

Voices from the Community: A Multi-Dimensional Analysis of Tech Career Discourse on Reddit

Matheus S. Azevedo

Departamento de Computação e Sistemas,
Universidade Federal de Ouro Preto
João Monlevade, MG, Brazil
matheus.azevedo@aluno.ufop.edu.br

Carlos H. G. Ferreira

Departamento de Computação, Universidade Federal
de Ouro Preto
Ouro Preto, MG, Brazil
chgferreira@ufop.edu.br

Geovana S. Oliveira

Departamento de Computação e Sistemas,
Universidade Federal de Ouro Preto
João Monlevade, MG, Brazil
geovana.so@aluno.ufop.edu.br

Helen C. S. C. Lima

Departamento de Computação e Sistemas,
Universidade Federal de Ouro Preto
João Monlevade, MG, Brazil
helen@ufop.edu.br

ABSTRACT

The technology labor market has experienced major changes in recent years, driven by the consolidation of remote work, the adoption of generative AI, and successive corporate layoffs. As these changes reshaped career paths and employment prospects, professionals increasingly turn to online communities to exchange experiences and discuss their concerns. In this work, we present a longitudinal, multidimensional analysis of career discussions across three technology-focused Reddit communities (subreddits) covering 2018–2025. We collected Reddit discussions and applied a two-stage BERTopic pipeline. The first stage identified relevant publication domains, yielding a curated dataset of 143,933 publications and 2,015,841 associated comments, while the second modeled the thematic structure of their discussions. The analysis was complemented by sentiment analysis, emotion classification, temporal analysis, and engagement metrics. Our findings reveal nine major thematic areas spanning technical skills, education, job seeking, workplace conditions, and broader labor-market disruptions. Discussions of market instability, including layoffs, unemployment, and the disruptive effects of generative AI, generated high engagement, while social and labor-rights discussions produced the most sustained conversational activity. In contrast, technical and educational discussions were more strongly associated with knowledge sharing and positive emotions. We further observe a marked increase in career-related discussions after 2022, coinciding with major shifts in the technology labor market and the emergence of generative AI as a salient concern in career-related discourse.

KEYWORDS

social media analysis, topic modeling, emotion detection, technology market, labor, remote work

1 INTRODUCTION

The technology market has undergone profound transformations in recent years, reshaping work models, valued skills, organizational strategies, and career perspectives. These changes have been driven by different factors, including the consolidation of remote work after the COVID-19 pandemic [2, 10], rapid adoption of generative artificial intelligence tools [30], and successive waves of layoffs in large companies [7]. Together, these phenomena have intensified the need for organizational and professional adaptation, increasing concerns about stability, reskilling, and career development.

Professionals increasingly turn to online communities to understand these transformations, share experiences, and seek career guidance. Exposure to others' experiences broadens knowledge of career trajectories and influences expectations, aspirations, and perceptions about one's own professional future, although it may also encourage social comparison [12]. These communities also provide spaces for mentorship and peer support, where users exchange advice, experiences, and emotional support when facing professional challenges [6, 32].

Among these platforms, Reddit has become a major repository of user-generated discussions about technology careers. Its topic-oriented communities (subreddits) enable continuous conversations around shared interests, while the volume of content makes manual analysis impractical. Consequently, Natural Language Processing (NLP) techniques have been widely used to extract structure from large text collections, identifying topics, sentiment patterns, and temporal dynamics in online discourse.

Recent studies have applied computational approaches to online discussions about the technology labor market, examining perceptions of artificial intelligence and its impact on employment [1, 14, 25], experiences of specific professional groups [26, 27, 29], job satisfaction and retention [20, 22], and trends observed in job advertisements and recruitment practices [3, 9, 35]. Together, these studies highlight the potential of computational analysis to capture different dimensions of the technology job market, although they typically focus on isolated aspects of the problem.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2026.

© 2026 SBC – Brazilian Computing Society.
ISSN 2966-2753

The literature therefore provides a fragmented view of technology career discussions. Most studies examine specific phenomena through limited analytical dimensions, leaving unclear how these discussions are jointly structured in online communities, evolve over time, and relate to users' emotions, sentiments, and engagement.

Given this gap, this work investigates discussions about technology careers published in three Reddit communities between 2018 and 2025. Specifically, we address the following research questions:

- **RQ1:** How do thematic discussion structures, combined with platform engagement metrics, characterize user interaction dynamics and perceptions of the technology job market?
- **RQ2:** How do core career-related themes evolve over time, and what are the prevalent emotional states and sentiment polarities expressed by users across these themes?

To answer these questions, we constructed a longitudinal dataset comprising publications and comments from three prominent technology-career Reddit communities, known as subreddits, spanning 2018–2025. The analysis pipeline integrates BERTopic-based topic modeling, sentiment and emotion analysis, temporal analysis, and engagement metrics to characterize how discussions evolve and how users react to career-related topics.

Our work makes three main contributions: (i) a two-stage topic modeling and affective pipeline linking high-level publication domains to fine-grained comment interactions; (ii) a large-scale longitudinal characterization (2018–2025) coupling career-topic dynamics with engagement, sentiment polarity, and discrete emotions; and (iii) actionable insights to support online community administrators and labor market platforms in identifying emerging workforce concerns and structuring targeted peer-support mechanisms.

2 RELATED WORK

Research on technology-related discussions in online communities has explored how professionals talk about artificial intelligence, working conditions, and career experiences across different digital platforms. These studies suggest that such discussions are strongly influenced by context, including platform design, timing of events, and community structure. At the same time, they differ widely in scope and methodology, which makes it difficult to build a unified view of how technology career discourse unfolds online.

Perception of AI and the labor market. Recent social-media studies have examined how AI-related discussions reflect changing perceptions of work and employment. Reddit-based research reports persistent concerns about job displacement and regulation, alongside increasing attention to reskilling after 2022 [1]. Similarly, online discussions have shifted from AI replacement toward more collaborative views of human-AI work, often accompanied by more positive sentiment [14], while discussions of ChatGPT remain largely positive despite ethical concerns [25]. Beyond online discourse, studies of technology professionals report concerns about job security,

potential displacement, and adaptation to generative AI [23, 33], while related work connects these changes to evolving software development practices and professional roles [15]. However, social-media research has largely focused on specific AI technologies or AI-related changes in work, rather than the broader and evolving concerns surrounding technology careers.

Professional platforms and labor discourse. Work on professional and semi-professional platforms shows how labor-related discussions are shaped by both platform design and economic context. Analyses of Blind, an anonymous professional community, for example, indicate that conversations often mirror industry events such as layoffs and shifts to remote work [8]. Studies based on job reviews and postings highlight organizational culture, flexibility, and management as key factors in job satisfaction [20, 22]. At a broader level, research also points to the difficulty of collective organization in the technology sector, even in a context of high perceived prestige [31]. These studies provide useful insights into workplace perceptions, but they tend to rely on structured platforms or aggregated data, rather than open community discussions.

