Mohammadi Kia, ORCID iD: 0000-0003-2225-4450
Shahid Beheshti Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, o_mohammadikia@sbu.ac.ir
Mahmood Neshati, ORCID iD: 0000-0002-3953-4857
Shahid Beheshti Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, m_neshati@sbu.ac.ir