Reddit-based studies of professional communities. On Reddit, previous work shows that the characteristics of each community strongly influence how people express opinions and emotions. Discussions about internships, for instance, tend to be positive and focused on entry into the job market and professional growth [27]. In contrast, gender-focused communities often show more negative emotional tones, linked to experiences of bias, isolation, and career stagnation [29]. This variation suggests that subreddit context plays an important role in shaping discourse, but most studies still focus on specific communities or themes, rather than broader patterns of career-related discussions.

Computational approaches in related domains. Outside the labor market literature, topic modeling and sentiment analysis have been widely used to study large-scale online discussions. Examples include analyses of misinformation and health communication in messaging platforms [13] and studies of narrative structures in online entertainment contexts [28]. These works show that combining topic-level and sentiment analysis is effective for understanding complex discourse, but they are usually applied in separate domains rather than specifically to technology career discussions.

Despite these advances, the literature is still fragmented. Most studies focus on specific phenomena or rely on a single analytical perspective, often with limited methodological scope. As a result, there is still no unified understanding of how discussions about technology careers are structured, how they evolve over time, and how emotional responses relate to engagement patterns.

To address this gap, we conduct a longitudinal analysis of discussions about technology careers across three Reddit communities between 2018 and 2025. We combine BERTopic-based topic modeling applied to publications and comments with sentiment analysis, temporal analysis, and engagement metrics to build a multi-dimensional view of these discussions.

Overall, our work provides a broad picture of how technology professionals discuss and respond to recent changes in the labor market.

3 METHODOLOGY

This section describes our methodological pipeline, covering data collection, curation, topic modeling, and sentiment and emotion classification.

3.1 Data Collection and Curation

We utilized Reddit submissions and comments dumps curated by *Watchful1* [34] and distributed through Academic Torrents platform¹. At the time of data collection, the archive provided historical Reddit data from the 40,000 most active subreddits extracted from Pushshift [4], spanning 2005–2025. During the publication process, however, the archive was discontinued, and the dump became unavailable from its original source. Similar Reddit datasets derived from *Watchful1*'s dumps have supported studies of online social discourse, hate speech, political polarization, and health-related communities [5, 19, 21], demonstrating its relevance for large-scale social behavior research. To ensure transparency, the complete source code used in this study is available in a GitHub repository². The original dataset is not redistributed, and exact replication may be limited because the original source is no longer publicly accessible. Since all analyses were conducted exclusively at the aggregate level, no individual users are identified or reported in this study.

We targeted three popular subreddits focused on technology careers and labor dynamics: *r/cscareerquestions*, *r/ITCareerQuestions*, and *r/womenintech*. Although the historical data spanned 2011–2025, we restricted the analysis to 2018–2025 due to the substantial growth in publication volume during this period, ensuring a dense representation of contemporary market dynamics.

Text preprocessing removed null and duplicate records and automated placeholders (e.g., "[deleted]", "[removed]", and "[deleted by user]"). Comments shorter than 25 characters were also discarded to remove low-effort interactions and spam. The resulting foundational dataset comprised 390,321 publications and 6,160,644 comments.

3.2 Two-Stage Topic Modeling Pipeline

To identify discussion themes, we developed a two-stage BERTopic pipeline capturing publication-level domains and fine-grained comment topics.

Stage 1: Domains Discovery (Publications). We first applied BERTopic to publications titles to identify overarching core domains in the target subreddits. Texts were embedded using the pre-trained *bge-small-en-v1.5* sentence transformer³.

Dimensionality reduction was performed via UMAP, configured with 5 neighbors, 10 dimensions, minimum distance of 0.01, and cosine distance as the metric. Density-based clustering was executed applying HDBSCAN with a minimum cluster size of 150 and 5 minimum samples. The vocabulary was restricted to the 10,000 most frequent terms, excluding tokens appearing in fewer than 0.1% or more than 90% of the document corpus. From this step, we selected 24 highly relevant tech-career topics, accounting for 36.87% of total publications. Filtering out off-topic and generalized meta-discussions yielded **143,933** publications and **2,015,841** comments. This reduction also mitigated memory and processing constraints, making subsequent dense comment modeling feasible on standard hardware.

The 24 topics were aggregated into five broad domains using *gemini-3.5-flash* model with structured system instructions (see Section A.1). To reduce reliance on automated synthesis, the resulting domains were manually reviewed by two authors. The review considered the most representative keywords identified across the representations provided to the LLM, as well as the top representative documents associated with each constituent topic. Domain assignments or labels were revised when necessary to improve semantic and thematic coherence, with disagreements resolved through consensus. The resulting domains and constituent topics are shown in Figure 1 and define the high-level structure of the publication corpus used in subsequent analyses.

Stage 2: Discussion Structure Discovery (Comments). We applied BERTopic a second time, specifically targeting the comments linked to the previously filtered publications. In contrast to Stage 1, topic modeling was employed to identify fine-grained discussion subjects emerging from user interactions within the selected domains. To accommodate their larger volume, length, and lexical diversity, UMAP used 20 neighbors and 5 dimensions, while HDBSCAN used a minimum cluster size of 500 and 5 minimum samples. The vocabulary was expanded to 20,000 terms. Due to the resulting granularity, we once again leveraged the *gemini3.5-flash*, using a dedicated structured prompt designed for fine-grained thematic synthesis (see Section A.2) to aggregate the identified topics into cohesive Macro-Topics (MTs). The same manual review procedure used in Stage 1 was applied to the resulting MTs. This consolidation enables structured analysis of large-scale discussion trends, quantified in Section 4.

3.3 Sentiment and Emotion Classification

To capture users' perceptions and emotional expressions across the identified discussions, we employed two transformer-based sequence classification models. For sentiment analysis, we used the *twitter-roberta-base-sentiment-latest* model⁴, which is fine-tuned for short social media texts and classifies comments into three polarity classes: *Positive*, *Neutral*, and *Negative*. To capture more fine-grained affective

¹Academic Torrents is a distributed repository designed for researchers to share large scientific datasets via BitTorrent.

²<https://github.com/matsa0/voices-from-tech-community-multidimensional-analysis-on-reddit>

³<https://huggingface.co/BAAI/bge-small-en-v1.5>

⁴<https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>

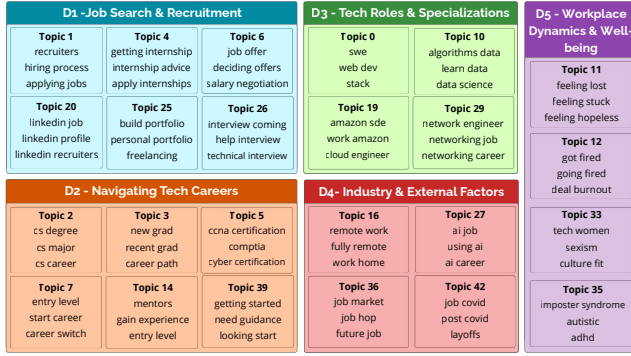


Figure 1: Discovered domains in publications titles of the corpus

states, we employed the `emotion-roberta-large` model⁵, which predicts the six basic emotions proposed by Ekman (*Anger*, *Disgust*, *Fear*, *Joy*, *Sadness*, and *Surprise*), in addition to the *Neutral* class. Together, these models enable a complementary analysis of sentiment polarity and emotional responses across different technology career-related discussions.

4 RESULTS & DISCUSSIONS

This section presents findings from the large-scale analysis of Reddit comments. We first examine comment-level topics and engagement patterns, followed by sentiment distributions and emotional dynamics.

4.1 Macro-Topics Characterization & Engagement metrics

To address **RQ1**, we analyze comments within the discovered domains. While publication titles provide concise discussion contexts, comments capture richer interactions, including technical guidance, personal experiences, and community perspectives. As detailed in Section 3.2, BERTopic extracted **31** comment-level topics, accounting for 77.02% of the analyzed corpus. To facilitate higher-level interpretation, these topics were subsequently consolidated into nine macro-topics (MT1–MT9), representing major thematic dimensions of discussions surrounding the contemporary technology labor market.

Table 1 presents the topics' structural alignment, five representative keywords, comment counts, and percentage distribution within the curated dataset. Table 2 reports engagement metrics used to examine user behavior across macro-topics, while Figure 2 shows the proportional distribution of domains within each macro-topic.

To assess statistical differences across macro-topics, Kruskal-Wallis tests were systematically applied to all continuous and binary engagement metrics in Table 2. For metrics with significant variation ($p < 0.001$), Dunn post-hoc comparisons with Holm's correction [18] were conducted. These results support the characterizations below by distinguishing behavioral differences from random variation:

⁵<https://huggingface.co/j-hartmann/emotion-english-roberta-large>

MT1 — Software Engineering & Tech Stacks: As the largest macro-topic, accounting for more than 23% of all comments, MT1 represents the technical foundations of technology labor-market discussions across the analyzed subreddits. Rather than focusing exclusively on programming languages, tools, and frameworks, MT1 reveals a strong prevalence of discussions related to Software Development, DevOps practices, Quality Assurance, cloud platforms, and infrastructure management. This thematic baseline is structurally anchored in the distribution of two major domains visualized in Figure 2: *Tech Role & Specializations* (32%) and *Navigating Tech Careers* (31%). As detailed in Table 2, the intersection of these domains stands out as the primary driver of recurrent user participation across the entire corpus, yielding the statistically highest rate of comments per user ($C/U = 3.53, p < 0.001$), alongside a highly elevated density of comments per publication ($C/P = 4.69, p < 0.001$) and a stable reply rate ($RR = 51.98\%$). Collectively, these metrics demonstrate that career-oriented technical architectural debates, form a highly active, persistent background conversation that continuously engages the community.

MT2 — Interviewing & Entry-Level Market: Primarily associated with *Job Search & Recruitment* (34%) and *Navigating Tech Careers* (25%), MT2 captures discussions surrounding technical interviews, coding assessments, internships, bootcamps, and challenges faced by early-career professionals entering the technology workforce. MT2 involves 73,021 users and exhibits a significantly higher reply rate than MT1 ($RR = 54.07\%, p < 0.001$). Post-hoc tests also show significantly higher median publication scores ($Med.Pub. = 31, p < 0.001$) and mean comment scores ($Avg.Com. = 7.50, p < 0.001$) than MT1. These results reinforce that employability, concerns and early-career hurdles are broadly shared across the community, transforming this discussions into interactive exchange of experiences, advice, and strategies for overcoming barriers to entry in an increasingly competitive labor market.

MT3 — Academic Education: Despite being composed of only two elemental topics and leaning predominantly toward the *Navigating Tech Careers* domain (61%), MT3 aggregates

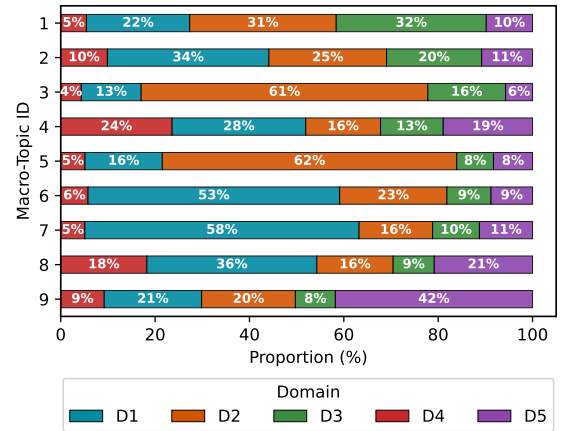


Figure 2: Domains distribution per Macro-Topic

Table 1: Thematic aggregation of comment-level latent topics into Macro-Topics.

ID	Name	(ID) Keywords	# Comments (%)
MT1	Software Engineering & Tech Stacks	(00) devops, msp, sysadmin, software development, qa	239,255 (15.40)
		(09) python, java, sql, languages, javascript	52,374 (3.37)
		(10) aws, linux, azure, github, kubernetes	48,956 (3.15)
		(22) react, typescript, angular, android, frameworks	26,436 (1.70)
		(51) spring boot, java, frameworks, learn java, kotlin	1,266 (0.08)
MT2	Interviewing & Entry-Level Market	(03) scrum, agile, chatgpt, interview, grind leetcode	116,121 (7.47)
		(07) bootcamp, interns, new grads, unpaid, graduation	84,290 (5.42)
		(28) level jobs, entry, require experience, experience years, qualifications	2,305 (0.14)
MT3	Academic Education	(01) phd, wgu, gpa, cs degree, programs	178,730 (11.51)
		(18) semester, went school, time study, teacher, community college	28,416 (1.83)
MT4	Macroeconomics & Industry Trends	(08) layoffs, startup, companies, outsourcing, recession	74,689 (4.81)
		(12) unemployment, pip, getting fired, severance, layoff	41,772 (2.69)
		(20) automation, think ai, humans, ai replacing, productivity	26,836 (1.72)
		(27) rates, recession, inflation, cuts, trump	14,473 (0.93)
		(46) saturated, different field, people field, competitive field, good field	2,966 (0.19)
MT5	Specialized Sectors & Credentials	(04) ccna, certs, cybersecurity, network security, comptia	92,466 (5.95)
		(17) healthcare, pandemic, health insurance, medicine, nursing	31,596 (2.03)
		(26) security clearance, contractor, government jobs, air force, private sector	15,544 (1.00)
MT6	Job Application Strategies	(05) profile, referrals, glassdoor, background checks, recruiting	89,091 (5.73)
		(14) start applying, apply jobs, looking new, rejections, hunting	36,800 (2.37)
		(32) build portfolio, good portfolio, site, showcase, projects	9,957 (0.64)
MT7	Compensation & Career Advancement	(02) negotiate, compensation, promoted, new offer, job hop	120,337 (7.75)
		(35) mentors, coach, volunteering, career advice, help grow	7,986 (0.51)
		(43) senior engineers, good engineers, lot engineers, junior, talented	4,557 (0.29)
		(50) soft skills, tech skills, social skills, hard skills, skills important	2,305 (0.14)
MT8	Workplace Culture & Well-being	(06) pto, overtime, burnout, mental health, worklife balance	86,826 (5.59)
		(21) remote work, want remote, fully remote, hybrid, rto	26,560 (1.71)
		(23) adhd, imposter syndrome, therapy, autistic, insecurity	21,781 (1.40)
MT9	Social Issues & Labor Rights	(16) diversity, gender, women, sexism, misogyny	33,934 (2.18)
		(25) ageism, boomers, age discrimination, older people, millennials	16,669 (1.07)
		(39) unions, capitalism, rights, corporations, tech workers	6,258 (0.40)

a substantial volume of discussions (49,959 publications and 207,146 comments) centered on computer science degrees, universities, postgraduate education, and academic dynamics. Compared to most macro-topics, educational discussions receive lower publication and comment scores however, MT3 exhibits the absolute shortest median response time across the analyzed corpus ($Med.Hrs. = 4.10$), coupled with a significantly elevated density of comments per publication ($C/P = 4.15, p < 0.001$). Collectively, these metrics indicate that academic threads function primarily as agile, crowd-sourced support and information-sharing spaces rather than polarized or crisis-driven debates, where community members actively exchange guidance regarding educational decisions and early career preparation.

Table 2: Engagement metrics distribution per Macro-Topic.

ID	#Pub.	#Com.	#Usr.	#Rep.	RR(%)	Med. Pub.	Avg. Com.	C/U	C/P	Med. Hrs.
MT1	78,583	368,287	104,275	191,426	51.98	16	5.60	3.53	4.69	4.88
MT2	53,764	213,834	73,021	115,630	54.07	31	7.50	2.93	3.98	4.83
MT3	49,959	207,146	69,714	103,932	50.17	6	4.66	2.97	4.15	4.10
MT4	31,658	160,736	58,032	96,322	59.93	131	10.27	2.77	5.08	5.27
MT5	37,650	139,606	51,750	71,437	51.17	10	4.38	2.70	3.71	4.53
MT6	40,703	135,848	54,727	70,219	51.69	23	6.84	2.48	3.34	4.51
MT7	37,284	135,185	50,689	62,047	45.90	24	7.96	2.67	3.63	4.24
MT8	29,662	135,167	55,615	66,178	48.96	60	8.15	2.43	4.56	5.31
MT9	13,313	56,861	28,156	36,408	64.03	122	7.60	2.02	4.27	6.58

Note: #Pub. = Publications; #Com. = Comments; #Usr. = Unique Users; #Rep. = Comment Replies; RR(%) = Reply Rate (% of nested comment-to-comment replies); Med. Pub. = Median Post Score (Upvotes – Downvotes); Avg. Com. = Mean Comment Score; C/U = Comments per User; C/P = Comments per Publication; Med. Hrs. = Median Response Time in Hours.

MT4 — Macroeconomics & Industry Trends: MT4 captures structural events and transformations affecting the technology labor market, including mass layoffs, unemployment, recession, inflation, market saturation, and the disruptive influence of generative AI. These economic and technological concerns intersect with job searching (*Job Search & Recruitment* – 28%), workplace volatility (*Workplace Dynamics & Well-being* – 19%), and broader industry shifts (*Industry & External Factors* – 24%). In terms of ecosystem dynamics, MT4 exhibits a high reply rate of ($RR = 59.93\%, p < 0.001$), the highest median publication score ($Med.Pub. = 131, p < 0.001$) and the highest mean comment score ($Avg.Com. = 10.27$). Furthermore, it drives the highest density of comments per publication ($C/P = 5.08$). Post-hoc tests indicate that its C/P distribution is statistically comparable to MT2 ($p = 0.072$), suggesting similar interaction patterns between industry-wide disruptions and entry-level market challenges. These findings demonstrate that macroeconomic instability paired with high-impact technological shifts such as AI-transition, act as a powerful catalyst for collective anxiety, highlighted by intense voting behavior and prolonged discussions within the comment sections

MT5 — Specialized Sectors & Credentials: Heavily dominated by the *Navigating Tech Careers* domain (62%), MT5 focuses primarily on professional credentials and technical specializations within the cybersecurity and network security fields (e.g., CCNA, CompTIA), alongside career transitions into sectors such as government and healthcare. Regarding engagement, MT5 shows lower visibility than most macro-topics, with the lowest mean comment score ($Avg.Com. = 4.38, p <$

0.001). However, its comments-per-user rate ($C/U = 2.70$) is statistically comparable to MT4 ($p = 0.4552$), indicating that specialized certifications and career niches sustain consistent, focused exchanges of career guidance despite lower overall visibility.

MT6 — Job Application Strategies: Strongly associated with *Job Search & Recruitment* domain (53%), MT6 focuses on practical aspects of the job acquisition process, including job applications, referrals, recruitment platforms and portfolio construction. Like MT2, it portrays entry into the technology workforce as a multi-stage process involving networking, skill demonstration, and persistent applications. Despite lower comment volume, MT6 presents a higher number of publications than MT4 and MT5. Furthermore, its low median response time ($Med.Hrs. = 4.51, p < 0.001$) indicates that users actively and quickly exchanges job-search advices. Its reply rate ($RR = 51.69$) is statistically comparable to MT1 ($p = 0.0693$), suggesting similar interaction depth between practical job-search discussions and core software engineering topics. These findings demonstrate that the concern in keeping a good profile and portfolio for applications, as well as referrals, indicates that community members often discuss ways to improve visibility and competitiveness in the labor market.

MT7 — Compensation & Career Advancement: MT7 covers salary negotiation, promotions, career growth, professional development and the importance of different types of skills for career success. Primarily associated with *Job Search & Recruitment* domain (58%), it also shows that job offers, compensation, and job hopping are major concerns among technology workers. This finding is consistent with previous studies showing that changing jobs voluntarily is very common among software developers [24]. MT7 exhibits the lowest reply rate ($RR = 45.90, p < 0.001$) among all macro-topics, yet a fast median response time ($Med.Hrs. = 4.24$). The post-hoc pairwise comparison reveal that its response speed is statistically comparable to MT3 ($p = 0.7873$), implying that as well as academic discussions, users respond quickly to salary, negotiation and job discussions. Although MT4 presents the highest average number of comments per post ($C/P = 5.08$), MT7 does not differ significantly from it ($C/P = 3.63, p = 0.6042$). Combined with its low reply rate, this suggests that compensation discussions tend to receive multiple direct responses rather than extended back-and-forth exchanges.

MT8 — Workplace Culture & Well-being: This macro-topic encompasses discussions on workplace conditions and their implications for employee well-being. The identified topics cover work arrangements, including remote, hybrid and return-to-office policies, as well as quality-of-life factors such as paid time off (PTO), overtime and work-life balance. In parallel, users frequently discuss mental health and neurodiversity, including burnout, attention-deficit/hyperactivity disorder (ADHD), autism and imposter syndrome. The co-existence of these themes suggests that users often frame workplace conditions and psychological well-being as interconnected aspects of technology careers, a relationship that has

also been reported in previous studies on burnout, imposter syndrome, ADHD and autism among technology professionals [11, 16, 17]. Although MT8 contains fewer publications than previously discussed macro-topics, it generates a disproportional number of comments (135,167), involving more users than MT5, MT6 and MT7, ($C/P = 4.56, p < 0.001$) comments per publication, an elevated median publication score ($Med.Pub. = 60, p < 0.001$) and a high mean comment score ($Avg.Com. = 8.15$), indicating that individual posts frequently attract substantial community interaction. The predominance of discussions surrounding burnout, work-life balance, and remote work suggests that quality-of-life considerations have become increasingly important aspects of career decision-making.

MT9 — Social Issues & Labor Rights: MT9 encompasses discussions on diversity, gender inequality, age discrimination, labor rights, unionization, and broader social issues affecting technology professionals. The identified topics reveal concerns about sexism and misogyny experienced by women in the industry, age-related discrimination involving both younger and older professionals, and collective actions such as labor unions aimed at protecting workers' rights. Although this is the smallest macro-topic in the corpus, comprising only 13,313 publication and 56,861 comments, it exhibits the highest reply rate among all macro-topics ($RR = 64.03, p < 0.001$), indicating that these discussions frequently evolve into extended conversational threads. In addition, MT9 records the second-highest median publication score ($Med.Pub. = 122, p < 0.001$) and ($C/P = 4.27$) comments per publication. Pairwise comparisons revealed no statistically significant difference between MT9 and MT8 ($p = 0.5657$), suggesting that discussions in both macro-topics receive comparable levels of community appreciation. Similarly, no significant difference was observed in comments per publication between MT9 and MT6 ($p = 0.6042$), indicating comparable levels of discussion activity. These findings reinforce that, despite its relatively low prevalence, MT9 consistently attracts substantial engagement from the community. Altogether, these engagement indicators suggest that, while social issues are discussed by a comparatively smaller community (28,156 users), they consistently stimulate sustained interaction and collective debate, reflecting the sensitivity and relevance of these topics within the technology workforce.

4.2 Temporal Dynamics Across Macro-Topics

To address part of **RQ2**, Figure 3 presents the temporal evolution of the discovered topics and their corresponding macro-topics. Since Macro-Topics represent broader discussion domains, we focus primarily on their temporal dynamics, while individual topics are used to explain specific temporal variation whenever necessary. Each heatmap row was standardized using z -score normalization⁶, which highlights the relative prominence of each topic across time while controlling for differences in absolute comment volume. Positive values indicate periods of above-average discussion intensity

⁶The z -score is defined as $z = \frac{x - \mu}{\sigma}$.

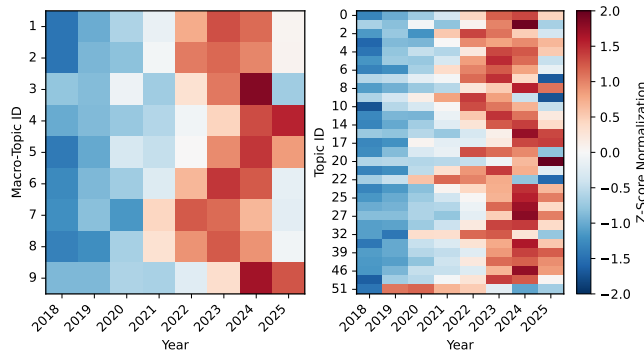


Figure 3: Temporal distribution of Macro-Topics and Topics.

within a given topic, whereas negative values denote periods of comparatively lower activity.

A first observation is that most macro-topics exhibit above-average discussion intensity from 2022 onward, indicating a simultaneous increase in the relative prominence of multiple technology labor market subjects. Instead of just independent evolutions, several macro-topics follow remarkably similar temporal trajectories. In particular, MT1, MT2, MT6, MT7 and MT8 present highly synchronized patterns with sustained prominence between 2022 and 2024. Although these macro-topics addresses different aspects of professional life, they collectively encompass the technology career life-cycle, ranging from technical skills acquisition and hiring processes to career progression and workplace conditions. Their similar temporal behavior suggests that discussions surrounding technology careers responded simultaneously to broader transformations in the labor market.

The topic-level heatmap further reveals that these aggregate trends are driven by different underlying discussion subjects. For example, the prominence of MT3 in 2024 is primarily explained by Topic 01, which concentrates discussions about computer science degrees, graduate programs and PhD studies, whereas Topic 18 remains comparatively stable throughout the analyzed period. Likewise, the temporal increase observed in MT5 from 2023, is largely associated with Topic 04, which focuses on professional certifications in security field such as CCNA and CompTIA, rather than with healthcare or government-related occupations represented by the remaining topics. These examples illustrate how the topic-level representation complements the macro-topic analysis revealing which specific discussion subjects are responsible for broader temporal movements.

In contrast, MT4 and MT9 exhibit a distinct temporal behavior, reaching their highest prominence during 2024 and 2025. This increase is mainly driven by discussions on layoffs, unemployment, artificial intelligence, market saturation, diversity, age discrimination and labor rights (Topics 08, 12, 16, 20, 25 and 27). Unlike the previous group, which reflects career development and professional progression, these macro-topics capture broader structural transformations affecting the technology labor market.

Finally, comparing both heatmaps reveals an important distinction. While the macro-topics display relatively smooth and coherent temporal trajectories, the underlying topics exhibit considerably greater heterogeneity, with several topics presenting isolated peaks (Topics 01, 16, 20, 27 and 46) or gradual declines (Topics 02, 04, 09, 10, 21, 22, 32 and 51). Taken together, these patterns suggest a gradual shift in the community's attention. Earlier discussions were predominantly centered on entering and progressing in technology careers, whereas more recent conversations increasingly focus on understanding how external economic conditions, emerging technologies and social issues are reshaping the technology labor market itself.

4.3 Sentiment polarization and Emotions per Macro-Topic

This section address **RQ2** by investigating how sentiment polarization and emotions responses characterize the tech-market domains explored in this study. Figure 4 exhibits the distribution of *Positive*, *Negative* and *Neutral* sentiments across the macro-topics and Figure 5 depicts the distribution of six emotion categories (*Anger*, *Disgust*, *Fear*, *Joy*, *Sadness* and *Surprise*). Since *Neutral* predictions account for 60.16% of the corpus, they were excluded from the emotion analysis to facilitate the interpretation of emotion patterns.

To determine whether emotional distributions differed significantly across macro-topics, we applied a Chi-square test of independence. The test revealed a highly significant association between macro-topics and emotions ($\chi^2 = 37048.09$, $df = 40$, $p < 0.001$). Although the overall association measured by Cramér's V was relatively small ($V = 0.1095$), adjusted standardized residuals enable local analysis of emotions are over- or under-represented within each macro-topic. Given the large sample size of the corpus, only residuals with absolute values greater than three ($|z| > 3$) were interpreted as meaningful deviations.

Considering sentiment polarization together with the emotional distributions, MT1 (Software Engineering & Tech Stacks), MT3 (Academic Education), and MT5 (Specialized Sectors & Credentials) exhibit remarkably similar sentiment distribution patterns. Adjusted residuals indicate that these discussions are strongly characterized by the over-representation of *joy*, particularly in MT1 (58.44), MT3 (38.22) and MT5 (28.00) accompanied by significant levels of *surprise* in MT1 (22.98) and MT3 (16.28). Conversely, *anger* is substantially under-represented in MT1 (-46.87), MT3 (-45.42), and MT5 (-35.81). Together, these patterns suggest that discussions centered on technical skills, university education, professional certifications, and specialized sectors are primarily associated with knowledge sharing, technical guidance, and career achievements. This interpretation is illustrated by comments such as "*After months of applying, I finally landed a well-paid IT position at a great company. I*

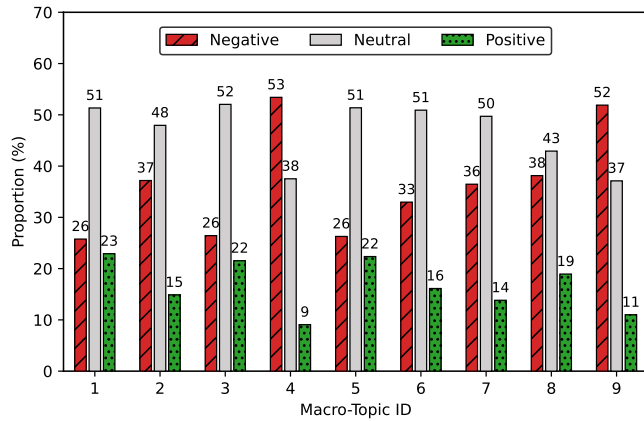


Figure 4: Sentiments proportion across Macro-Topics

almost gave up after weeks of negative thoughts, but perseverance paid off."⁷ and "I'm about to finish my degree and I'm genuinely happy to have finally broken into the tech industry."

Moving the analysis to macro-topics MT2 (Interviewing & Entry-Level Market), MT6 (Job Application Strategies), and MT7 (Compensation & Career Advancement), the sentiment polarity reveal greater imbalance between positive and negative dimensions compared to the previous group, shifting decisively toward the negative spectrum. Adjusted standardized residual analysis indicates that this thematic cluster is marked by a systemic predominance of *anger* across MT2 (6.04), MT6 (3.72), and, most heavily, MT7 (29.50). Furthermore, *fear* stands out prominently in MT2 (16.95) and MT7 (13.02), capturing the insecurity of early-career professionals alongside corporate instability. Meanwhile, MT6 is strongly driven by *surprise* (25.40), a reflection of broken expectations regarding recruitment feedback and selection processes. Conversely, *joy* is severely suppressed across all three domains (MT2: -17.49; MT6: -9.75; MT7: -17.77), showing that these spaces serve as places to vent and share frustrating career experiences, as seen in MT2: "My biggest concern is being a graduate looking for work 6+ months after graduation and not finding it [...]".

Although MT8 (Workplace Culture & Well-being) presents a similarly polarized sentiment profile, its emotional composition differs substantially from the previous group. Here, the under-representation of *anger* (-23.20) and *surprise* (-54.65) coexists with strong over-representations of both *sadness* (62.93) and *joy* (28.81). This emotion duality reflects the heterogeneous nature of discussions within MT8, which encompass both positive experiences related to workplace flexibility and work-life balance, as well as conversations about burnout, neurodiversity, and mental health challenges. This contrasting landscape are delineated with user comments such as: "Work-Life balance is amazing! They don't expect you to work overtime. I have a lot of time for me." and "OP

⁷For privacy reasons, all comments reported here were paraphrased into English by the authors.

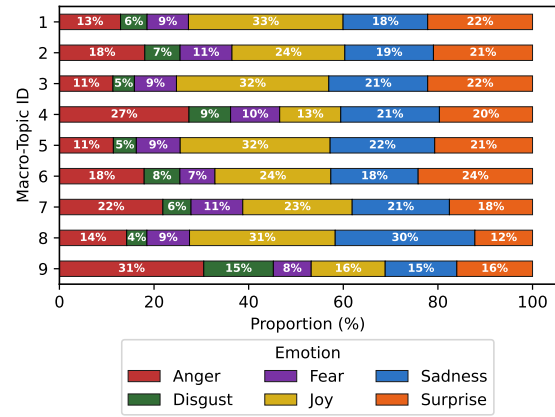


Figure 5: Emotion proportion across Macro-Topics

is struggling with a level of burnout that a lot of you already faced regarding being in this field."⁸

The most negative emotional profiles are observed in MT4 (Macroeconomics & Industry Trends) and MT9 (Social Issues & Labor Rights), which exhibit the highest proportions of negative sentiment among all macro-topics, exceeding even their neutral proportions. Adjusted residuals further reveal a pronounced under-representation of *joy*, reaching the largest negative deviations of the study in MT4 (-90.48) and MT9 (-42.60). Instead, these discussions are dominated by *anger* (MT4: 79.01; MT9: 61.28), accompanied by strong over-representations of *disgust* (MT4: 27.18; MT9: 58.04). Although both macro-topics share this negative emotional profile, their underlying discussions differ considerably. In MT4, the predominance of *anger* and *fear* (10.03) appears to be associated with structural concerns regarding the technology labor market including large-scale layoffs, market saturation, and uncertainty surrounding AI-driven automation, as illustrated by comments such as: "I've been following AI for years, and the pace of recent advances is genuinely scary. [...] Even if only a fraction of software jobs disappear, the impact on the economy and future recruitment could be severe. Personally, I wouldn't want to be starting a career in tech today.". By contrast, MT9 exhibits the strongest over-representation of *disgust* in the entire corpus (58.04), suggesting that discussions extend beyond economic dissatisfaction to ethical concerns surrounding workplace culture, including harassment, age discrimination, sexism, and other forms of unfair treatment. This interpretation can be visualized in comments like: "It's upsetting to realize that I'm not the only one who has experienced harassment and discrimination in the workplace. [...] I've seen favoritism, sexism, and inappropriate behavior that made the work environment feel deeply toxic and unacceptable.".

⁸OP (Original Poster) is an acronym used in Reddit comments to refer to the original author of the post.

5 CONCLUSION & FUTURE WORK

This work presented a longitudinal, multi-dimensional analysis of technology career discussions across three Reddit communities, r/cscareerquestions, r/ITCareerQuestions, and r/womenintech, from 2018 to 2025. Using BERTopic-based topic modeling together with sentiment and emotion analysis, temporal analysis, and engagement metrics, we analyzed 143,933 publications and 2,015,842 comments to examine how technology professionals discuss careers amid labor market changes.

Our analysis identified nine macro-topics spanning technical skills, education, certifications, job searching, macroeconomic instability, and social issues. These topics exhibited distinct engagement patterns, with macroeconomic and social issues consistently attracting the highest participation, suggesting that engagement is closely tied to the personal and emotional relevance of each topic.

The temporal analysis revealed a marked increase in career-related discussions after 2022 across most macro-topics, indicating a collective response to recent labor market transformations. Affective analyses also distinguished discussion themes: learning and professional development were predominantly associated with positive emotions, whereas job searching, compensation, economic uncertainty, and social issues showed more negative profiles, particularly anger, fear, and disgust.

Beyond characterizing discourse, our pipeline offers practical applications for online platforms. First, tracking multi-dimensional patterns can inform labor intelligence systems, helping institutions monitor credential demands and workforce concerns. Second, for platform designers and moderators, detecting high-reply threads coupled with intense negative emotions (e.g., anger or disgust in crisis topics) can power automated mechanisms that route distressed users toward targeted peer-support or mental health resources.

This study presents limitations that motivate future work. First, regarding platform and demographic bias, our data is exclusively drawn from Reddit, whose user base predominantly reflects Western, English-speaking corporate cultures and anonymous interaction norms; future work should validate these dynamics across platforms such as YouTube, LinkedIn and X and in localized contexts, including the Brazilian technology ecosystem. Second, our Ekman-based emotion model captures sentence-level semantics but may overlook finer psycholinguistic constructs; complementary resources such as LIWC or dimensional emotion frameworks could provide additional perspectives. Finally, although our approach captured generative AI within broader macroeconomic discussions (MT4), future work could examine emerging narratives of developer deskilling and LLM-driven job displacement.

REFERENCES

- [1] I. Baholli, G. Tigno, F. Hidri, et al. 2026. Tracking Topic Drift in AI and Workforce Narratives on Reddit: A Longitudinal Visualization from 2010–2024. *International Journal of Innovative Technology and Interdisciplinary Sciences* (2026).
- [2] Y. Bal and N. E. Bulgur. 2023. Remote work: A paradigm shift in the modern workplace and its impact on the workforce. In *Enhancing employee engagement and productivity in the post-pandemic multigenerational workforce*. IGI Global.
- [3] B. A. P. Barata, A. F. Justino, A. F. L. Junior, and othes. 2023. What about data science? An analysis of the market based on Job posts. In *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*.
- [4] J. Baumgartner, S. Zannettou, B. Keegan, et al. 2020. The pushshift reddit dataset. In *Proc. Int. AAAI Conf. on Web and Social Media*.
- [5] P. Bento, Y. A. Aquino, A. Buzelin, et al. 2025. Mapping Hate and Extremism: How Brazilian Reddit Communities Spread Toxic Discourse.
- [6] J. Carpenter, C. McDade, and S. Childers. 2018. Advice seeking and giving in the Reddit r/Teachers online space. In *Society for Information Technology & Teacher Education International Conference*. AACE.
- [7] M. B. Çetin and S. Gündüz. 2024. From Boom to Bust: Unravelling the Global Tech Layoffs Phenomenon. In *Economic Uncertainty in the Post-Pandemic Era*. Routledge.
- [8] T. Chaudhary, P. Malhotra, R. Mamidi, et al. 2023. Blind Leading the Blind: A Social-Media Analysis of the Tech Industry. In *Proc. of the 20th International Conference on Natural Language Processing (ICON)*.
- [9] T. Chumwatana and A. Hpone. 2025. Bridging the IT skill gap with industry demands: An AI-driven text mining approach to job market trends using large language model. *Journal of Theoretical and Applied Information Technology* (2025).
- [10] M. Colak and B. C. Saridogan. 2023. Exploring the Remote Work Revolution: A Managerial View of the Tech Sector's Response to the New Normal. *International Journal of Contemporary Management* (2023).
- [11] C. Consiglio, N. Massa, V. Sommovigo, et al. 2022. Techno-Stress Creators, Burnout and Psychological Health among Remote Workers during the Pandemic: The Moderating Role of E-Work Self-Efficacy. *International Journal of Environmental Research and Public Health* (2022).
- [12] S. Corvite and J. Hui. 2024. Social Media as a Lens into Careers During a Changing World of Work. *Proc. of the ACM on Human-Computer Interaction* (2024).
- [13] J. H. dos Santos, J. C. S. Reis, P. F. Melo, et al. 2025. Caracterização da (Des) Informação das Narrativas sobre Vacinação em Grupos Públicos no WhatsApp no Brasil. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC.
- [14] S. Fodor and B. Boncz. 2025. InGPT we trust: perceptions of the future of work with artificial intelligence on online forums. *Data Science and Management* (2025).
- [15] Stephen L. France. 2024. Navigating software development in the ChatGPT and GitHub Copilot era. *Business Horizons* (2024). <https://www.sciencedirect.com/science/article/pii/S0007681324000697>
- [16] K. Gama, G. Liebel, M. Goulão, et al. 2025. A Socio-Technical Grounded Theory on the Effect of Cognitive Dysfunctions in the Performance of Software Developers with ADHD and Autism. In *Proc. Int. Conf. on Software Engineering: Software Engineering in Society (ICSE-SEIS)*.
- [17] P. Guenes, R. Tomaz, B. Trinkenreich, et al. 2025. Impostor Phenomenon Among Software Engineers: Investigating Gender Differences and Well-Being. In *6th Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering (GEICSE)*.
- [18] M. Hollander, D. A. Wolfe, and E. Chicken. 2013. *Nonparametric statistical methods*. John Wiley & Sons.
- [19] E. Kaloudis, V. Kouti, F. Triantafyllou, et al. 2025. AI-Powered Analysis of Weight Loss Reports from Reddit: Unlocking Social Media's Potential in Dietary Assessment. *Nutrients* (2025).
- [20] M. Kosar and F. Lee. 2024. Using topic modeling to identify factors influencing job satisfaction in the it industry. *J Inf Syst Appl Res* (2024).
- [21] M. S. Locatelli, A. S. da Costa, V. Thome, et al. 2025. From Inclusion to Contention: Analyzing DEI and “Woke” Narratives on Reddit. In *Int. Conf. on Advances in Social Networks Analysis and Mining*. Springer.
- [22] C Madding, A. G. Ansari, C. Ballenger, and A. Thota. 2020. Topic Modeling to Understand Technology Talent. *SMU Data Science Review* (2020).

- [23] P. S. Malheiros, R. R. Lima, and A. C. Oran. 2024. Impact of Generative AI Technologies on Software Development Professionals' Perceptions of Job Security. In *Proc. of the XXIII Brazilian Symposium on Software Quality*. ACM.
- [24] T. Massoni, N. Ginani, W. Silva, et al. 2019. Relating voluntary turnover with job characteristics, satisfaction and work exhaustion: an initial study with Brazilian developers. In *Proc. Int. Workshop Cooperative Human Aspects Softw. Eng. (CHASE)*. IEEE.
- [25] S. Z. S. Naing and P. Udomwong. 2024. Public opinions on ChatGPT: An analysis of Reddit discussions by using sentiment analysis, topic modeling, and SWOT analysis. *Data Intelligence* (2024).
- [26] S.E. V. S. Pillai, K. Polimetla, R. Avacharmal, and year = 2022 others, journal = Journal of Recent Trends in Computer Science and Engineering (JRTCSE). [n.d.]. Mental health in the tech industry: Insights from surveys and NLP analysis. ([n.d.]).
- [27] S. Popoola, A. Vollem, and I. K. Nti. 2025. What do Computing Interns Discuss Online? An Empirical Analysis of Reddit Posts. In *Proc. ACM Conf. on Cybersecurity & IT Education (SIGCITE)*.
- [28] G. Prenassi, A. Machado, E. Souza, et al. 2025. Narrativas de Jogos de Azar em Plataformas de Vídeo: Um Estudo Linguístico-Temático sobre o jogo do Tigrinho no YouTube. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC.
- [29] J. Ribas, J. Carneiro, and T. andothers Sousa. 2025. Pain in a Safe Space: Mapping Emotions and Discourse in theWomenintech Subreddit. In *Simpósio Brasileiro de Engenharia de Software (SBES)*.
- [30] N. Salari, M. Beiromvand, A. Hosseini-Far, et al. 2025. Impacts of generative artificial intelligence on the future of labor market: A systematic review. *Computers in Human Behavior Reports* (2025).
- [31] C. M. Sum, A. Konvicka, M. Wang, and S. E. Fox. 2025. The future of tech labor: How workers are organizing and transforming the computing industry. *Proc. of the ACM on Human-Computer Interaction* (2025).
- [32] M. Tomprou, L. Dabbish, R. E. Kraut, et al. 2019. Career Mentoring in Online Communities: Seeking and Receiving Advice from an Online Community. In *Proc. of the CHI Conf. on Human Factors in Computing Systems*. ACM.
- [33] T. S. Vaillant, F. D. Almeida, P. A. M. S. Neto, et al. 2024. Developers' Perceptions on the Impact of ChatGPT in Software Development: A Survey. arXiv:2405.12195 <https://arxiv.org/abs/2405.12195>
- [34] Watchful1. 2025. Subreddit Comments/Submissions (2005–2025). Academic Torrents. <https://academictorrents.com/details/3d426c47c767d40f82c7ef0f47c3acacedd2bf44> Dump no longer publicly available.
- [35] L. Yang and M. Pigultong. 2024. Sentiment Analysis on Job Descriptions in the Technology Sector: Measuring Positive and Negative Perceptions of Companies Using Natural Language Processing Techniques. *International Journal for Applied Information Management* (2024).

A USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

In this work, `gemma-3.5-flash` was used to create the thematic organization of BERTopic two-stage application outputs (prompts documented in the appendix). Additionally, it served to construct the engagement metric distribution per macro-topic, as shown in Table 2. Nano Banana 2 was employed strictly as a conceptual visual aid for (Figure 1, which was manually designed by the authors. Finally, a GPT model assisted with grammar refinement and text translation.

A.1 LLM Prompt for Domain Discovery

The LLM received each BERTopic topic through three complementary representations extracted from Reddit publication titles: the default topic representation, KeyBERT-based keywords, and a diversity-enhanced representation generated using *Maximal Marginal Relevance* (MMR, $\lambda = 0.8$). These

representations were provided jointly to support decisions based on both central topic terms and alternative representative terms.

The model was instructed to identify 4–6 broad themes representing the overall discussion ecosystem. Each topic had to be assigned to exactly one theme, with themes grouping multiple topics around a shared discussion area. Instructions emphasized high-level, semantically coherent groupings, concise academic-style names (2–4 words), and a consistent level of abstraction across themes. The model was also required to use all three representations jointly rather than assigning topics based solely on keyword overlap.

Expected output: Theme: `<name>`; Description: `<one-sentence summary>`; Topics: `<topic ids>`.

A.2 LLM Prompt for Macro-Topic Grouping

The LLM received the same three complementary BERTopic representations described above, extracted from Reddit comments. It was instructed to group topics according to their underlying discussion subjects rather than surface-level keyword overlap. Topic membership was determined by jointly considering all representations to identify semantically coherent discussion areas.

The prompt requested 6–10 macro-topics, each containing at least two topics, while requiring every topic to be assigned to exactly one group. Instructions emphasized preserving meaningful analytical distinctions between related topics and prioritizing interpretability over maximum semantic compression. Macro-topic names were required to be concise and descriptive, and groupings were expected to reflect coherent higher-level discussion structures.

Expected output: Macro-Topic: `<name>`; Description: `<one-sentence summary>`; Topics: `<topic ids>`.

B ETHICAL CONSIDERATIONS

This study used textual data that were publicly available at the time of data collection, without direct clinical or experimental intervention involving human participants, complying with applicable provisions of Brazilian National Health Council Resolution No. 510/2016.

In accordance with the Brazilian General Data Protection Law (LGPD – Law No. 13,709/2018), measures were adopted to reduce the risks of user identification and exposure. The data were analyzed exclusively at an aggregate level, and no individual users are identified or reported in this study. No personal data were collected or used for the purpose of identifying users. Furthermore, the raw dataset is not publicly available or redistributed and was stored in a secure environment with access restricted to the authors.